CODEN: JASMAN 4966

THE JOURNAL of the Acoustical Society of America

Vol. 106, No. 3, Pt. 1 September 1999

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

ACOUSTICS RESEARCH LETTERS ONLINE

CUMULATIVE AUTHOR INDEX

NOTES CONCERNING ARTICLE ABSTRACTS

1. The number following the abstract copyright notice is a Publisher Item Identifier (PII) code that provides a unique and concise identification of each individual published document. This PII number should be included in all document delivery requests for copies of the article.
2. PACS numbers are for subject classification and indexing. See June and December issues for detailed listing of acoustical classes and subclasses.
3. The initials in brackets following the PACS numbers are the initials of the JASA Associate Editor who accepted the paper for publication.

A frequency domain stereophonic acoustic echo canceler exploiting the coherence between the channels

Jacob Benesty

Bell Laboratories, Murray Hill, New Jersey 07974
jbenesty@bell-labs.com

André Gilloire

France Telecom CNET, France
andre.gilloire@cnet.francetelecom.fr

Yves Grenier

ENST/SIG, France
grenier@sig.enst.fr

Abstract: Stereophonic acoustic echo cancellation (AEC) is typically intended for use in high quality teleconferencing systems and multi-participant desktop conferencing that implement sound transmission through two channels. Because of the high correlation between the two channel signals, rapidly converging adaptive filter algorithms (such as the two-channel fast recursive least-squares - FRLS) are required. In this paper, we propose a new and efficient frequency domain adaptive algorithm that achieves this goal.

PACS numbers: 43.60.Gk, 43.60.Lq

1. Introduction

Stereophonic AEC can be viewed as a straightforward generalization of the single-channel acoustic echo cancellation principle.[1] Figure 1 shows this technique for one microphone in the *receiving* room (which is represented by the two echo paths, h_1 and h_2, between the two loudspeakers and the microphone). The two reference signals, x_1 and x_2, from the transmission room are obtained by two microphones in the case of teleconferencing. These signals are derived by filtering from a common source, which gives rise to a non-uniqueness problem that does not arise for the single-channel AEC.[1,2] As a result, conventional two-channel least mean square (LMS) type adaptive algorithms converge very slowly to the solution, and the two-channel FRLS algorithm (which is very complex and unstable) must be used. This requirement implies a high level of computational complexity, so that a real-time implementation of this algorithm is difficult.

It is important to devise an adaptive algorithm that takes into account the high correlation between the two channel signals or, equivalently, the coherence in the frequency domain to speed up its convergence rate. In this paper, we propose a new frequency domain adaptive algorithm that converges to a low level misalignment and is much less complex than the two-channel FRLS.

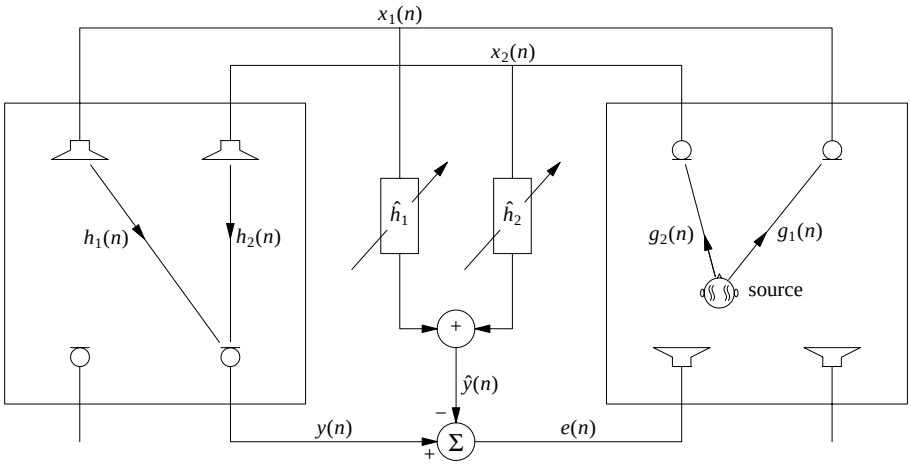


Fig. 1. Schematic diagram of stereophonic acoustic echo cancellation.

2. The extended least mean square (ELMS) algorithm

We recall here the ELMS, which is a direct approximation of the RLS algorithm; for more details see[3]. The error signal at time n is given by

$$e(n) = y(n) - \hat{y}_1(n) - \hat{y}_2(n), \tag{1}$$

where

$$\hat{y}_i(n) = \hat{\mathbf{h}}_i^T(n-1)\mathbf{x}_i(n),\ i = 1, 2, \tag{2}$$

$$\hat{\mathbf{h}}_i(n) = \left[\ \hat{h}_{0,i}(n) \quad \hat{h}_{1,i}(n) \quad \cdots \quad \hat{h}_{L-1,i}(n)\ \right]^T,$$

$$\mathbf{x}_i(n) = \left[\ x_i(n) \quad x_i(n-1) \quad \cdots \quad x_i(n-L+1)\ \right]^T.$$

L is the length of the modeling filters. The two filters are updated as follows:

$$\hat{\mathbf{h}}_1(n) = \hat{\mathbf{h}}_1(n-1) + \mu r_1^{-1}[\mathbf{x}_1(n) - \rho r_{1,2} r_{2,2}^{-1}\mathbf{x}_2(n)]e(n), \tag{3}$$

$$\hat{\mathbf{h}}_2(n) = \hat{\mathbf{h}}_2(n-1) + \mu r_2^{-1}[\mathbf{x}_2(n) - \rho r_{2,1} r_{1,1}^{-1}\mathbf{x}_1(n)]e(n), \tag{4}$$

with

$$r_{i,j}(n) = \mathbf{x}_i^T(n)\mathbf{x}_j(n),\ i, j = 1, 2, \tag{5}$$

$$r_i(n) = r_{i,i}(n)[1 - \rho^2 k^2(n)],\ i = 1, 2, \tag{6}$$

where

$$k(n) = \frac{r_{1,2}(n)}{\sqrt{r_{1,1}(n) r_{2,2}(n)}} \tag{7}$$

is the cross-correlation coefficient. We suppose in the following that $0 < \mu < 1$ and $0 \leq \rho \leq 1$. This algorithm is interesting because it introduces the cross-correlation coefficient between the two input signals. That can easily be exploited, as we will see later, to derive a frequency domain adaptive filter that takes into account the coherence function between the channels at different frequencies.

3. Block version of the ELMS algorithm

In this section, we derive a block version of the ELMS, which will be the foundation of the frequency domain algorithm. Let N be an integer number. We assume that L is an integer multiple of N, i.e., $L = KN$. The block error of the ELMS is:

$$\mathbf{e}(m) = \mathbf{y}(m) - \hat{\mathbf{y}}_1(m) - \hat{\mathbf{y}}_2(m), \tag{8}$$

where m is the block time index, and

$$\begin{aligned}
\mathbf{e}(m) &= \left[\begin{array}{cccc} e(mN) & \cdots & e(mN+N-2) & e(mN+N-1) \end{array} \right]^T, \\
\mathbf{y}(m) &= \left[\begin{array}{cccc} y(mN) & \cdots & y(mN+N-2) & y(mN+N-1) \end{array} \right]^T, \\
\hat{\mathbf{y}}_j(m) &= \left[\begin{array}{ccc} \mathbf{x}_j(mN) & \cdots & \mathbf{x}_j(mN+N-1) \end{array} \right]^T \hat{\mathbf{h}}_j(m-1).
\end{aligned}$$

We can easily show that[4,5]

$$\hat{\mathbf{y}}_j(m) = \sum_{i=0}^{K-1} \mathbf{T}_{i,j}(m)\hat{\mathbf{h}}_{i,j}(m-1), \ j = 1, 2, \tag{9}$$

where

$$\mathbf{T}_{i,j}(m) = \left[\begin{array}{ccc} x_j(mN-Ni) & \cdots & x_j(mN-Ni-N+1) \\ \cdots & \cdots & \cdots \\ x_j(mN-Ni+N-1) & \cdots & x_j(mN-Ni) \end{array} \right]$$

is an $(N \times N)$ Toeplitz matrix and

$$\hat{\mathbf{h}}_{i,j} = \left[\begin{array}{cccc} \hat{h}_{Ni,j} & \hat{h}_{Ni+1,j} & \cdots & \hat{h}_{Ni+N-1,j} \end{array} \right]^T, \ i = 0, 1, ..., K-1, \ j = 1, 2,$$

are the sub-filters of $\hat{\mathbf{h}}_1$ and $\hat{\mathbf{h}}_2$. Suppose for now that the coefficients $r_{i,j}$ are fixed, so that the filters are adapted according to

$$\hat{\mathbf{h}}_{i,1}(m) = \hat{\mathbf{h}}_{i,1}(m-1) + \mu_{\mathrm{b}} r_1^{-1}[\mathbf{T}_{i,1}^T(m) - \rho r_{1,2} r_{2,2}^{-1} \mathbf{T}_{i,2}^T(m)]\mathbf{e}(m), \tag{10}$$

$$\hat{\mathbf{h}}_{i,2}(m) = \hat{\mathbf{h}}_{i,2}(m-1) + \mu_{\mathrm{b}} r_2^{-1}[\mathbf{T}_{i,2}^T(m) - \rho r_{2,1} r_{1,1}^{-1} \mathbf{T}_{i,1}^T(m)]\mathbf{e}(m). \tag{11}$$

From this formulation, the frequency domain adaptive algorithm is straightforward.

4. The proposed algorithm

It is well known that by doubling its size a Toeplitz matrix $\mathbf{T}$ can be transformed to a circulant matrix

$$\mathbf{C} = \left[\begin{array}{cc} \mathbf{T}' & \mathbf{T} \\ \mathbf{T} & \mathbf{T}' \end{array} \right]$$

where $\mathbf{T}'$ is also a Toeplitz matrix. Using circulant matrices, the block ELMS can be rewritten equivalently.
Block error signal:

$$\left[\begin{array}{c} \mathbf{0}_{N\times 1} \\ \mathbf{e}(m) \end{array} \right] = \left[\begin{array}{c} \mathbf{0}_{N\times 1} \\ \mathbf{y}(m) \end{array} \right] - \mathbf{W}[\hat{\mathbf{y}}_1'(m) + \hat{\mathbf{y}}_2'(m)], \tag{12}$$

where

$$\mathbf{W} = \begin{bmatrix} \mathbf{0}_{N\times N} & \mathbf{0}_{N\times N} \\ \mathbf{0}_{N\times N} & \mathbf{I}_{N\times N} \end{bmatrix},$$

$$\hat{\mathbf{y}}'_j(m) = \sum_{i=0}^{K-1} \mathbf{C}_{i,j}(m) \begin{bmatrix} \hat{\mathbf{h}}_{i,j}(m-1) \\ \mathbf{0}_{N\times 1} \end{bmatrix}, \ j = 1, 2, \tag{13}$$

$$\mathbf{C}_{i,j}(m) = \begin{bmatrix} \mathbf{T}'_{i,j}(m) & \mathbf{T}_{i,j}(m) \\ \mathbf{T}_{i,j}(m) & \mathbf{T}'_{i,j}(m) \end{bmatrix},$$

and

$$\mathbf{T}'_{i,j}(m) = \begin{bmatrix} x_j(mN - Ni - N) & \cdots & x_j(mN - Ni + 1) \\ \cdots & \cdots & \cdots \\ x_j(mN - Ni - 1) & \cdots & x_j(mN - Ni - N) \end{bmatrix}.$$

Adaptation:

$$\begin{bmatrix} \hat{\mathbf{h}}_{i,1}(m) \\ \mathbf{0}_{N\times 1} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{h}}_{i,1}(m-1) \\ \mathbf{0}_{N\times 1} \end{bmatrix} + \mu_{\mathrm{b}} \mathbf{W} r_1^{-1} [\mathbf{C}^T_{i,1}(m) - \rho r_{1,2} r_{2,2}^{-1} \mathbf{C}^T_{i,2}(m)] \begin{bmatrix} \mathbf{0}_{N\times 1} \\ \mathbf{e}(m) \end{bmatrix}, \tag{14}$$

$$\begin{bmatrix} \hat{\mathbf{h}}_{i,2}(m) \\ \mathbf{0}_{N\times 1} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{h}}_{i,2}(m-1) \\ \mathbf{0}_{N\times 1} \end{bmatrix} + \mu_{\mathrm{b}} \mathbf{W} r_2^{-1} [\mathbf{C}^T_{i,2}(m) - \rho r_{2,1} r_{1,1}^{-1} \mathbf{C}^T_{i,1}(m)] \begin{bmatrix} \mathbf{0}_{N\times 1} \\ \mathbf{e}(m) \end{bmatrix}. \tag{15}$$

It is also well known that a circulant matrix is easily decomposed as follows: $\mathbf{C} = \mathbf{F}^{-1}\mathbf{D}\mathbf{F}$, where $\mathbf{F}$ is the Fourier matrix and $\mathbf{D}$ is a diagonal matrix whose elements are the Fourier transform of the first column of $\mathbf{C}$. Using this decomposition and the following time shift relationship[4]

$$\mathbf{D}_{i,j}(m) = \mathbf{D}_{i-1,j}(m-1), \ i = 1, 2, ..., K, \ j = 1, 2, \tag{16}$$

(where $\mathbf{D}_{i,j}(m) = \mathbf{F}\mathbf{C}_{i,j}(m)\mathbf{F}^{-1}$) we deduce an efficient frequency domain adaptive algorithm:
filtering:

$$\underline{\hat{\mathbf{y}}}'_j(m) = \sum_{i=0}^{K-1} \mathbf{D}_{0,j}(m-i) \underline{\hat{\mathbf{h}}}_{i,j}(m-1), \ j = 1, 2. \tag{17}$$

error signal:

$$\underline{\mathbf{e}}(m) = \underline{\mathbf{y}}(m) - \mathbf{F}\mathbf{W}\mathbf{F}^{-1}[\underline{\hat{\mathbf{y}}}'_1(m) + \underline{\hat{\mathbf{y}}}'_2(m)]. \tag{18}$$

adaptation:

$$\underline{\hat{\mathbf{h}}}_{i,1}(m) = \underline{\hat{\mathbf{h}}}_{i,1}(m-1) + \mu_{\mathrm{b}} \mathbf{F}\mathbf{W}\mathbf{F}^{-1}\mathbf{S}_1^{-1}[\mathbf{D}^*_{i,1}(m) - \rho \mathbf{S}_{1,2}\mathbf{S}_{2,2}^{-1}\mathbf{D}^*_{i,2}(m)]\underline{\mathbf{e}}(m), \tag{19}$$

$$\underline{\hat{\mathbf{h}}}_{i,2}(m) = \underline{\hat{\mathbf{h}}}_{i,2}(m-1) + \mu_{\mathrm{b}} \mathbf{F}\mathbf{W}\mathbf{F}^{-1}\mathbf{S}_2^{-1}[\mathbf{D}^*_{i,2}(m) - \rho \mathbf{S}_{2,1}\mathbf{S}_{1,1}^{-1}\mathbf{D}^*_{i,1}(m)]\underline{\mathbf{e}}(m), \tag{20}$$

where we have used the frequency domain quantities:

$$\begin{aligned}
\hat{\underline{\mathbf{y}}}'_j(m) &= \mathbf{F}\hat{\mathbf{y}}'_j(m), \\
\hat{\underline{\mathbf{h}}}_{i,j}(m) &= \mathbf{F}\begin{bmatrix} \hat{\mathbf{h}}_{i,j}(m) \\ \mathbf{0}_{N\times 1} \end{bmatrix}, \\
\underline{\mathbf{e}}(m) &= \mathbf{F}\begin{bmatrix} \mathbf{0}_{N\times 1} \\ \mathbf{e}(m) \end{bmatrix}, \\
\underline{\mathbf{y}}(m) &= \mathbf{F}\begin{bmatrix} \mathbf{0}_{N\times 1} \\ \mathbf{y}(m) \end{bmatrix}.
\end{aligned}$$

Note that if $\mathbf{S}_i = r_i\mathbf{I}_{2N\times 2N}$ and $\mathbf{S}_{i,j} = r_{i,j}\mathbf{I}_{2N\times 2N}$, then this algorithm is strictly equivalent to the block ELMS. For $\rho = 0$, we find the basic two-channel *multidelay filter* (MDF).[5]

The advantage of this approach is that the coefficients of the diagonal matrices $\mathbf{S}_i,\ i = 1, 2$, can be computed individually according to the power spectrum ($\mathbf{S}_{i,i}$), cross-power spectrum ($\mathbf{S}_{i,j}$), and coherence in the corresponding frequency bin. Hence:

$$\mathbf{S}_{i,j}(m) = \beta\mathbf{S}_{i,j}(m-1) + (1-\beta)\mathbf{D}^*_{0,i}(m)\mathbf{D}_{0,j}(m),\ i,j = 1,2, \tag{21}$$

$$\mathbf{S}_i(m) = \mathbf{S}_{i,i}(m)[\mathbf{I}_{2N\times 2N} - \rho^2\mathbf{U}^*(m)\mathbf{U}(m)],\ i = 1,2, \tag{22}$$

where β is a smoothing factor and

$$\mathbf{U}(m) = [\mathbf{S}_{1,1}(m)\mathbf{S}_{2,2}(m)]^{-1/2}\mathbf{S}_{1,2}(m) \tag{23}$$

is the (diagonal) coherence matrix. It is interesting to compare Eq. (22) and Eq. (6). This algorithm has a strong link to the two-channel RLS.[6] There are much more sophisticated ways to estimate Eq. (21). One is to use the Welch method, which gives very good results in practice with speech signals.[5]

Arithmetic complexity:
Suppose that the block length N is a power of 2, i.e., $N = 2^b$. We also assume that the FFT is computed using the split radix algorithm.[7] Then the number of operations to be performed per output point for the proposed algorithm is roughly: $4K(b+12)+4b$ real multiplications and $4K(3b+10)+12b$ real additions. For example with $L = 1024$ and $N = 256$ the computational cost of the proposed algorithm is about 8 times smaller than the two-channel NLMS.

The proposed algorithm can be easily generalized to the *generalized multidelay filter* (GMDFα),[8] where α is the overlap factor. This structure is very useful in the context of adaptive filtering, because the filter coefficients are adapted more frequently (every $M = N/\alpha$ samples instead of every N samples). As a result, a faster convergence rate and a better tracking are expected. However, the complexity is increased by a factor α. Of course, we can also derive an unconstrained version to reduce the complexity.[9]

5. Conclusions

We presented a new and efficient (at least in theory) frequency domain adaptive algorithm exploiting the coherence between the input channels to reduce the detrimental effect of their correlation. The low complexity and the good behavior of the proposed algorithm will probably make it a very good candidate for stereophonic and, more generally, multichannel acoustic echo cancellation.

References

[1] M. M. Sondhi, D. R. Morgan, and J. L. Hall, "Stereophonic acoustic echo cancellation—An overview of the fundamental problem," IEEE Signal Processing Lett. **2**, 148-151 (1995).

[2] J. Benesty, D. R. Morgan, and M. M. Sondhi, "A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation," IEEE Trans. Speech Audio Processing **6**, 156-165 (1998).

[3] J. Benesty, F. Amand, A. Gilloire, and Y. Grenier, "Adaptive filtering algorithms for stereophonic acoustic echo cancellation," in *Proc. IEEE ICASSP, 1995*, pp. 3099-3102.

[4] J. Benesty and P. Duhamel, "Fast constant modulus adaptive algorithm," IEE Proc.-F **138**, 379-387 (1991).

[5] J.-S. Soo and K. K. Pang, "Multidelay block frequency domain adaptive filter," IEEE Trans. Acoust., Speech, Signal Processing **38**, 373-376 (1990).

[6] J. Benesty, P. Duhamel, and Y. Grenier, "Multi-channel adaptive filtering applied to multi-channel acoustic echo cancellation," in *Proc. EUSIPCO, 1996.*

[7] P. Duhamel, "Implementation of split-radix FFT algorithm for complex, real and real-symmetric data," IEEE Trans. Acoust., Speech, Signal Processing **34**, 285-295 (1986).

[8] E. Moulines, O. Ait Amrane, and Y. Grenier, "The generalized multidelay adaptive filter: structure and convergence analysis," IEEE Trans. Signal Processing **43**, 14-28 (1995).

[9] D. Mansour and A. H. Gray, JR., "Unconstrained frequency-domain adaptive filter," IEEE Trans. Acoust., Speech, Signal Processing **30**, 726-734 (1982).

Evolution of the Journal: Early history and overview

Allan D. Pierce
Acoustical Society of America, Office of the Editor-in-Chief, P.O. Box 323, East Sandwich, Massachusetts 02537

(Received 8 July 1999; accepted for publication 8 July 1999)

The new Editor-in-Chief of the Acoustical Society of America gives an editorial discussion of the history of the *Journal of the Acoustical Society of America*. Emphasis is given to the earlier years in the *Journal*'s 70 year history and to the evolution of the Office of Editor-in-Chief. A dominant driving force throughout the history of the *Journal* has been that of fulfilling the Society's purpose—to increase and diffuse the knowledge of acoustics and to promote its practical applications. © *1999 Acoustical Society of America.* [S0001-4966(99)05309-6]

PACS numbers: 43.05.Dr, 43.10.Gi [ADP]

INTRODUCTION

R. Bruce Lindsay, Editor-in-Chief of the Acoustical Society of America from 1957 to 1985, once wrote[1]

> *Students have a right and probably an obligation to know the historical origins of...*

This is taken out of context, but Lindsay would probably concur with its use here. Various members of the Society frequently raise questions (1) as to why the *Journal* is as it is today and (2) as to what might be changed to improve it. The first question is comparatively easy to answer, although still a difficult one. It is not absolutely necessary that the first question be addressed first, but knowing the answer in at least a broad sense should be helpful when one addresses the second question. Answering the first requires that one know something about the history. Liberal acceptance of Lindsay's philosophy implies in turn that the members have a right and probably an obligation to know the history of the *Journal*. The Editor-in-Chief probably has the obligation to transmit the salient historical details in a form that is palatable to the members.

I. ORIGINS OF THE SOCIETY

In the summer of 1928, three individuals (Waterfall, Watson, and Knudsen) met and initiated a plan of action that quickly resulted in the formation[2,3] of the Acoustical Society of America. An organizational meeting with forty attendees took place, in conjunction with a national meeting of the American Association for the Advancement of Science, on December 27, 1928.

Just how the discussion progressed at that historic meeting is now largely a matter of conjecture, but the question of a journal was undoubtedly discussed. Many years later, E. C. Wente, one of the prominent early members of the Society, on the occasion[4] of the presentation of the Gold Medal to Harvey Fletcher, wrote the following:

> ''*I don't know what was discussed at this meeting but I do know that [Harvey Fletcher, who was to become the Society's first President,] was strongly in favor of the organization of an acoustical society broad enough in its objectives to embrace practically all fields of acoustics, patterned after the American Physical Society with regularly scheduled meetings and a regularly issued journal. Some of his advisors felt that he might be going out on a limb with such an ambitious program. They warned that, while there was a certain amount of accumulated material worthy of presentation and publication which might carry the society along for a while, thereafter it might face starvation.*''

Optimism along with a faith in the founders' steadfastness of purpose must have prevailed. The actual substantive outcome of the meeting was that a tentative slate of officers was elected and a committee of ten was appointed to draft a constitution and by-laws.

The constitutional committee subsequently met and carried out their delegated task. Of considerable relevance to the present discussion is Article II in the Constitution,[5] which stated

> *The purpose of the Society is to increase and diffuse the knowledge of acoustics and promote its practical applications.*

Harvey Fletcher later[6] stated: ''It was only after considerable debate and trimming that we were able to reduce the purpose ... to a single sentence.'' This result of their labors is, in this writer's opinion, magnificent—a rallying cry that has guided the actions of the world's first and greatest professional society devoted to acoustics. The Constitution and the By-Laws have been amended many times over the years, but the Purpose has never been changed. (The present author would have preferred the wording ''and *to* promote'' rather than ''and promote'', but every Liberty Bell worth preserving is entitled to at least one crack.)

The Society held its first meeting on May 9–11, 1929. This meeting was in conjunction with a meeting of the Society of Motion Picture Engineers. (By all accounts, the common misconception that the Acoustical Society of America is an offshoot of the American Physical Society (APS) is erroneous; what is apparently true is that its organization was patterned somewhat after the APS.) At the first business session[7] of the May 1929 meeting, on the morning of May 10, the draft Constitution and By-Laws were presented. At this time the Constitution already had 376 signatures. There was a lively discussion concerning the By-Laws, but both

the Constitution and the By-Laws were adopted largely as written.

In this first version of the Constitution and the By-Laws, there was no mention of a journal. However, at the end of the continuation of the business session, which took place immediately after lunch, the newly elected President, Harvey Fletcher, called for a meeting of the newly elected Executive Council for the same day. ''There are a large number of questions to be considered, one being the question of the publication of a journal. I feel sure that we have sufficient financial support already so that we can promise you that the proceedings of this meeting will come out as the first publication of the Society. What the future plans will be, of course, will be put in the hands of a committee.''

At the subsequent Executive Council meeting, a committee[8] consisting of Wallace Waterfall (Chair), Floyd R. Watson, Paul E. Sabine, and Charles F. Stoddard was appointed to supervise the publication of a journal. (All of these persons were members of the Executive Council; Stoddard was the Society's first Treasurer; Waterfall was its first Secretary.)

The first issue (Volume 1, Number 1) of the *Journal of the Acoustical Society of America* appeared in October 1929. Much of this historic issue was devoted to the archiving of the creation of the Society, but there were eight technical articles. These were somewhat interdisciplinary, but a gross classification yields one on architectural acoustics, three on speech, two on instrumentation, and one on psychoacoustics. One of the speech papers was also concerned with architectural acoustics, another overlapped with physical acoustics; one of the instrumentation papers was concerned with musical acoustics, the other with noise. In any event, the first issue set the tone for a journal that was relating strongly to practical problems of the day.

II. THE FIRST FOUR YEARS

The *Journal* was initially a quarterly, with issues appearing in July (No. 1), October (No. 2), January (No. 3), and October (No. 4). The first volume, because the first issue appeared in October, had only three issues. The first issue of Volume 2 was in July 1930. A typical issue had of the order of 100 pages, these appearing in a single column format during the first four years. There were of the order of 8 to 10 technical articles in a typical issue. Several of the issues had a part 1 and a part 2. The second parts were sometimes numbered separately, and sometimes had technical articles. The text of the *Journal* is unclear as to the reason for the two parts and as to whether there was a clear-cut distinction between the nature of the technical articles that appeared in the two parts. However, most of the papers that did appear in the second parts were authored[9] by people affiliated with the Bell Telephone Laboratories. A remark published many years later suggests that some of these may have been reprintings of articles published in the *Bell Laboratories Record* or the *Bell System Technical Journal*. One paper, authored by Harvey Fletcher, which appeared in Volume 3, Number 2, part 2, was a paper that had previously appeared in *Reviews of Modern Physics*.

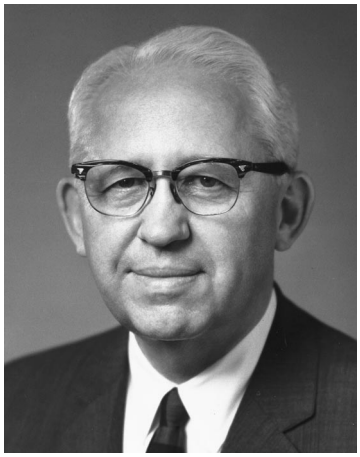


FIG. 1. Wallace Waterfall, Chair of the original committee to supervise the publication of a journal; managing editor for the Acoustical Society, 1929–1933.

The Society had no officer who carried the title of Editor during the first four years. Wallace Waterfall (Fig. 1) held the office of Secretary during those years (he held the office from 1929 to 1962) and carried out some of the functions of an editor. The only stated duties for the Secretary stipulated by the By-Laws of that time were that (Article 9) ''the Secretary shall attend all meetings of the Society and the Executive Council and shall keep a true and faithful record thereof.'' Waterfall, moreover, had another assigned task in that he was Chair of the original committee appointed at the first Executive Council meeting to ''supervise the publication of the Journal.'' Waterfall also apparently played a major role in the garnering of financial support for the launching of the *Journal*. That the founders of the Society were very successful in this respect is evidenced by Fletcher's statement mentioned above that ''I feel sure that we have sufficient financial support already so that we can promise you that the proceedings of this meeting will come out as the first publication of the Society.'' Financial support obtained for the *Journal* over the first two years from outside sources amounted to $3700. To put this in perspective, the dues then were $5.00 for members (and $7.50 for fellows), and the initial membership was of the order of 450 persons. In regard to the financial strength of the *Journal*, one should also note that the *Journal* did not then accept advertising.

Waterfall's role in regard to the *Journal* during those early years has sometimes been referred to as that of a *managing editor*. He carried out many tasks that in the present day are carried out on behalf of the Society by the American Institute of Physics. Waterfall prepared papers for the printer; he was responsible for the printing of the *Journal*, for its mailing, and for the collection of bills. He was also responsible for the design of the front cover of the *Journal*, which is largely the same as that of the front cover of the *Journal* of today.

The actual selection of the material that went into the *Journal* was largely the responsibility of an Editorial Board of which Waterfall was a member. The Society held its second meeting in December 1929, only two months after the appearance of the first issue of the *Journal*. At that meeting,

FIG. 2. Floyd R. Watson, Chair of the Editorial Board, 1929–1939; first person to hold the Office of Editor (1933–1939).

the Executive Council again took up the subject of the *Journal*. Floyd R. Watson (Fig. 2), who was then a member of the Council, was appointed Chair of an Editorial Board, who in turn selected E. C. Wente, Paul E. Sabine, H. W. Lamson, Frederick A. Saunders, and Wallace Waterfall as the other members of the Board. One should note that Waterfall and Watson were two of the three persons who were responsible for the initial impetus that led to the formation of the Society.

Some insight into just who did what in regard to the selection of the content of the *Journal* in these very early years may possibly be obtained from a consideration of the biographies[10,11] of Watson and Waterfall. Floyd Watson was born on April 23, 1872, while Wallace Waterfall was born on May 27, 1900. Thus, at the time of the historical three-man meeting in the summer of 1928, Watson was 56 and Waterfall was 28 years of age. Watson was, for virtually all of his professional life, a professor in the Physics Department at the University of Illinois; Waterfall graduated from Illinois with a bachelor's degree in engineering physics in 1923, and Waterfall gave considerable credit to Watson for awakening his early interest in acoustics. The two of them probably functioned very much as a team, with Waterfall supplying a great amount of creativity, energy, and commitment, and with Watson supplying a mature judgement and the accumulated wealth of experience of one of the world's most senior and most accomplished researchers and teachers in acoustics. [Our most recent Editor-in-Chief, Daniel Martin, was also a student of Floyd Watson.]

The membership of the Editorial Board changed frequently over the successive years, but Floyd Watson remained as its Chair until June 30, 1939. Watson sought to appoint Board members who represented ''as far as possible the various interests in acoustics.'' There was no standard definition as what these interests were, and individual members were likely to have relatively wide interests.

In those idyllic years, the ''selection'' of papers for publication in the *Journal* was anything but passive. A primary source of papers for publication was the Society's semiannual meetings. There were typically of the order of 20 to 30 technical talks at each such meeting, and there were no parallel sessions. The members of the Editorial Board attended the meetings and presumably each member heard a considerable percentage of the presentations. The Board met immediately after the Society's meeting, discussed the talks presented at the meeting, and made decisions as to what papers were to be solicited for publication. They apparently were proactive in inviting persons to present papers at the meetings and to subsequently publish written versions in the *Journal*.

Unsolicited papers were not precluded, and peer review then was far less rigid than is the common practice today. Papers by eminent senior persons, especially those by well-recognized authorities in other countries, generally sailed through with only a cursory review. The Editorial Board evidently took the ''diffuse and increase the knowledge'' clause of the *Purpose Statement* very seriously. A guiding principle was that ''the Acoustical Society was formed to stimulate activities in acoustics, even those abroad.'' In those cases where an author was perceived as either inexperienced or as not well-known, the submitted paper was read by a member of the Editorial Board, who suggested any changes that appeared necessary. Watson also strongly felt that one of the functions of the Editorial Board was to encourage young investigators and to help them get started in their work. Reviewers from outside the Board were occasionally consulted, but this was typically informal. Wente, for example, drew extensively on the advice of his immediate colleagues at the Bell Telephone Laboratories.

III. AFFILIATION WITH THE AIP

The American Institute of Physics (AIP) was established in 1931, and the Acoustical Society of America was one of its four original founding member societies. The other three were the American Physical Society (founded 1899), the Optical Society of America (founded 1916), and the Society of Rheology (founded 1929). (Some literature gives the American Association of Physics Teachers as also being an original AIP member, but a 1932 article,[12] written by the Director of the AIP, seems to give definitive evidence that the AAPT was not a member in 1931.) The AIP was originally intended as a service organization, much in the same spirit as a farm cooperative to which various farmers might have belonged. The principal service that the AIP offered was the management of the publication of scientific journals. The member societies would of course have to pay for any services at a cost commensurate with the magnitude of the services that were provided. The Executive Council at the time found the affiliation to be attractive because it solved the problem of the management of the publication of the *Journal* at a reasonable cost. As a consequence all the publication management functions that had originally been undertaken by Wallace Waterfall were taken over by the American Institute of Physics.

Beginning with Volume 5 (first issue was that of July 1933), the *Journal* displayed the statement—*published for the Acoustical Society of America by the American Institute of Physics*. The format of the Journal also changed with this issue; one noticeable feature was that the technical articles

were published with two columns on each page rather than one column.

IV. INSTITUTING THE OFFICE OF EDITOR

More or less simultaneously with the transferring of the task of publication to the AIP, the Society formally created the office of Editor. At the May 1933 meeting,[13] amendments were passed to the Constitution and By-Laws[14] that made the Editor of the *Journal* an elected officer of the Society and a member of the Executive Council. The Editor was empowered to select an Editorial Board consisting of five members, with the membership subject to approval by the Executive Council. At that time all of the officers were elected officers, including the Secretary and the Treasurer. The office of Editor differed somewhat from the others in that the term was three years rather than one year. The amendment that created the office of Editor was the first amendment to the Society's Constitution.

In view of the early history of the *Journal* as depicted in the previous portion of this article, it is apparent that Floyd Watson had been functioning as the editor of the *Journal* in the sense that the term is used today. He was a voting member of the Executive Council for the first three years of the Society's existence and the By-Laws stipulated that he was ineligible for immediate reelection. The problem thus arose that the de-facto editor of the *Journal*, who had been on the Executive Council from the very beginning, was suddenly no longer on the Executive Council. A plausible hypothesis for the impetus in the creation of the amendment is consequently that the Executive Council felt that much of their deliberations concerned the *Journal* and that they needed the editor's presence at their meetings. Given this line of reasoning, it is not surprising that the Executive Council did not wait on the electoral process to fill the office of Editor. They made use of their constitutional prerogative to appoint Watson for a term that would expire with the beginning of the first elected term. Watson was subsequently elected Editor for a term that began July 1, 1934.

V. SUCCESSION OF EDITORS

Floyd Watson was re-elected as Editor for the term which began July 1, 1937 and which was to end on June 30, 1940. He served only two of these three years because he was elected President of the Society for a term that was to begin July 1, 1939. There was no clause in the Constitution or By-Laws that precluded a person from holding both offices simultaneously; it is likely that Watson felt that he did not have the time to do justice to both offices simultaneously. His tenure as Editor consequently ended June 30, 1939. He did, however, remain on the Editorial Board, in an *ex officio* capacity, while he was President.

The Executive Council appointed Floyd A. Firestone[15] (Fig. 3) to fill Watson's unexpired term. Firestone, who was a Charter Member of the Society and who was a former student of Dayton Miller, the Society's second President, was born on 21 September 1898 and consequently was 41 years of age at the time he was appointed Editor. He was also the current Vice-President of the Society and consequently a

FIG. 3. Floyd R. Watson, Editor of the *Journal of the Acoustical Society of America,* 1939–1956; Editor-in-Chief, 1956–1957.

member of the Executive Council. In contrast with Watson, Firestone did hold two offices at the same time. He was Vice-President from 1938–1940, the second year being the same as the year he was initially appointed to serve as Editor. Firestone was also President of the Society from 1943–1945, which was a term that fell within his much longer term as Editor. Prior to being Vice-President, he had served on the Executive Council from 1932 to 1935, coming onto the Council at the same time that Watson's term had expired. Firestone's service on the Editorial Board dated back at least as far as April 1935. His externally-visible service on the Board began with the creation of a department of the *Journal* titled *Current Publications in Acoustics*, which was announced[16] in the January 1937 issue and appeared frequently with Firestone's by-line starting with the April 1937 issue (Vol. 8, No. 4). This department carried the section *References to Contemporary Papers in Acoustics*, which was originally compiled by Firestone and which continues to be a part of the *Journal* to the present date.

Chronicling all the pertinent events here that took place during Firestone's tenure (1939 to 1957) as Editor would result in a unacceptable diversion from the purpose of the present section, but there are four events whose immediate relation appears mandatory. Two of these concern the frequency of appearance of the *Journal*. Volume 18 had only two issues (July and October 1946). Starting with January 1947 and Volume 19, the *Journal* had 6 issues a year (January, March, May, July, September, and November). The six issues per year frequency ended in 1956. Beginning with January 1957 (Volume 29), the *Journal* was published with 12 issues per year.

The third event was a wholesale delegation of responsibilities to the members of the Editorial Board, who thereafter were listed as Associate Editors. This new listing appeared for the first time on the inside front cover of the January 1950 issue (Vol. 22, No. 1). The technical areas that were used in classifying articles for the *Journal*'s index then had 11 major categories and each of the five associate editors was responsible for one or more (typically two) of these categories. For example, Robert Bruce Lindsay, who was eventually to succeed Firestone as Editor, was responsible for the

categories of Ultrasonics and Underwater Sound. The March 1950 issue[17] carried a historic announcement, the text of which is partially repeated here:

> *In order to decrease the burden of mail on the Editor's office, papers for publication should henceforth not be sent to the Editor but should be sent to that particular Associate Editor having cognizance of the field of subject matter into which the paper falls ... The Associate Editor will transmit your paper to the Editor after it has been refereed and placed in final form for publication.*

Incidentally, the misconception that individual associate editors have a direct relationship to our current Technical Committees is another that is erroneous. In 1950, there was no Technical Council and no formal structure of technical committees.

The fourth event was the change of office title from Editor to Editor-in-Chief. This was apparently occasioned by anticipation of the Society's publication[18] starting with January 1955 of a second journal. This was to have its own editor, but there was some feeling that a single person should oversee all of the Society's publications. By this time[19] the original Constitution and Bylaws had been somewhat overhauled and recast into a single operating code with the simple designation, ''Bylaws.'' Consequently, a Bylaws change was proposed in which the office of Editor was to be replaced by that of the office of Editor-in-Chief. The most applicable new clause[20] was:

> *The Editor-in-Chief shall be responsible for the editorial management of all publications of the Society. He shall be assisted by such other editors and editorial boards as, upon a recommendation, are approved by the Executive Council.*

[The insensitive use here of the word *he* was of course rectified in a subsequent revision of the Bylaws.] The new Bylaws were passed[21] at the June 1955 meeting of the Society, and Firestone became the first Editor-in-Chief of the Society, continuing on with an elected term that he had begun as Editor.

The office of Editor-in-Chief still remained an elective office, and the first election for this position was by mail ballot in the spring of 1956, one year after the Bylaws change. The nominations committee[22] submitted a slate of three candidates for Editor-in-Chief, and Firestone was elected.

Floyd Firestone resigned[23] his elected position of Editor-in-Chief one year later at a regular semi-annual meeting of the Society ''to devote all his efforts to scientific work and the handling of his patent affairs.'' (He was an extremely successful inventor.) The meeting summary[24] carries statements to the effect that, at the business session, Secretary Waterfall announced that, after eighteen years of devoted and unusually successful service, Floyd A. Firestone had resigned as Editor-in-Chief and that the Executive Council had elected (or appointed) R. Bruce Lindsay (Fig. 4) to fill the unexpired term of Floyd Firestone.

As best the present writer can discern from what was printed in the *Journal* at that time, Firestone's resignation

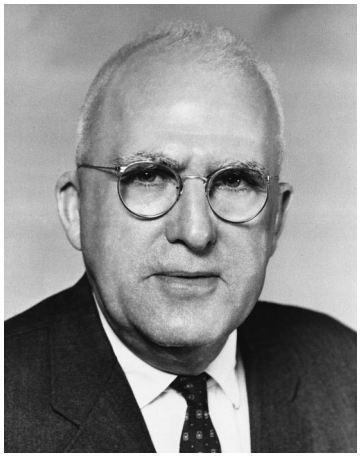


FIG. 4. R. Bruce Lindsay, Editor-in-Chief of the Acoustical Society of America, 1957–1985.

was somewhat unanticipated. It is likely that the phenomenal growth of the *Journal* over the eighteen years of his service had gotten to the point where Firestone was starting to make an assessment of whether he really wanted to continue with the commitment of such a large fraction of his time. Some criticisms[25] from the other members of the Executive Council as to the style with which he had recently handled the Acoustical Society News section may have precipitated the decision. In any event, the Executive Council had to appoint a temporary successor relatively abruptly, and the choice as stated above was Lindsay. An account of the Executive Council meeting where this occurred was related[26] some 28 years later by R. T. Beyer.

> *''I served on the Council during [Lindsay's] presidency in 1956-57. As a result, I was there ... when Bruce, as presiding officer, took up the problem of the vacancy in the position of Editor-in-Chief that had come about because of the resignation of Floyd Firestone. The Society had received one application, from a past president, and the discussion was on the floor. I remember some voice—possibly that of Dick Cook—saying ''What about Dr. Lindsay here?'' At that, Bruce excused himself from the rest of the discussion and left the room. A few minutes later, we welcomed him back as President and Editor-in-Chief.''*

Thus it is likely that Lindsay had never even contemplated serving as Editor-in-Chief. It may have also been that he took on the job with no particular thought of serving beyond the time when a new Editor-in-Chief was duly elected by the membership. Lindsay was, by all accounts, a logical choice, as in addition to his impressive credentials as an author, researcher, and administrator, he had been an associate editor of the Journal since 1950 and was continuing in this capacity at the time he was serving as President. His appointment consequently assured a certain degree of continuity, and the issues published at the time show no appreciable sign of disruption. Lindsay's name appeared on the masthead as Editor-in-Chief for the first time in July 1957.

The question of Lindsay's continuation beyond a temporary appointment was settled by the Society's election of May 1959. The nominating committee[27] proposed a slate of

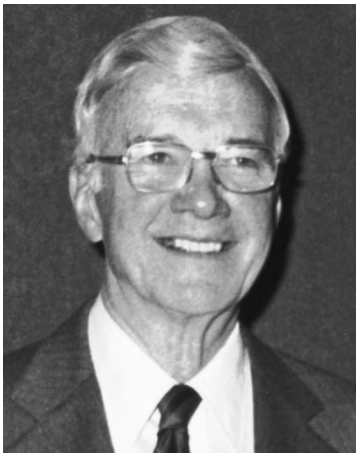


FIG. 5. Daniel W. Martin, Editor-in-Chief, 1985–1999.

two candidates for Editor-in-Chief, and Lindsay received the most votes. He began his tenure at the age of 57 and served continuously for 28 years until his death.

The position of Editor-in-Chief along with that of Treasurer ceased to be an elected office in 1969. This was voted upon at a business meeting of the Society on 9 April 1969. The new Bylaws[28] took effect[29] on 11 April at 1:00 p.m. and at that time the Executive Council, using its new authority, appointed R. Bruce Lindsay Editor-in-Chief for the first of a succession of many three year terms.

Twenty-eight years after beginning his service as Editor-in-Chief, R. Bruce Lindsay died unexpectedly on 2 March 1985. Most of the members learned of this either via an announcement[30] in the April 1985 issue or via an announcement at the plenary session that took place at the Society's meeting in April. Robert T. Beyer, who was Treasurer of the Society at the time and a long-term associate of Bruce Lindsay, assumed the job of Acting Editor-in-Chief. His name as such appears on the masthead of the May, June, July, and August issues. The September issue[31] carried the notice that Daniel W. Martin (Fig. 5) had been appointed by the Executive Council as Editor-in-Chief for a term of three years beginning with September 1, 1985.

The course of the *Journal* during the past 14 years attests to the fact that the Council's decision was a very good one. Martin, who was born in 1918, was 67 years of age at the time and was recently retired from a distinguished career at Baldwin Piano, so he was able to make a considerable commitment of time to the job of Editor-in-Chief. His involvement with the *Journal* dated back to 1951, when he first became a patent reviewer. He had been on the Editorial Board since 1977 as the Associate Editor for the Patents Review section. A long, productive, and varied record of effective service and leadership had resulted in his serving as President of the Society at the time of Lindsay's death. The nature of his professional activities and his service with the Acoustical Society and other organizations had given him an unusually broad knowledge of acoustics, and this was to prove very helpful in dealing with the variety of problems that were to present themselves in the years that followed.

The events leading to the appointment of the present Editor-in-Chief, as best he can ascertain, are as follows. At the time of the Seattle ASA/ICA meeting in June 1998, Daniel Martin made it known to the Executive Committee and to the Editorial Board that he would like to eventually step down from the Office of Editor-in-Chief. There was no expressed need for immediate action, but Martin apparently wanted to give the Council ample time to pick a successor, and he also wanted his successor to have the opportunity to work with him on an interim basis so that there would be a smooth transition. In some unpublished minutes of the Editorial Board meeting that took place at the time, a transitional period of 2 years was mentioned. The Executive Council, following the tradition that it had followed in more recent appointments of Society officers, such as that of the recent Treasurer, appointed a search committee, with the charge that the Committee identify a number of prospective candidates and that it make recommendations to the Council at some later date. An announcement of the search subsequently was printed[32] in the August 1998 issue of the Journal. The present writer did not notice this announcement and was unaware that the search was in progress until he received a letter from the Search Committee stating that he had been nominated for the position and stipulating what was required should he choose to be a candidate.

The Search Committee apparently had collected a number of applications by the time of the Norfolk meeting of the Society in October 1998. Given that the program for the Berlin meeting, that was to take place in March 1999, scheduled two meetings for the Search Committee, it is likely that the original intent was that interviews were to take place then. Possibly because of an awareness that Martin's health had taken an unexpected down-turn or possibly because of the complications of interviews carried out at a location outside the United States, the decision was made to carry out the interviews in conjunction with the semiannual as a Officers and Managers meeting in early December 1998, to take place in New York City. The actual date of the interviews was December 6, and each of the selected candidates met with the Search Committee and with a subcommittee of the Executive Council. A week later the present writer learned that Daniel Martin was in critical condition and that the Executive Council was moving rapidly to complete the selection of the next Editor-in-Chief. This was carried out by mail ballot of the Executive Council with the determination completed on December 21.

The initial plan was that the present writer was to assume the Office of Editor-in-Chief sometime towards the end (not the beginning) of 1999. Although he had previously served with Daniel Martin on both the Executive Council and the Editorial Board, he was not currently a member of either at the time of the selection. He was looking forward to an apprentice period with Dr. Martin and to learning about the issues currently confronting the publications of the Society. But, as we all now know,[33] Daniel Martin died on January 7. To some, this may not have been unexpected; to this writer, however, it was unexpected. Regardless of how much all of us wished it to take place at a considerable later date, the job of Editor-in-Chief abruptly fell upon Allan Pierce (Fig. 6) on January 7, 1999.

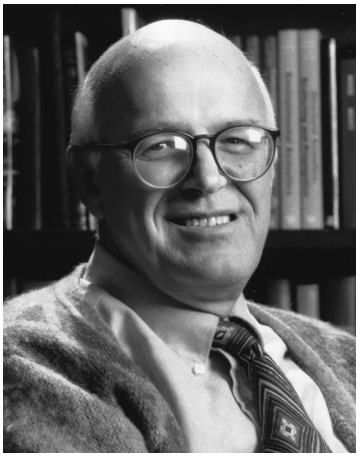


FIG. 6. Allan D. Pierce, Editor-in-Chief, as of January 1999.

VI. CONCLUDING REMARKS

Much more can of course be written (and perhaps should be written) about the history of this journal. The one enduring impression the writer has obtained from the effort whose results are portrayed here is that, up to now, the Society has been ably served by a succession of individuals that have brought its *Journal* to its present dominant role as the world's leading journal on acoustics. How we can do even better and whether we can do so remains a topic of continuing concern.

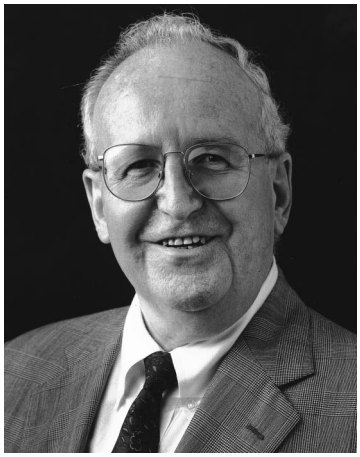
[1] R. B. Lindsay, in ''Book Reviews,'' J. Acoust. Soc. Am. **70**, 1548 (1981).
[2] W. Waterfall, ''History of the Acoustical Society of America,'' J. Acoust. Soc. Am. **1**, 5–8 (1929).
[3] W. Waterfall, ''Events leading to the formation of the Acoustical Society of America,'' J. Acoust. Soc. Am. **11**, 11–12 (1939).
[4] E. C. Wente, ''Review of Dr. Fletcher's Achievements,'' in ''1957 Gold Medal Award to Dr. Harvey Fletcher,'' J. Acoust. Soc. Am. **29**, 1036–1038 (1957).
[5] ''Constitution and By-Laws as adopted by the Society, May 10, 1929,'' J. Acoust. Soc. Am. **1**, 9–12 (1929).
[6] H. Fletcher, ''The Acoustical Society of America. Its aims and trends,'' J. Acoust. Soc. Am. **11**, 13–14 (1939).
[7] ''Proceedings of the Business Meeting,'' J. Acoust. Soc. Am. **1**, 39–46 (1929).
[8] F. R. Watson, ''The Journal of the Acoustical Society of America,'' J. Acoust. Soc. Am. **11** (part 1), 15–20 (1939).
[9] ''Reprinting of exhausted back numbers,'' J. Acoust. Soc. Am. **19**, 506 (1947).
[10] V. O. Knudsen, ''Biographical sketch of Floyd Rowe Watson,'' J. Acoust. Soc. Am. **32**(2), 286–287 (1960).
[11] R. H. Bolt, ''Wallace Waterfall, 1900–1974,'' in ''Acoustical News—USA,'' J. Acoust. Soc. Am. **56**, 1932–1933 (1974).
[12] H. A. Barton, ''The American Institute of Physics,'' J. Acoust. Soc. Am. **4** (part 1), 94–95 (1932).
[13] F. R. Watson, ''Meeting of the Society,'' in ''News and notes,'' J. Acoust. Soc. Am. **5**, 61 (1933).
[14] ''Constitution and By-Laws and Membership List of the Acoustical Society of America,'' J. Acoust. Soc. Am. **7** (part 2), 151–166 (1935).
[15] W. Waterfall, in ''1955 Gold Medal Award to Dr. Floyd A. Firestone,'' in ''Acoustical Society News,'' J. Acoust. Soc. Am. **27**, 976–977 (1955).
[16] ''Acoustical Society News,'' J. Acoust. Soc. Am. **8**, 204 (1937).
[17] ''Send your paper to the cognizant associate editor, not to the editor,'' J. Acoust. Soc. Am. **22**, 292 (1950).
[18] L. L. Beranek, ''New Acoustical Society publication, Noise Control,'' in ''Acoustical Society News,'' J. Acoust. Soc. Am. **26**, 906 (1954).
[19] ''By-Laws (as amended 1947),'' J. Acoust. Soc. Am. **19** (part 2), 740–742 (1947).
[20] W. Waterfall, ''Proposed amendments to the Bylaws,'' J. Acoust. Soc. Am. **27**, 388 (1955).
[21] F. V. Hunt, ''The Penn State Meeting,'' J. Acoust. Soc. Am. **27**, 979 (1955).
[22] W. Waterfall, ''Announcement of forthcoming election,'' in ''Acoustical Society news,'' J. Acoust. Soc. Am. **28**, 134–135 (1956).
[23] L. L. Beranek, H. F. Olson, and R. W. Young, ''Floyd A. Watson resigns office of Editor-in-Chief,'' J. Acoust. Soc. Am. **29**, 1035–1036 (1957).
[24] R. B. Lindsay, ''The fifty-third meeting of the Society,'' J. Acoust. Soc. Am. **29**, 1036 (1957).
[25] F. A. Firestone, ''Reverberation,'' J. Acoust. Soc. Am. **27**, 978 (1955).
[26] R. T. Beyer and others, ''In Memoriam—R. Bruce Lindsay: Memories of R. Bruce Lindsay,'' J. Acoust. Soc. Am. **78**, 1463–1474 (1985).
[27] W. Waterfall, ''Announcement of forthcoming election,'' J. Acoust. Soc. Am. **30**, 1167–1168 (1958).
[28] W. Waterfall, ''Proposed new bylaws,'' J. Acoust. Soc. Am. **44**(6), 1755–1757 (1968).
[29] E. Shaw, ''77th Meeting; 1969; Philadelphia,'' J. Acoust. Soc. Am. **46** (part 2), 222–223 (1969).
[30] ''R. Bruce Lindsay, 1900–1985,'' J. Acoust. Soc. Am. **77**, unnumbered page, appearing immediately behind the front cover of the April issue, bound between pp. 1294–1295 (1985).
[31] B. Goodfriend, ''New Editor-in-Chief Appointed,'' J. Acoust. Soc. Am. **78**, v (September 1985).
[32] ''Editorial Announcement,'' J. Acoust. Soc. Am. **104** (part 2), 601 (1998).
[33] J. E. West, ''A message from the President,'' Echoes **9**, 3 (Winter 1999). The text of this message was originally sent out as a broadcast e-mail message to all ASA members on 15 January 1999.

ACOUSTICAL NEWS—USA

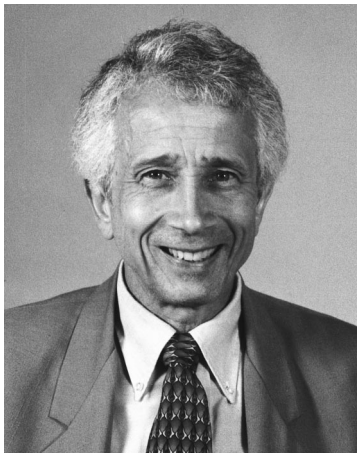
Elaine Moran
Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502

Editor's Note: *Readers of this Journal are asked to submit news items on awards, appointments, and other activities about themselves or their colleagues. Deadline dates for news items and notices are 2 months prior to publication.*

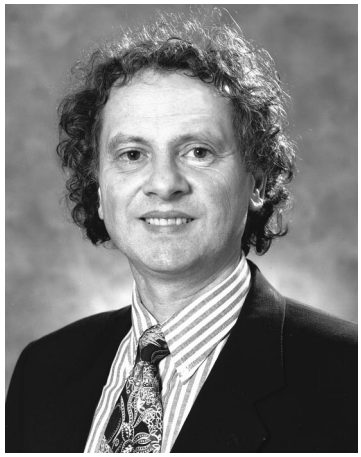
New Fellows of the Acoustical Society of America



Robert E. Greene, Jr.—For contributions to acoustical nondestructive evaluation of materials.



Robert D. Hellweg, Jr.—For contributions to noise measurement standards and applications.



Peter B. Nagy—For contributions to ultrasonic materials characterization and nondestructive evaluation.

Elliott H. Berger elected a Fellow of American Industrial Hygiene Association

Elliott H. Berger, Fellow of the Acoustical Society, has been elected a Fellow of the American Industrial Hygiene Association (AIHA). AIHA fellows are nominated by a local section or standing committee for their recognized achievement in the field of industrial hygiene, either through research, leadership, publications, education, or service to AIHA. Founded in 1939, AIHA is an organization of more than 13 000 professional members dedicated to the anticipation, recognition, evaluation, and control of environmental factors arising in or from the workplace that may result in injury, illness, impairment, or affect the well-being of workers and members of the community.

ASA Presents Awards to Young Scientists at the 50th International Science and Engineering Fair

The Acoustical Society of America (ASA) presented five special awards at the 50th International Science and Engineering Fair (ISEF) which was held 2–8 May in Philadelphia, to the students whose projects were most outstanding in the field of acoustics. The ISEF, now in its 50th year, is the only international science project competition for students in grades 9–12. This year the Fair attracted almost 1300 attendees from more than 47 countries. The students were competing for a variety of governmental, industrial, and educational awards and scholarships and special awards from organizations including ASA.

The top winning prize of $300 went to Jeffrey S. Martzall, 17, of Northrop High School, Fort Wayne, IN, for his project entitled ''2-D Audio Tracking for Aviation Guidance Systems.'' Jeff designed and constructed a model of an aircraft fully controlled by acoustic transmitters and receivers. The second prize of $200 was awarded to Matthew O. Hart, 17, of Alfred M. Barbe High School, Lake Charles, LA, who presented a project entitled ''Design and Construction of a Pseudovision Device for the Sight Impaired.'' Matt's device makes use of echo-ranging sonar and Matt hopes that it will significantly improve the ability of the sight-impaired population to become more mobile and independent. Several Honorable Mention awards

were also presented: Jeffrey J. Ratusznik, 16, of Marine City High School, Marine City, MI received an award for his project ''Optimization of a Helicopter Blade Using Piezoelectric Ceramic as a Camber Deformer.'' Nikolas D. Bohne, 16, of Nicolet High School, Glendale, WI was recognized for his project ''Acoustic Resonance Spectroscopy'' and Michelle R. Griffith, 16, Elaine A. Pearson, 17, and Lacey Marae Drouet, 16, all of Summer High School, Summer, WA, were recipients of the Honorable Mention award for their team project ''Ultrasound: Effective Management Tool in Sheep Production.''

FIG. 1. ASA judges Richard Lec (front row left) Peter Lewin (front row center) and Felicia Doggett (front row right) with ASA award recipients.

In addition, all seven award winners (see Fig. 1) received a one-year student membership in ASA. The winners were among 46 projects determined by the ASA judging team to be within the Society's areas and interest. The ASA judging team included Felicia M. Dogget, President, Metropolitan Acoustics, Philadelphia, and Richard M. Lec and Peter A. Lewin (Chief Judge), both of Drexel University, Philadelphia.

PETER A. LEWIN

INCE issues ACTIVE 99/INTER-NOISE 99 invitation to participate

An ''Invitation to Participate'' in the 1999 events organized by the Institute of Noise Control Engineering has been posted on the Internet. ACTIVE 99 is the 1999 International Symposium on Active Control of Sound and Vibration—a symposium being held for the third time. It will take place in Fort Lauderdale, FL on 2–4 December 1999. It will be followed by INTER-NOISE 99, the 1999 International Congress and Exposition on Noise Control Engineering which will be held in the same location on 6–8 December 1999. INTER-NOISE 99 is the 27th in a series of congresses that have been field annually since 1972.

More than 120 abstracts have been received for ACTIVE 99, and more than 420 abstracts have been received for INTER-NOISE 99. The INTER-NOISE 99 Equipment Exposition is almost completely booked.

Information on these two meetings has been posted on the Internet. Interested parties can go to the INCE home page, ⟨http:users.aol.com/inceusa/ince.html⟩, and find the link to the meetings at the top of the page. The web site contains general information, the titles of all of the abstracts received, hotel information, and registration information.

A 16-page flyer is available, and will be sent to those who inquire about the meetings. Contact INCE, P.O. Box 3206 Arlington Branch, Poughkeepsie, NY 12603, USA. Telephone: (914) 462-4006; FAX: (914) 463-0201, E-mail: hq@ince.org.

Technical Committee Reports

Acoustical Oceanography

This past year has been an active and fruitful one for Acoustical Oceanography (AO). Starting with meetings, both the Norfolk and Berlin meetings were well attended, and featured extremely interesting technical programs. In Norfolk, AO featured dual sessions on Monday on Shallow Water Geoacoustic Inversions, co-chaired by Ross Chapman and Alex Tolstoy. Tuesday's fare was two joint sessions with Underwater Acoustics, the first on Acoustics and Ocean Processes, chaired by Stanley Dosso, and the second on Internal Waves, chaired by Ahmad Abawi. In Berlin, one of the biggest ASA meetings in years, the technical program abounded with special sessions. On Tuesday, Ken Foote, David Demer, Chris Feuillade, and David MacLennan co-chaired dual sessions on Innovations in Fish and Plankton Acoustics, a subject of intense interest in Europe that is receiving increasingly more attention in the U.S. These sessions were followed by a full poster session on the subject on Wednesday, indicating just how popular the topic is. On Wednesday, dual sessions on Acoustic Tomography and Thermometry were conducted, chaired by Uwe Send and Jim Lynch. Thursday featured even more dual sessions, these on Acoustic Seafloor Classification, chaired by Larry Mayer and Eric Pouliquen. This is a plethora of special sessions for a relatively small technical committee! Perhaps due to this, at the AO technical committee meeting in Berlin, the attendees voted to hold only one special session in Columbus. That will be the joint AB, UW, and AO session on The Effects of Manmade Noise on Marine Mammals, a topic of vital interest to all three technical committees. We also have a large number (27) of papers for contributed sessions, so that Columbus should be a lively meeting yet again for AO.

Getting to the Columbus meeting, there will be two rather special events. First, Hank Medwin, AO's ''Godfather,'' will be launching a little celebration of AO's 10th anniversary as a technical committee. This will be held at the technical committee meeting and it really should be fun! Second, we will give our Oceanography ''Mini-Tutorial'' concept its first outing there. Rob Pinkel and Lev Ostrovsky have each agreed to give one hour talks on deep water and shallow water internal waves (respectively), and these should be a great kick-off for the series.

Speaking of fairly new initiatives, the Best Student Paper competition has finally arrived as a regular feature for the AO committee. In Norfolk, we had our first competition, and our congratulations go to the three winners: Thomas Berger (SIO), Tracianne Neilsen (ARL/UT), and Gopu Potty (URI). Berlin interrupted this intra-AO competition, due to having a special format adapted to the international meeting, but in Columbus the competition should resume again in its original form.

Getting to our ''usual business,'' things are proceeding quite reasonably. Communication of committee business via e-mail seems to be working quite well; responses to queries, requests for volunteers, etc. all are being addressed by having the AO chair communicate intermittently with the committee via electrons. If AO'ers don't mind some occasional ''Spam'' from the committee chair, this seems to be the most efficient way of dealing with things. (This still doesn't replace our physical meetings, however!)

Much more can be said, addressing this detail and that. However, brevity is also appreciated, and so the chairman will respectfully sign off with the message that AO is alive, well, Y2K compliant, and looking forward to another exciting year in 2000.

JAMES. F. LYNCH
Chair 1998–2001

Animal Bioacoustics

The number of attendees for the Animal Bioacoustics (AB) sessions at the Acoustical Society of America (ASA) meeting in Norfolk, VA, was modest. A half-day session was held on Wednesday afternoon and a full day of sessions was held on Thursday. An important milestone was achieved at the Norfolk meeting with the presentation of the first silver medal awarded to a member of the Animal Bioacoustics group. Whit Au, the chair of the AB technical committee, was presented with the silver medal for his contribution in understanding the dolphin sonar.

The joint meeting of the ASA, European Acoustics Association (EAA), and the German Acoustics DAGA Conference in Berlin was a great success for the ASA and for the Animal Bioacoustics Committee. It was probably the most successful meeting for AB, with five sessions over five days. There were 62 verbal presentations scheduled along with 15 posters. Special sessions were conducted in animal electrophysiological acoustics, physics of animal bioacoustics, evolution in hearing and sound production, and the acoustics of echolocating animals. Phillip Jen from the US and Alexander Supin from Russia organized the animal electrophysiological acoustics session. Roland Aubauer, Whit Au, and Christine Erbe organized the physics of animal bioacoustics session. The evolution in hearing and sound production session was organized by Art Popper, Darlene Ketten, Robert Dooling, and Lee Miller. Finally, Hans Urick Snitzler and Whit Au

organized the acoustics of echolocating animals session. The success of the AB sessions can be attributed in a large part by the large number of European participation. Approximately 30% to 40% of the participants were from Europe.

Although the AB sessions were very successful, the AB technical committee meeting was not. A very small number of people were in attendance. One apparent reason for the low attendance at the AB technical committee may be attributed to the good beer at the very successful social which did not have a clear ending time and continued into the time devoted for the technical committee meeting. The second reason can be attributed to some individuals having difficulties finding the room location of the technical committee meeting. Due to the low attendance at the technical committee meeting, nothing of substantial importance was achieved.

The AB committee under the leadership of Dr. Charles Greene has assembled a subcommittee, including representatives from UW and AO, to consider specific recommendations that ASA should make to Congress concerning the Marine Mammal Protection Act. The committee members include John Potter, Peter Worcester, Chris de Moustier, Sam Marshall, John Richardson, and Ellen Livingston.

WHITLOW W. L. AU
Chair 1997–2000

Architectural Acoustics

1998 was a very busy year in Architectural Acoustics both in terms of papers presented as well as activity in the other areas supported by initiatives from the Technical Council.

The first session of the year was held jointly with the International Congress on Acoustics and the Acoustical Society of America in Seattle, WA, 20–26 June. Since this was a joint session there was participation from many countries around the world.

A special workshop was held on Opera House Acoustics jointly with CHARM, MOCHA and ASA. Also included as part of the workshop was a student design competition in Opera House design. Following is a list of topics that were covered as part of the workshop:

Opera House Acoustics Workshop (joint with CHARM and MOCHA)

- Sound Field for the Audience—13 papers,
- State of the Art—4 papers,
- Sound Field for the Singer and Orchestra—13 papers,
- Opera House Design Case Studies—10 papers,
- Opera House Acoustics Student Design Competition—12 entries from 5 schools, with a total of 15 students involved in the competition.

Listed below are the other sessions held at the Seattle meeting:

- Room Acoustic Measurement Techniques Absorption and Diffusion (poster session)—32 posters,
- Measuring and Modeling Noise Flanking Transmission—6 papers,
- Impact of Trains on Performance Halls—5 papers,
- Case Study of the New Tokyo Performing Arts Center—6 papers,
- Case Study of the New Tokyo International Forum—4 papers and panel discussion,
- Speech Intelligibility—4 papers,
- Case Studies of Performance Spaces—12 papers,
- Performance Spaces (poster session)—20 posters,
- Classroom Acoustics—5 papers.

Special Sessions included

- Vern Knudsen Distinguished Lecture—Concert Hall Research: Balancing Complexity with Practicality—John Bradley, and
- Plenary Lecture—Sound Transmission Through Suspended Ceilings— F. M. Mechel.

There were many sessions that were held jointly with the Noise committee which are listed below:

- Novel Approaches to Noise Control Barriers—13 papers,
- Acoustics of Buildings—6 papers,
- Environmental Noise from Combined Sources—9 papers,
- Low-Frequency Noise—4 papers,
- Progress Report and Discussion on the Continuing Activity on ASA's Role in Noise and Its Control.

A Graduate Student Reception was also held (jointly with MU/SC/P&P/AB).

For those in Architectural Acoustics there was a Technical Tour of special interest: the tour of Benaroya Hall (which was nearing completion at that time).

The Seattle session proved to be excellent with almost too many choices of what paper/poster sessions to take in. There were a total of 134 papers/posters presented in the Architectural Acoustics area.

The Fall meeting of the ASA was held in Norfolk, VA, 12–16 October. The Architectural Acoustics group had a number of sessions as well as a combined tour and session in Acoustics for Organ Spaces which was one of the highlights of the meeting.

The following is a list of sessions that were presented at the Norfolk meeting:

- Reverberation Topics—3 papers,
- Recording Microphones—2 papers and panel discussion,
- Acoustics for Office Spaces—8 papers,
- Acoustics for Organ Performance Spaces—18 papers,
- Characteristics of Scattering—6 papers.

Total papers—37 papers.

The Technical Tour mentioned above included organ concerts and demonstrations at two Norfolk churches which were held in conjunction with the session on Acoustics of Organ Performance Spaces.

At the conclusion of the Norfolk meeting there was a Short Course held on Auralization.

Technical Initiatives that were sponsored by the Technical Committee included

- travel funding assistance for students involved in the design comptition at the Seattle meeting;
- funding for translation for some sessions given in Japanese at the Seattle meeting;
- the development of materials for a course in Classroom Acoustics to be used to help educate those outside of the field of acoustics of the need for and proper acoustical design of classrooms;
- funding to place the data that had been collected as part of the Concert Hall Research Group measurements onto a CD-ROM;
- part of the technical initiative was used to help sponsor the Graduate Student Reception in Seattle;
- funding for the organ concerts and demonstrations that were part of the Acoustics of Organ Performance Spaces at the Norfolk meeting;
- funding was provided, but was not used for the development of the Architectural Acoustics section of the ASA web site.

Of special interest in 1998 was the progress made in the area of Classroom Acoustics both from within the ASA and teaming up with other groups outside of the ASA. After a Request for Information on Acoustics was received from the Office of Technical and Information Services from the Architectural and Transportation Barriers Compliance Board, a response was prepared by the ASA and submitted back to the Compliance Board.

The activity in Architectural Acoustics continues to be very strong. I would like to thank the presenters, chairs and co-chairs, organizers, TCAA members, and all those who have participated in 1998 for their hard work and effort to make it a successful year. I would also encourage any of those interested in Architectural Acoustics to attend the ASA meetings and become involved in the Architectural Acoustics group. The TCAA meetings are an excellent way to get started.

RON FREIHEIT
Chair 1998–2001

Engineering Acoustics

The Engineering Acoustics Technical Committee (EATC) met two times since last June; at Norfolk, VA, in October 1998 and at Berlin in March 1999. The meeting in Norfolk had 21 members present while the meeting in Berlin had 15.

EATC paper sorting for the Norfolk meeting was done by James Tressler. The meeting had 40 papers in four EATC sessions including two special sessions. The first special session was organized by P. K. Raju and was titled ''Nondestructive Evaluation of Materials using Acoustical Techniques.'' This session included three invited papers and 11 contributed speakers. The second special session, titled ''In-air Actuators,'' was organized by Thomas Howarth and included five invited presentations and two contributed papers. A student paper contest was organized and conducted by Hal Robinson.

The paper sorting for the Berlin Meeting was done by Stan Ehrlich and Roger Richards. EATC had a total of 93 papers in ten EATC sessions including six special sessions. The first special session was organized by Neil Shaw and Diemer de Vries and was titled ''3D Loudspeaker Reproduction Methods.'' This session included three invited papers and nine contributed speakers. Our second special session, ''Silicon Sensors,'' was organized by Ilene Busch-Vishniac and Gerhard Sessler and included five invited presentations and six contributed papers. The third special session was organized by Dehua Huang, Bill Marshall, and Lars Kofoed. This session was titled ''Recent Advances in Underwater Transducers I,'' and included six invited papers and seven contributed. The EATC fourth special session was organized by Steve Thompson, Wolfgang Doering, and James Powers. This session was titled ''Digital Signal Processing for Hearing Aids I'' and was cosponsored with the Interdisciplinary Technical Group on Signal Processing in Acoustics. There were six invited papers and six contributed papers for this session. The fifth special session, ''Implications of Recent Standards,'' was organized by Victor Nedzelnitsky and Dieter Gottlob and cosponsored with the Technical Committee on Noise. This session had five invited papers and one contributed presentation. The sixth and final EATC special session was ''SAW Devices'' and was organized by Hal Robinson and Daniel Hauden. This session included three invited papers and two contributed presentations.

Our Berlin TC Meeting revealed that there were nine EATC-based student papers competing in the overall Meeting competition. Hal Robinson continued as the EATC Student Paper award representative. We also started a new outreach committee as a liaison to Signal Processing with Stan Ehrlich as the inaugural EATC representative.

THOMAS R. HOWARTH
Chair 1997–2000

Biomedical Ultrasound/Bioresponse to Vibration

The 1998–1999 year was a good one for the B&B technical committee, a year in which our visibility and membership grew. The Norfolk, VA, meeting (TPOM rep. Doug Mast) saw a well-attended special session on ''Models for tissue-ultrasound interaction'' co-chaired by Doug Mast and Carr Everbach. The technical committee meeting at Norfolk engaged topics such as the rising cost of meetings (especially their included social functions and audiovisual equipment rentals), and preregistration for conferences. But the main B&B effort was the joint ASA/EAA/DEGA meeting in Berlin, Germany, in March 1999. Five B&B sessions, the last three of which were special sessions, were generally well attended: HIFU and Thermal Effects (Ron Roy and Ibrahim Hallaj, co-chairs), Modeling and Imaging (Carr Everbach and Doug Mast, co-chairs), Lithotripsy (Robin Cleveland and Michael Delius, co-chairs), Safety Guidelines for Diagnostic Ultrasound (Francis Duck and Wesley Nyborg, co-chairs), and Low Intensity/Low Frequency Ultrasound (Carr Everbach and Gail ter Haar, co-chairs). In addition, B&B member and founder Henning von Gierke received a gold medal for his lifetime of accomplishments. At the technical committee meeting in Berlin, which boasted record attendance, it was generally agreed that B&B special sessions should be encouraged in the Spring meeting each year and not the Fall. The rationale for this change in emphasis (which would begin following the Columbus meeting) is to allow those B&B members with other professional society duties a stronger reason to attend at least one ASA meeting per year. There would still be contributed paper sessions at the Fall ASA meetings. For the Columbus, OH, meeting in November 1999, three B&B special sessions are expected: ''Signal Processing in Biomedical Ultrasound'' (joint with SP)—Shira Broschat and Hua Lee, organizers, ''Therapeutic Applications of Ultrasound''—Inder Makin, organizer, and ''Modeling of Nonlinear Ultrasonic Fields'' (joint with PA)—Robin Cleveland, organizer.

For the Atlanta meeting, five special sessions are currently under development. Finally, at the Berlin ASA meeting, the baton was passed from Ron Roy to Carr Everbach, whose term as chair of the B&B will span through 2002. Contact him at (610) 328-8079 or via e-mail at ceverba1@swarthmore.edu with comments or questions.

E. CARR EVERBACH
Chair 1999–2002

Musical Acoustics

The 1998–1999 year for the Technical Committee on Musical Acoustics (TCMU) has included national and international meetings. The ASA meeting in Norfolk in October 1998 formed the occasion in which the ASA Silver Medal in Musical Acoustics was presented to Neville Fletcher. His contributions were cited in understanding sound production, and especially the role of nonlinear processes, in strings, wind, and percussion musical instruments. The Norfolk meeting was reminiscent of older ASA meetings, modest in scope and intimate in style, providing opportunities for extensive discussions among colleagues. A contributed session at Norfolk chaired by Doug Keefe included papers on a wide range of topics in musical acoustics. Musical acoustics cosponsored a special session on acoustics in multimedia, which was organized by David Havelock on behalf of the Interdisciplinary Technical Group on Signal Processing in Acoustics.

The joint meeting of the 2nd Convention of the European Acoustics Association and the ASA was held in Berlin, Germany, in March 1999. Not to be outdone by recent international meetings of musical acousticians, this Forum Acusticum 99 included more than 100 presentations on musical acoustics. Each special session was jointly organized by EAA and ASA co-chairs. Murray Campbell and Bill Strong organized a session on sound production of wind instruments. This was followed by a session on modeling versus measurements of wind instruments, organized by Mico Hirschberg and Shigeru Yoshikawa. The emphasis in a session organized by Anders Askenfelt and Tom Rossing was on music acoustics and the musician. A paired set of special sessions followed on musical instruments and structural acoustics. The first session chaired by Antoine Chaigne and Uwe Hansen featured presentations on experimental studies, theoretical models, and numerical analysis, and the second session chaired by Ingolf Bork and Isao Nakamura concerned the piano and related instruments. A named special session in honor of the contributions of Reinier Plomp was jointly sponsored by Psychological and Physiological Acoustics, Speech Communication, and Musical Acoustics (co-chairs Tammo Houtgast and Bob Shannon). René Caussé and Johan Sundberg organized a special session on the quality of musical instruments and human voice. Exploring research in subjective musical acoustics, Roger Kendall and Stephen McAdams organized a session on mapping multiple physical and perceptual attributes to musical structures. A final contributed abstracts session featured both lectures and posters, and was chaired by James Beauchamp and Helmut Fleischer. In something of an innovation in musical acoustics, Forum Acusticum 99 was the second meeting (with ISMA 98) in a year that included poster presentations. Peter Hoekje was acting Chair of the Berlin meeting of the TCMU, and the evening concluded with a joint gathering of the DEGA colleagues and TCMU members to discuss collaborative projects and future musical acoustics meetings.

Efforts to increase student involvement in musical acoustics have continued through best student presentation contests and joint sponsorship of student socials. The winner of the Norfolk contest for Best Student Paper in Musical Acoustics was Keith D. Martin (MIT), who presented with Youngmoo E. Kim on ''Musical instrument identification: A pattern-recognition approach.'' Other education outreach this year included a technical initiative in support of educational workshops on acoustics for teachers conducted by Uwe Hansen. There were three entrants in the best student paper in musical acoustics competition in Norfolk and ten entrants (from musical acoustics) in the best student paper competition in Berlin.

The representatives to the Technical Program Organizing Meetings were Charles Schmid for the Norfolk meeting, and Jürgen Meyer and Doug Keefe for the Berlin meeting. Representation from the TCMU on other ASA committees continues with Uwe Hansen on Membership Committee, Roger Hanson on Medals and Awards Committee, and Ian Lindevald on the ASA-COS Standards Committee; *ex officio* members on the TCMU include JASA Associate Editors Bill Strong and Stan Ehrlich. New members on the TCMU include George A. Bissinger, Annabel J. Cohen, Diana Deutsch, Uwe J. Hansen, Roger J. Hanson, Peter L. Hoekje, Douglas H. Keefe, Ian M. Lindevald, Stephen E. McAdams, Gary P. Scavone, Marc-Pierre Verge, and Gabriel Weinreich. The web site for the TCMU maintains current information on musical acoustics at http://www.public.coe.edu/~jcotting/tcmu.

This 1998–1999 year concluded with the election of James P. Cottingham as Chair of the Technical Committee on Musical Acoustics.

DOUGLAS H. KEEFE
Chair 1996–1999

Noise

The TC Noise met two times since last summer: October in Norfolk, VA, and in March in Berlin. The Berlin TC meeting had significantly fewer attendees and proportionally fewer regular attendees.

In Norfolk we discussed the amorphous committee structure and the purpose of the committee. It was a clear consensus that the group was working well without structure or clear purpose and would not be changed soon. Nevertheless, with no officers, no formal committee structure, and no clear purpose, the Chairperson believes the Committee will need to be more formalized at some time in the future.

The committee has worked hard at organizing sessions for the ASA meetings. The Berlin ASA/EAA meeting took a considerable amount of volunteer time. Probably the most visible item to discuss is the efforts several members made on the classroom acoustics task force. Funding, measurement standards, and criteria are all developing from this effort.

We now distribute all minutes by E-mail only. This seems efficient and definitely a cost savings.

The qualifications for Committee ''membership,'' if any remain, are too vague. As a result, the membership list grows because of the reluctance to drop long-time, active, members. It appears that people participate on the Committee independent of their membership status. That is, the conferring of ''member'' status does not assure attendance at the TC Noise meetings.

We have had help from many people, from session organizers and program planners, to invited speakers, to people bringing refreshments, to workshop organizers, and to contributors to the lively discussions at the meeting. We continue to welcome new people to our open meetings and solicit their participation.

RICHARD J. PEPPIN
Chair 1997–2000

Psychological and Physiological Acoustics

The Norfolk and Berlin ASA meetings presented quite a contrast for Psychological and Physiological Acoustics (P&P). The technical program at Norfolk was small, consistent with P&P's deemphasis of the fall meeting. There was reasonable attendance at the sessions, however, and opportunities to explore other technical areas. The highlight of the meeting was the presentation of the von Békésy Silver Medal to Murray B. Sachs. Our congratulations to Murray for this well-deserved honor, and our thanks to Eric Young for his opening remarks at the ceremony. In a surprise move, Robert Bilger announced his retirement while chairing one of the sessions. Although an exception to avoiding special sessions at fall meetings, a session honoring Bob has been organized for Columbus by Larry Feth (who is also Technical Program Chair) and Walt Jesteadt. At Norfolk, the discussion at the evening P&P meeting focused on costs associated with meetings. P&P continues to advocate for lower registration costs and author-dictated A/V support (including computer projection), giving up some amenities if necessary. There was support for replacing one social with refreshments at the Technical Committee meetings to help encourage attendance. Support continued for one ASA meeting per year, although a formal vote of the Technical Council confirmed that this remains a minority position. During Associate Editors' reports, Joseph Hall was thanked for his three years of excellent work, and Sid Bacon was welcomed as a new Associate Editor, beginning January 1999. With Peggy Nelson as our representative, P&P continues to participate in the efforts to improve the acoustics of classrooms. A standard is being prepared, and there is optimism for federal legislation.

The spring meeting in Berlin was unique in many ways. It was large even by P&P standards for spring meetings, with 35% more abstracts than Seattle last year. P&P has rarely scheduled so many special sessions, but this proved successful as a strategy to guarantee a strong core to the program. There was a good balance of ASA and EAA contributions. P&P was the primary sponsor for six sessions, thanks to the hard work of the organizers: Torsten Dau, Bill Yost, Mary Florentine, Brian Moore, Ruth Litovsky, Steve Colburn, Tammo Houtgast, Bob Shannon, John Culling, Gerald Kidd, Hugo Fastl, Manfred Kalivoda, and Rich Peppin. P&P also cosponsored three additional sessions. Two eminent European members of P&P were honored at Berlin. The many outstanding contributions of Reinier Plomp were honored at a special session co-sponsored by P&P, Speech Communication, and Musical Acoustics. Another high point was the presentation of the Helmholtz–Rayleigh Interdisciplinary Silver Medal to Jens P. Blauert. Congratulations to Dr. Blauert for this recognition of the breadth and importance of his work, and our thanks to Constantine Trahoitis for his introductory remarks honoring Dr. Blauert. Although the university venue presented some problems, those who attended the meeting seemed enthusiastic about the strong technical program and opportunities for interactions between ASA and EAA colleagues. Several sessions addressed topics that are not routinely part of our P&P programs, such as sound quality and auditory display. Twelve P&P students entered the all-ASA student-paper contest at Berlin, requiring seven (anonymous) P&P judges recruited by Mary Florentine. Our students also enjoyed a crowded student reception, held jointly by several technical areas and funded with Technical Initiative money. Attendance at the evening P&P meeting was small, but those present elected the six new members of our Technical Committee for next year: Barbara Shinn-Cunningham, Virginia Richards, Patrick Zurek, Pierre Divenyi, Richard Freyman, and Wesley Grantham. My thanks to the outgoing Committee members, who prepared the ballot: Leslie Bernstein (Election Chair), Amy Donahue, Gerald Kidd, Ruth Litovsky, Dennis McFadden, and Robert Shannon. Discussion of A/V policy and costs continued. Those present had little objection to transparencies as the default for paper presentation. For computer projection or other special needs, the majority favored sharing these costs across all attendees rather than charging individual authors. Robert Frisina provided a report on fundraising by the new ASA Foundation. With targeted giving, the Foundation might provide a vehicle for P&P-specific initiatives. More generally, the Technical Council has funding for innovative projects. P&P members are encouraged to pursue ideas with the Chair. In the second part of the meeting, we met jointly with EAA colleagues to discuss topics of mutual interest.

In general news, this has been a good year for P&P representation in ASA. William Hartmann served as Vice President, and he will be a candidate for President next spring—be sure to vote! Congratulations also to William Yost, who was elected to Executive Council this spring. Bill's replacement on the Medals and Awards Committee is Marjorie Leek. The mass e-mail list for P&P developed last year was used several times this year with apparent success. If you did not receive these mailings and wish to, please send your e-mail address to neff@boystown.org. Finally, my term as Chair ended with Berlin, and Neal Viemeister has been elected as the new Chair. Personally and on behalf of P&P, I thank all those who worked on special sessions, projects, various committees, or otherwise gave generously of their time. I particularly thank Dennis McFadden, Bill Yost, and the 30 members of the P&P Committee these past three years, who were always quick with good advice and help.

DONNA NEFF
Chair 1997–1999

Signal Processing in Acoustics

Signal and image processing are the underlying technologies that encompass many of the efforts in acoustics through the necessity to model, acquire, analyze, and extract the required information from noisy measurement data. It is for these reasons that the Interdisciplinary Technical Group on Signal Processing (ITG-SP) has and continues to foster a growing interaction with each of the Technical Committees (TC) through the sponsoring of joint sessions and tutorials. This effort was made by the various signal-processing-related sessions throughout the Norfolk and Berlin meetings.

At the 136th ASA meeting in Norfolk the ITG-SP sponsored or cosponsored five sessions: two contributed sessions and three special sessions. These sessions were entitled

- ''Time-Frequency Techniques is Acoustic Applications I and II,'' Chair: J. V. Candy,
- ''Acoustics in Multimedia-Content Issues,'' (MU), Chair: D. I. Havelock,
- ''Performance of Active Noise Control Systems in Real-World Applications,'' (SA, EA) Co-chairs: A. J. Brammer and S. C. Thompson, and
- ''General Topics in Signal Processing in Acoustics,'' Chair: K. B. Fisher.

Also the ''Gallery of Acoustics,'' sponsored by the ITG-SP and compiled, set up, and displayed by Randy Smith, was one of the highlights of the meeting. There will be galleries set up at the Columbus and Atlanta meetings. The Gallery provides ''creative'' acousticians the opportunity to display the results of their efforts in a more creative light.

The 137th ASA meeting in Berlin, Germany, featured more joint ITG-SP sponsored/cosponsored sessions with 2 contributed and 13 special sessions. These sessions were

- ''Auditory Displays,'' (PP) Chair: E. Wenzel, Co-chair: J. Blauert,
- ''Acoustic Image Synthesis for Arbitrary Listeners,'' (SC) Co-chairs: A. Brammer and A W. Bronkhorst,
- ''Smart Microphones and Audio for Teleconferencing,'' (EA) Co-chairs: G. W. Elko and J. Blauert,

- ''Signal Processing,'' Chair: J. Impagliazzo,
- ''Localization, Estimation, and Detection in Underwater Acoustics,'' (UW) Chair: D. E. Smith,
- ''Underwater Communication and Signal Processing,'' (UW) Co-chairs: C. Jones and A. Maguer,
- ''Binaural Technology I & II,'' (EA) Co-chairs: M. Burkhart and J. Blauert,
- ''Verification of Auralization and Modeling Programs I & II,'' (AA) Co-chairs: R. H. Campbell and U. P. Svensson,
- ''Digital Signal Processing for Hearing Aids I & II,'' (EA) Co-chairs: S. Thompson and W. H. Doering,
- ''Acoustic Signal Processing for Systems with 2D/3D Sensor Arrays I & II,'' Co-chairs: S. Stergiopoulos and D. Kraus,
- ''Acoustics in Multimedia,'' Co-chairs: H. Moller and D. I. Havelock, and
- ''General Signal Processing in Acoustics,'' Chair: K. Benjamin.

In a concerted effort to enhance the communications between the ITG-SP and the other Technical Committees, Technical Liaisons (TL) were created for each TC. A Technical Liaison is a member of the ITG-SP who represents the interests of the group to other Technical Committees (e.g., Underwater Acoustics). The TL would attend the ITG-SP meeting as well as his particular Technical Committee. He would primarily facilitate the CO-SPONSORING of joint sessions with ITG-SP (our charter) and the Committee, assuring no technical overlap with other sessions. The following ITG-SP TL were assigned:

Acoustical Oceanography	Open
Animal Bioacoustics	Brian Ferguson
Architectural Acoustics	Open
Biomedical Ultrasound/ Bioresponse to Vibration	Shira Broschat/Stergios Stergiopoulos
Engineering Acoustics	Stan Ehrlich
Musical Acoustics	John Impagliazzo
Noise	Dave Evans/Joe Pope
Physical Acoustics	Dave Havelock/Kent Lewis
Psychological and Physiological Acoustics	Open
Speech Communication	Open
Structural Acoustics and Vibration	Stuart Bolton
Underwater Acoustics	Eliza Michalopoulou

We also would like to recognize our JASA Associate Editor for Acoustical Signal Processing, John C. Burgess, for performing a wonderful job in removing the serious backlog of papers he inherited in a timely and efficient manner. This effort is often unnoticed until a crisis (such as this) appears and somehow magically disappears. We are attempting to create a list of interests of our members to help in this review process.

We would also like to recognize Marti Bartlet for creating, updating, and maintaining our website, another task that requires much effort but yields little return. Surf to our site and get all of the current tidbits of information and links to other processing websites. Our address is http://www.arlut.utexas.edu/asaspweb/

JAMES V. CANDY
Chair 1997–2000

Speech Communication

Productivity for the Speech Technical Committee has been high this year with a number of individuals providing substantial contributions. In the following paragraphs, a brief account will summarize the activities of the Committee and describe key issues which are being considered by the Committee. Feedback on any issue is welcomed and may be addressed to any member of the Committee. Current members of the Committee are Frederica Bell-Berti, Lynne Bernstein, Timothy Bunnell, Carol Espy-Wilson, Marios Fourakis, Sadaoki Furui, Kenneth Grant, John Hawks, William Katz, Patricia Keating, Jody Kreiman, Patricia Kuhl, Paul Luce, Richard McGowan, Susan Nittrouer, Douglas O'Shaughnessy, Robert Port, Deborah Rekart, Astrid Schmidt-Nielsen, Juergen Schroeter, Mitchell Sommers, Janet Stack, Winifred Strange, Mario Svirsky, Ann Syrdal, Rosalie Uchanski, Gary Weismer, and Doug Whalen. The *ex officio* members of the Committee are James Hieronymus, Anders Lofqvist, James Hillenbrand, Winifred Strange, Maureen Stone, and Diane Kewley-Port. If you are interested in becoming a member of the Committee, please contact Emily Tobey.

Web Page Improvements

Bob Port has done a tremendous job developing the web page for the Speech Communication Committee ⟨asa.aip.org/committees.html⟩. A new feature added this year is *Dennis Klatt's History of Speech Synthesis* with audio clips taken from his 1987 JASA article. The audio samples illustrate early speech synthesis of interest to students and professionals. The page also describes the mission and scope of our division within the Society, lists interesting web sites related to speech communication, provides a listing of upcoming meetings related to speech communication, and describes student activities sponsored by our Committee. Please contact Bob Port if you have ideas regarding new additions to the page.

Student Activities

The Committee sponsors two activities, an evening reception and competition with a cash award for best oral or poster presentation, geared especially for students in Speech Communication at each ASA meeting. Attendance at the Norfolk meeting student reception was excellent with approximately 30 students participating in the event. Assistance in organizing and executing the reception was given by Marios Fourakis and John Hawks. The reception for students at the Berlin meeting was ''standing-room'' only with over 50 attendees from Speech Communication, Psychological and Physiological Acoustics, Animal Bioacoustics, and Musical Acoustics.

Susan Nittrouer has headed a subcommittee whose responsibility is to judge student presentations at the meetings. Members of the Speech Communication community serve as judges to independently rate the presentations over the course of the meeting. The outstanding student presentation in Norfolk went to Alexander Francis from the University of Chicago for a paper entitled ''Perceptual learning of synthetic speech.''

The Education in Acoustics Committee sponsored a Society-wide student competition at the Berlin meeting. One hundred and nineteen papers were submitted for consideration of the student award. Fourteen of the student presentations were from Speech Communication. Opinions regarding a Society-wide versus Speech Communication-only award varied. Most of our members at the Berlin meeting preferred a Speech Communication-only competition so we will continue to sponsor this event at future meetings.

Editors

Allan Pierce of Boston University has been selected to serve as the new editor-in-chief of our Journal. James Hillenbrand and Anders Lofqvist, Associate Editors, will be relinquishing their editorships in the Journal over the summer months. Both editors have done a fine job and new appointments will be named by Editor Pierce.

Raymond H. Stetson Scholarship in Phonetics and Speech Production

Through donations made by Arthur Benton and Mac Pickett, a scholarship has been established in honor of Raymond Stetson. Members who are interested in contributing to this fund should send their donations to the Acoustical Society Foundation and specify the Stetson Scholarship. Students who are interested in motor speech production (motor coordination of speech organs, kinematics of speech articulation, aerodynamics of speech, or combined acoustic/kinematic modeling of speech), phonetics (linguistic issues related to speech, articulatory phonology, or experimental phonology), or bio-developmental factors in speech (neural function or infant/child development of speech communication) are invited to apply. Patricia Keating chaired a subcommittee composed of Abeer Alwan, Mark Hasegawa-Johnson, and Peter Ladefoged to review the selection criteria and mechanics of selection. The subcommittee has revised the application form and streamlined the selection procedures. Students who applied for the initial competition of the award expressed research interests in a variety of topics including electromyography of voice disorders, speech control in Parkinson's Disease, an aerodynamic model of speech production, development of orofacial control for speech, the aging larynx, three-dimensional vocal fold models, velopharyngeal inadequacy, coordination in stuttering, development of motor coordination, foreign accents, relation of fundamental frequency to oral pressure, pitch accent, lexicon and development of speech fluency, and infant speech perception. The person selected to receive the 1999 Stetson Scholarship will be announced at the Fall ASA meeting.

Special Initiatives

At the Seattle meeting, our committee sponsored a special session entitled ''Speech Communication: A Half Century of Speech Research.'' The session hosted invited papers by Gunnar Fant, Ken Stevens, Jim Flanagan, Al Liberman, L. Chistovich, Mac Pickett, Bob Porter, Katherine Harris, Peter Ladefoged, Vicki Fromkin, and Hiroya Fujisaki. Video tapes were made of the presentations. This year, through a special ASA technical initiative, the University of Washington will work at making a composite tape available for members and explore the possibility of placing some of the samples on our Committee's web page. Any member who has an idea for a special ASA technical initiative in Speech Communication should contact one of the Committee members.

European Acoustics Association

Four members of the European Acoustics Association attended our Speech Technical Committee meeting. They were Bernd Moebius from the Univeristy of Stuttgart, Wolfgang Hess from the University of Bonn, Birge Kollmeier from the University of Oldenburg, and Klaus Fellbaum from the Technical University of Berlin. The members expressed interest in the activities of our group and interest in communicating further with interested individuals. If you are interested in further contact, they may be reached at Moebius (moebius@ims.uni-stuttgart.de), Hess (wgh@ikp.uni-bonn.de), Kollmeir (biko@medi.physik.uni-oldenburg.de), and Fellbaum (fellbaum@kt.tu-cottbus.de).

Fund Raising Efforts

As subscriptions for the Journal decrease, financial resources for new initiatives also decrease. In order to continue providing a financial avenue for Speech Communication to develop new ideas, the Committee has initiated discussions exploring avenues of income generation. Several ideas are being considered and we would welcome more. The Committee is exploring the possibilities of cosponsoring our current student events with commercial companies and initiating an active campaign to increase the Klatt and Stetson funds. Currently, funds from the teaching video tape produced by Maureen Stone and featuring several of our members are being channeled to the Klatt fund. More activities or teaching tools could be developed in the future with profits being moved to these accounts.

Medals, Awards, and Fellows

Winifred Strange is the Speech Communication representative to the Society's Medals and Awards Committee and Maureen Stone is our representative to the Membership Committee which makes decisions regarding Fellows of the Society. Please contact them if you wish to nominate an individual for either honor.

EMILY TOBEY
Chair 1998–2001

Underwater Acoustics

The Norfolk meeting featured special sessions on ''Acoustic Detection and Imaging of Buried Objects,'' chaired by Nicholas P. Chotiros, and ''Acoustics and Ocean Processes,'' chaired by Stanley E. Dosso and jointly sponsored by Acoustical Oceanography. Student activities included a graduate student reception, held jointly with Acoustical Oceanography, and a Best Student Paper Award competition. The winners were Sunny R. Khosla (University of Michigan)—first prize; Brian J. Sperry (Massachusetts Institute of Technology/Woods Hole Oceanographic Institution Joint Program)—second prize; and Thomas R. Hahn (Scripps Institution of Oceanography)—third prize. At the UW Technical Committee meeting, potential areas for expansion and growth were discussed and included oil exploration, underwater archeology, gas hydrates, and acoustics of thick suspensions. Mechanisms for expansion, such as joint meetings and invited speakers, were also discussed.

The highlight of the Berlin meeting was a special session on wave propagation held in honor of Russian academician Leonid M. Brekhovskikh and chaired by Peter N. Mikhalevsky and Nikolai Dubrovsky. American, European, and Asian colleagues paid tribute to his extraordinary scientific and personal accomplishments. The culmination of the event was the award of Honorary Fellowship to Brekhovskikh and his moving acceptance speech. It was truly a unique moment in the history of underwater acoustics.

There were several other interesting and timely special sessions in Berlin: ''Rapid Ocean Environmental Assessment,'' chaired by J. Sellschopp and R. A. Wagstaff, ''Scattering from Finite Objects Near Boundaries,'' chaired by J. A. Fawcett and H. G. Schneider; ''Ultra Low Frequency (<1 Hz) Ocean Acoustics,'' chaired by J.-P. Montagner and R. A. Stephen and jointly sponsored by Acoustical Oceanography; and ''Acoustics of the Comprehensive Test Ban Treaty,'' chaired by A. B. Baggeroer and M. Lawrence and cosponsored by Acoustical Oceanography and Physical Acoustics. The Berlin meeting also featured a lively and well-attended graduate student reception, held jointly with Acoustical Oceanography.

GEORGE V. FRISK
Chair 1997–2000

USA Meetings Calendar

Listed below is a summary of meetings related to acoustics to be held in the U.S. in the near future. The month/year notation refers to the issue in which a complete meeting announcement appeared.

1999

30 Sept.–2 Oct.	Seventh Annual Conference on the Management of the Tinnitus Patient, Iowa City, IA [Rich Tyler, Dept. of Otolaryngology, Head and Neck Surgery, The University of Iowa, 200 Hawkins Dr., #E230 GH, Iowa City, IA 52242-1078; Tel.: 319-356-2471; Fax: 319-353-6739; E-mail: rich-tyler@uiowa.edu, WWW: http://www.medicine.uiowa.edu/otolaryngology/news/news.html].
7–10 Oct.	Symposium on Occupational Hearing Loss, Philadelphia, PA [American Institute for Voice and Ear Research, Attn: Barbara-Ruth Roberts, 1721 Pine St., Philadelphia, PA 19103; Tel.: 215-545-2068; Fax: 215-735-2725].
1–5 Nov.	138th meeting of the Acoustical Society of America, Columbus, OH [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].
2–4 Dec.	ACTIVE 99, Fort Lauderdale, FL [Institute of Noise Control Engineering, P.O. Box 3206 Arlington Branch, Poughkeepsie, NY 12603; Tel.: 914-462-4006; Fax: 914-463-0201; E-mail: INCEUSA@aol.com;/users.aol.com/inceusa/ince.html].
6–8 Dec.	INTER-NOISE 99, Fort Lauderdale, FL [Institute of Noise Control Engineering, P.O. Box 3206 Arlington Branch, Poughkeepsie, NY 12603; Tel.: 914-462-4006; Fax: 914-463-0201; E-mail: INCEUSA@aol.com;/users.aol.com/inceusa/ince.html].

2000

30 May–3 June	139th meeting of the Acoustical Society of America, Atlanta, GA [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].
4–8 Dec.	140th meeting of the Acoustical Society of America, Newport Beach, CA [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].

Members of Technical and Administrative Committees and Technical Groups of the Acoustical Society of America

The Technical and Administrative Committees and Technical Groups listed below have been appointed by the Executive Council. These appointments, with such changes as may be made by the President from time to time, will be in effect until the Spring meeting of the Society in 2000.

Technical Committees 1999–2000

Acoustical Oceanography

	Term to
James F. Lynch, *Chair* to	2001
John A. Colosi	2002
Grant B. Deane	2002
Brian D. Dushaw	2002
David M. Farmer	2002
Alex E. Hay	2002
Kevin D. Heaney	2002
Stephen A. Reynolds	2002
David M. Rubenstein	2002
Holly A. Burch	2001
Marshall Hall	2001
Darrell R. Jackson	2001
Jeffrey A. Nystuen	2001
Lev A. Ostrovsky	2001
Robert Pinkel	2001
Peter F. Worcester	2001
Manell E. Zakharia	2001
Robert W. Farwell	2000
Gary J. Heald	2000
Anatoliy N. Ivakin	2000
Subramanian D. Rajan	2000
Barbara J. Sotorin	2000
Ralph A. Stephen	2000
Alexander G. Voronovich	2000
Suk Wang Yoon	2000

Ex officio:
John C. Burgess, Associate Editor of JASA
David L. Bradley, Associate Editor of JASA
Stanley L. Chin-Bing, Associate Editor of JASA
Michael J. Buckingham, member of Medals and Awards Committee
Christopher Feuillade, member of Membership Committee
Er-Chang Shang, member of ASACOS

Animal Bioacoustics

	Term to
Whitlow W. L. Au, *Chair* to	2000
William C. Burgess	2002
James J. Finneran	2002
Adam S. Frankel	2002
Darlene R. Ketten	2002
Larry L. Pater	2002
Peter L. Tyack	2002
Robert Hickling	2001
David A. Helweg	2001
Martin L. Lenhardt	2001
Peter M. Narins	2001
John R. Potter	2001
Daniel R. Raichel	2001
Robert H. Benson	2000
Ann E. Bowles	2000
William C. Cummings	2000
Charles R. Greene	2000
Mardi C. Hastings	2000
D. Vance Holliday	2000
David K. Mellinger	2000
Sam H. Ridgway	2000

Ex officio:
Whitlow W. L. Au, Associate Editor of JASA
David L. Bradley, Associate Editor of JASA
Stanley A. Chin-Bing, Associate Editor of JASA
Mardi C. Hastings, member of Medals and Awards Committee
Sam H. Ridgway, member of Membership Committee
Ann E. Bowles, member of ASACOS

Architectural Acoustics

	Term to
Ronald R. Freiheit, *Chair* to	2001
Christopher N. Blair	2002
John S. Bradley	2002
Christopher N. Brooks	2002
Angelo J. Campanella	2002
Robert C. Coffeen	2002
Peter D'Antonio	2002
Timothy J. Foulkes	2002
J. Christopher Jaffe	2002
Mendel Kleiner	2002
Edward L. Logsdon	2002
David Lubman	2002
Michael T. Nixon	2002
Jack E. Randorff	2002
H. Stanley Roller	2002
Noral D. Stewart	2002
Bennett M. Brooks	2001
Steven M. Brown	2001
Richard H. Campbell	2001
M. David Egan	2001
Richard M. Guernsey	2001
Mark A. Holden	2001
K. Anthony Hoover	2001
Dana S. Hougland	2001
David W. Kahn	2001
Gary S. Madaras	2001
Charles T. Moritz	2001
Paul B. Ostergaard	2001
Dennis A. Paoletti	2001
David J. Prince	2001
Neil A. Shaw	2001
Richard H. Talaske	2001
Ewart A. Wetherill	2001
George E. Winzer	2001
Michael R. Yantis	2001
Yoichi Ando	2000
David Braslau	2000
Dennis Fleisher	2000
John W. Kopec	2000
Robert F. Mahoney	2000
Richard J. Peppin	2000
Scott Pfeiffer	2000
Roy L. Richards	2000
Ludwig W. Sepmeyer	2000
Gary W. Siebein	2000
David Still	2000
Lily Wang	2000

Ex officio:
Courtney B. Burroughs, Associate Editor of JASA
J. David Quirt, Associate Editor of JASA
Steven M. Brown, member of Medals and Awards Committee
Gregory C. Tocci, member of Membership Committee
George E. Winzer, member of ASACOS

Biomedical Ultrasound/Bioresponse to Vibration

	Term to
E. Carr Everbach, *Chair* to 2002	2002
Robert E. Apfel	2002
Michalakis A. Averkiou	2002
Michael R. Bailey	2002
Andrew J. Coleman	2002
Floyd Dunn	2002
John Erdreich	2002
Christy K. Holland	2002
Wesley L. Nyborg	2002
Joseph E. Piercy	2002
K. Kirk Shung	2002
Mark P. Verge	2002
Paul E. Barbone	2001
Anthony J. Brammer	2001
Diane Dalecki	2001
J. Brian Fowlkes	2001
Leon Frizzell	2001
Alan K. Goble	2001
Mark Hollins	2001
Kullervo Hynynen	2001
T. Douglas Mast	2001
Doug Miller	2001
Pierre Mourad	2001
William D. O'Brien	2001
Stanley L. Bolanowski, Jr.	2000
Shira L. Broschat	2000
Robin O. Cleveland	2000
Inder R. S. Makin	2000
Janet M. Weisenberger	2000
Junru Wu	2000

Ex officio:
Floyd Dunn, Associate Editor of JASA
Sid P. Bacon, Associate Editor of JASA
Robert V. Shannon, Associate Editor of JASA
Wesley L. Nyborg, member of the Medals and Awards Committee
Anthony J. Brammer, member of the Membership Committee
Robin O. Cleveland, member of ASACOS

Engineering Acoustics

	Term to
Thomas R. Howarth, *Chair* to 2000	2000
Steven R. Baker	2002
David A. Brown	2002
Stephen C. Butler	2002
Robert D. Corsaro	2002
Stephen E. Forsythe	2002
Brian H. Houston	2002
W. Jack Hughes	2002
Stephen C. Thompson	2002
Arnie L. Van Buren	2002
Mahlon D. Burkhard	2001
James Christoff	2001
Dennis F. Jones	2001
Jan F. Lindberg	2001
Yushieh Ma	2001
Elizabeth A. McLaughlin	2001
Alan Powell	2001
Roger T. Richards	2001
Harold C. Robinson	2001
Kenneth D. Rolt	2001
Neil A. Shaw	2001
James Tressler	2001
Kim C. Benjamin	2000
Allan C. Cummings	2000
Gary W. Elko	2000
Robert D. Finch	2000
Guillermo C. Gaunaurd	2000
Gordon Hayward	2000
Dehua Huang	2000
Sung Hwan Ko	2000
Victor Nedzelnitsky	2000
Ahmet Selamet	2000
James E. West	2000
Oscar B. Wilson	2000
George S. K. Wong	2000

Ex officio:
Henry E. Bass, Associate Editor of JASA
Stanley L. Ehrlich, Associate Editor of JASA
Robert D. Finch, member of Medals and Awards Committee
Thomas R. Howarth, member of Membership Committee
Mahlon D. Burkhard, member of ASACOS

Musical Acoustics

	Term to
James P. Cottingham, *Chair* to 2002	2002
George A. Bissinger	2002
Annabel J. Cohen	2002
Diana Deutsch	2002
Uwe J. Hansen	2002
Roger J. Hanson	2002
Peter L. Hoekje	2002
Douglas H. Keefe	2002
Ian M. Lindevald	2002
Steven E. McAdams	2002
Gary P. Scavone	2002
Mark Pierre Verge	2002
Gabriel Weinreich	2002
R. Dean Ayers	2001
Judith C. Brown	2001
Courtney B. Burroughs	2001
Robert D. Collier	2001
Stephen F. Duncan	2001
George F. Emerson	2001
James H. Irwin	2001
Bozena Kostek	2001
Barry Larkin	2001
Thomas D. Rossing	2001
Julius O. Smith	2001
William J. Strong	2001
James W. Beauchamp	2000
René E. Caussé	2000
W. Jay Dowling	2000
William M. Hartmann	2000
Adrianus J. Houtsma	2000
Bruce A. Lawson	2000
James M. Pyne	2000
Shigeru Yoshikawa	2000

Ex officio:
Stanley L. Ehrlich, Associate Editor of JASA
William J. Strong, Associate Editor of JASA
Roger J. Hanson, member of Medals and Awards Committee
Uwe J. Hansen, member of Membership Committee
Ian M. Lindevald, member of ASACOS

Noise

	Term to
Richard J. Peppin, *Chair* to 2000	2000
Elliott H. Berger	2002
Stephen H. Bly	2002
Ann E. Bowles	2002
Frank H. Brittain	2002
Bennett M. Brooks	2002
Sanford Fidell	2002
Lawrence S. Finegold	2002
Lee D. Hager	2002
Murray R. Hodgson	2002
Peter C. Laux	2002
Jerry G. Lilly	2002
David Lubman	2002
George A. Luz	2002
Sally A. McInerny	2002
Luc Mongeau	2002
Michael T. Nixon	2002
Matthew A. Nobile	2002
Robert A. Putnam	2002
Mary M. Prince	2002
Jack E. Randorff	2002
Stephen I. Roth	2002
Paul D. Schomer	2002
Hideki Tachibana	2002
James K. Thompson	2002
Nancy S. Timmerman	2002
Laura A. Wilber	2002
Martin Alexander	2001
John P. Barry	2001
Leo L. Beranek	2001
Arno S. Bommer	2001
Anthony J. Brammer	2001
James O. Buntin	2001
Robert J. Cook	2001
Kenneth A. Cunefare	2001
Joseph P. Cuschieri	2001
Paul R. Donavan	2001
William D. Gallagher	2001
Gerald C. Lauchle	2001
Anna Maria Monslave	2001
Miomir Mijic	2001
Thomas Norris	2001
John P. Seiler	2001
Noral D. Stewart	2001
Michael R. Stinson	2001
Alice H. Suter	2001
Louis C. Sutherland	2001
Jiri Tichy	2001
Dennis Walton	2001
Keith Attenborough	2000
Sergio Beristain	2000
Robert D. Bruce	2000
John C. Burgess	2000
Angelo J. Campanella	2000
Robert J. Comparin	2000
T. James DuBois	2000
John J. Earshen	2000
Tony F. W. Embleton	2000
John Erdreich	2000
B. John Feng	2000
Robert D. Hellweg	2000
Robert M. Hoover	2000
Daniel L. Johnson	2000
Tor S. D. Kihlman	2000
Sonoko Kuwano	2000
Chantal Laroche	2000
Robert Lotz	2000
Alan H. Marsh	2000
Ralph T. Muehleisen	2000
Kenneth J. Plotkin	2000
Joseph Pope	2000
Brigitte Schulte-Fortkamp	2000
Jean Tourret	2000
Robert W. Young	2000

Ex officio:
Michael R. Stinson, Associate Editor of JASA
Stanley L. Ehrlich, Associate Editor of JASA
Elliott H. Berger, member of Medals and Awards Committee
Daniel L. Johnson, member of Membership Committee
Richard J. Peppin, member of ASACOS

Physical Acoustics

	Term to
Robert M. Keolian, *Chair* to 2002	2002
Michael R. Bailey	2002
Robin O. Cleveland	2002
Robert A. Hiller	2002
Thomas J. Hofler	2002
R. Glynn Holt	2002
Christopher C. Lawrenson	2002
Timothy G. Leighton	2002
Bart Lipkens	2002
G. Douglas Meegan	2002
Ralph T. Muehleisen	2002
Penelope Menounou	2002
Richard Raspet	2002
John S. Stroud	2002
Richard L. Weaver	2002
David T. Blackstock	2001
David A. Brown	2001
Gilles A. Daigle	2001
Bruce Denardo	2001
Steven L. Garrett	2001
Logan E. Hargrove	2001
Craig J. Hickey	2001
Andres Larraza	2001
William Moss	2001
George Mozurkewich	2001
Vladimir Ostashev	2001
Victor W. Sparrow	2001
Keith Attenborough	2000
Yves H. Berthelot	2000
James P. Chambers	2000
Robin O. Cleveland	2000
Kerry W. Commander	2000
D. Felipe Gaitan	2000
Mark F. Hamilton	2000
Steven G. Kargl	2000
Philip L. Marston	2000
Thomas J. Matula	2000
Kenneth J. Plotkin	2000
Harry Simpson	2000
D. Keith Wilson	2000

Ex officio:
Henry E. Bass, Associate Editor of JASA
Mack A. Breazeale, Associate Editor of JASA
Mark F. Hamilton, Associate Editor of JASA and member of Medals and Awards Committee
Courtney B. Burroughs, Associate Editor of JASA
Floyd Dunn, Associate Editor of JASA

Dale E. Chimenti, Associate Editor of JASA
Andrew N. Norris, Associate Editor of JASA
Paul J. Remington, Associate Editor of JASA
Louis C. Sutherland, Associate Editor of JASA
Steven L. Garrett, member of Membership Committee
Sameer I. Madanshetty, member of ASACOS

Psychological and Physiological Acoustics

	Term to
Neal F. Viemeister, *Chair* to 2002	2002
Huanping Dai	2002
Ann Clock Eddins	2002
Richard R. Fay	2002
Peggy B. Nelson	2002
Robert S. Schlauch	2002
Stanley E. Sheft	2002
Prudence Allen	2001
Robert P. Carlyon	2001
David F. Dolan	2001
Brent W. Edwards	2001
Nina Kraus	2001
John J. Rosowski	2001
David A. Eddins	2000
Mary Florentine	2000
John H. Grose	2000
Marjorie R. Leek	2000
Elizabeth A. Strickland	2000
Fan-Gang Zeng	2000

Ex officio:
D. Wesley Grantham, Associate Editor of JASA
Sid P. Bacon, Associate Editor of JASA
Robert V. Shannon, Associate Editor of JASA
Brenda L. Lonsbury-Martin, Associate Editor of JASA
Marjorie R. Leek, member of Medals and Awards Committee
Richard R. Fay, member of Membership Committee
Christine M. Rankovic, member of ASACOS

Speech Communication

	Term to
Emily A. Tobey, *Chair* to 2001	2001
Carol Y. Espy-Wilson	2002
Marios S. Fourakis	2002
John W. Hawks	2002
William F. Katz	2002
Patricia A. Keating	2002
Paul A. Luce	2002
Deborah M. Rekart	2002
Juergen Schroeter	2002
Janet W. Stack	2002
Mario A. Svirsky	2002
Rosalie M. Uchanski	2002
Gary G. Weismer	2002
Fredericka Bell-Berti	2001
H. Timothy Bunnell	2001
Astrid Schmidt-Nielsen	2001
Winifred Strange	2001
Douglas H. Whalen	2001
Lynne E. Bernstein	2000
Sadaoki Furui	2000
Kenneth W. Grant	2000
Jody E. Kreiman	2000
Patricia K. Kuhl	2000
Richard S. McGowan	2000
Susan N. Nittrouer	2000
Douglas D. O'Shaughnessy	2000
Robert F. Port	2000
Mitchell S. Sommers	2000
Ann K. Syrdal	2000

Ex officio:
Douglas D. O'Shaughnessy, Associate Editor of JASA
Anders Lofqvist, Associate Editor of JASA
Winifred Strange, member of Medals and Awards Committee
Maureen L. Stone, member of Membership Committee
Diane Kewley-Port, member of ASACOS

Structural Acoustics and Vibration

	Term to
Jerry H. Ginsberg, *Chair* to 2000	2000
John A. Burkhardt	2002
Kenneth D. Frampton	2002
Sunil Mehta	2002
Jeffrey S. Vipperman	2002
Richard L. Weaver	2002
Kuangcheng Wu	2002
Wen H. Lin	2001
Philip L. Marston	2001
James G. McDaniel	2001
Angie Sarkissian	2001
Benjamin Bard	2000
Alain C. Berry	2000
Hunter C. Cohen	2000
David Feit	2000
Guillermo C. Gaunaurd	2000
Karl Grosh	2000
Sabih I. Hayek	2000
Francis Kirschner	2000
Mauro Pierucci	2000
Andrew F. Seybert	2000
Scott D. Sommerfeldt	2000
Vasundara V. Varadan	2000
Richard D. Vogelsong	2000
Sean F. Wu	2000

Ex officio:
Paul J. Remington, Associate Editor of JASA
Courtney B. Burroughs, Associate Editor of JASA and member of Membership Committee
Gideon Maidanik, member of Medals and Awards Committee
Louis A. Herstein, member of ASACOS

Underwater Acoustics

	Term to
George V. Frisk, *Chair* to 2000	2000
Paul B. Baxley	2002
Shira L. Broschat	2002
Douglas H. Cato	2002
Peter H. Dahl	2002
Grant B. Deane	2002
Gerald L. D'Spain	2002
Garry H. Heard	2002
Charles W. Holland	2002
Samuel W. Marshall	2002
Kevin B. Smith	2002
Dajun Tang	2002
Christopher T. Tindle	2002
Alexandra I. Tolstoy	2002
Stephen N. Wolf	2002

Michael G. Brown	2001
Dennis B. Creamer	2001
Christian P. de Moustier	2001
Stanley E. Dosso	2001
Stewart A. L. Glegg	2001
Zoi-Heleni Michalopoulou	2001
Marshall H. Orr	2001
Gregory J. Orris	2001
James C. Preisig	2001
Daniel Rouseff	2001
William L. Siegmann	2001
David L. Bradley	2000
Curtis I. Caldwell	2000
William M. Carey	2000
Nicholas P. Chotiros	2000
Donald R. Del Balzo	2000
Frederick R. DiNapoli	2000
David R. Dowling	2000
Robert W. Farwell	2000
Roger C. Gauss	2000
Darrell R. Jackson	2000
Finn B. Jensen	2000
Roger W. Meredith	2000
John R. Preston	2000
Henrik Schmidt	2000
Ralph A. Stephen	2000
Robert D. Stoll	2000
Frederick D. Tappert	2000
Alexander G. Voronovich	2000
R. Jeffrey Wilkes	2000
Tsih C. Yang	2000
Robert A. Zingarelli	2000

Ex officio:
John C. Burgess, Associate Editor of JASA
Stanley A. Chin-Bing, Associate Editor of JASA
David L. Bradley, Associate Editor of JASA and member of Medals and Awards Committee
Peter H. Rogers, member of Membership Committee
Arnie L. Van Buren, member of ASACOS

Interdisciplinary Technical Group on Signal Processing in Acoustics

	Term to
James V. Candy, *Chair* to 2000	2000
David J. Evans	2002
Charles F. Gaumond	2002
Joseph Pope	2002
Leon H. Sibul	2002
Randall L. Smith	2002
Martin Barlett	2001
David H. Chambers	2001
Elmer Hixson	2001
Ning Xiang	2001
John C. Burgess	2000
Christian P. deMoustier	2000
Gary W. Elko	2000
David I. Havelock	2000
Sterios Stergiopoulos	2000
Julius O. Smith, III	2000
Yoshio Yamasaki	2000

Ex officio:
John C. Burgess, Associate Editor of JASA

Administrative Committees 1999–2000

Archives and History

	Term to
Henry E. Bass, *Chair* to 2000	2000
Leo L. Beranek	2002
William J. Cavanaugh	2002
Logan E. Hargrove	2002
James C. Saunders	2001
Harry Schecter	2001
Rosalie M. Uchanski	2001
Richard K. Cook	2000
John W. Kopec	2000
Richard J. Peppin	2000

Books^{+}

	Term to
Mohsen Badiey, *Chair* to 2002	2002
Nancy S. McGarr	2002
Jeffrey A. Nystuen	2002
Neil A. Shaw	2002
Emily A. Tobey	2002
Stanley L. Chin-Bing	2001
Gordon E. Martin	2001
Victor W. Sparrow	2001
Robert D. Frisina	2000
Jerry H. Ginsberg	2000
Philip L. Marston	2000
Joseph Pope	2000
Robert A. Walkling	2000
Stephen N. Wolf	2000

Ex officio:
Philip L. Marston, Associate Editor of JASA for Book Reviews

Education in Acoustics

	Term to
Victor W. Sparrow, *Chair* to 2000	2000
Courtney B. Burroughs	2002
Robin O. Cleveland	2002
Kenneth A. Cunefare	2002
D. Michael Daly	2002
Mary Florentine	2002
Roger J. Hanson	2002
Logan E. Hargrove	2002
Peter L. Hoekje	2002
Darrell R. Jackson	2002
Michel T. T. Jackson	2002
Murray F. Korman	2002
Luc Mongeau	2002
John S. Robertson	2002
James M. Sabatier	2002
Neil A. Shaw	2002
Kevin B. Smith	2002
Ralph A. Stephen	2002
James E. West	2002
Wayne M. Wright	2002
David A. Brown	2001
Robert D. Collier	2001
Corinne M. Darvennes	2001

Margaritis S. Fourakis	2001
Carole E. Gelfer	2001
Douglas R. Jones	2001
Sharon Y. Manuel	2001
Philip L. Marston	2001
Ana Maria Monsalve	2001
Ralph Muehleisen	2001
Andrew A. Piacsek	2001
Daniel R. Raichel	2001
Sally G. Revoile	2001
Thomas D. Rossing	2001
Ronald A. Roy	2001
Dawn R. Schuette	2001
Scott D. Sommerfeldt	2001
William Thompson, Jr.	2001
Robert A. Walkling	2001
George S. K. Wong	2001
Anthony A. Atchley	2000
Fredericka Bell-Berti	2000
E. Carr Everbach	2000
Uwe J. Hansen	2000
Elizabeth S. Ivey	2000
P. K. Raju	2000
Daniel A. Russell	2000
Emily A. Tobey	2000

Ethics and Grievances

	Term to
Robert E. Apfel, *Chair* to 2000	2000
Mardi C. Hastings	2001
William A. Yost	2001
William J. Cavanaugh	2000
Carol Espy-Wilson	2000

Investments

	Term to
John V. Bouyoucos and Leo L. Beranek, *Co-chairs* to 2001	2001
Ira J. Hirsh	2002
Ira Dyer	2002
Kenneth M. Eldred	2001

Paul B. Ostergaard, Treasurer, *ex officio*

Long Range Planning

	Term to
Anthony A. Atchley, *Chair* to 2000	2000
Scott D. Pfeiffer	2002
Stephen C. Thompson	2002
Nancy S. Timmerman	2002
Fredericka Bell-Berti	2001
Ervin R. Hafter	2001
Louis C. Sutherland	2001
Dana S. Hougland	2000
Scott D. Sommerfeldt	2000
Murray Strasberg	2000

Katherine S. Harris, President-Elect, *ex officio*

Medals and Awards

		Term to
D. Vance Holliday, *Chair* to 2000		2000
Steven M. Brown	Architectural Acoustics	2002
David L. Bradley	Underwater Acoustics	2002
Mardi C. Hastings	Animal Bioacoustics	2002
Gideon Maidanik	Structural Acoustics and Vibration	2002
Michael J. Buckingham	Acoustical Oceanography	2001
Robert D. Finch	Engineering Acoustics	2001
Roger J. Hanson	Musical Acoustics	2001
Winifred Strange	Speech Communication	2001
Elliott H. Berger	Noise	2000
Mark F. Hamilton	Physical Acoustics	2000
Marjorie R. Leek	Psychological and Physiological Acoustics	2000
Wesley L. Nyborg	Biomedical Ultrasound/ Bioresponse to Vibration	2000

Meetings

Gilles A. Daigle, *Chair* to 2002
Yves H. Berthelot, Spring 2000, Atlanta
Lawrence A. Crum, Spring 1998, Seattle
Gilles A. Daigle, Vice-President Elect
Mardi C. Hastings, Fall 1999, Columbus
Mauro Pierucci, Vice President
Kevin P. Shepherd, Fall 1998, Norfolk
Sigfrid D. Soli, Fall 2000, Newport Beach
Jiri Tichy, Spring 1999, Berlin

Elaine Moran, ASA Office Manager, *ex officio*
Charles E. Schmid, Executive Director, *ex officio*

Subcommittee on Exhibits

Martin A. Alexander, *Chair*
Tony F. W. Embleton
Robert Finnegan
Sabih I. Hayek
Noland L. Lewis

Membership

		Term to
Joseph W. Dickey, *Chair* to 2000		2000
Anthony J. Brammer	Biomedical Ultrasound/ Bioresponse to Vibration	2002
Courtney B. Burroughs	Structural Acoustics and Vibration	2002
Burton G. Hurdle	Foreign Members	2002
Daniel L. Johnson	Noise	2002
Christopher Feuillade	Acoustical Oceanography	2001
Steven L. Garrett	Physical Acoustics	2001
Peter H. Rogers	Underwater Acoustics	2001
Gregory C. Tocci	Architectural Acoustics	2001
Richard R. Fay	Psychological and Physiological Acoustics	2000
Uwe J. Hansen	Musical Acoustics	2000
Thomas R. Howarth	Engineering Acoustics	2000
Sam H. Ridgway	Animal Bioacoustics	2000
Maureen L. Stone	Speech Communication	2000

Public Relations

	Term to
Paul A. Baxley, *Chair* to 2000	2000
John Erdreich	2002
Logan E. Hargrove	2002
Burton G. Hurdle	2002
Ronald T. Kessel	2002
James F. Lynch	2002
Duncan E. McGehee	2002
Joanne L. Miller	2002
Joseph Pope	2002
Barbara J. Sotirin	2002
E. Carr Everbach	2001
Charles Gaumond	2001
Christy K. Holland	2001
David Lubman	2001
Andrew A. Piacsek	2001
Anthony J. Brammer	2000
William J. Cavanaugh	2000
T. James DuBois	2000
Blas Espinoza-Varas	2000
Holly S. Haggerty	2000
Helen Ann McCaffrey	2000
Victor Nedzelnitsky	2000
Mauro Pierucci	2000
Thomas D. Rossing	2000
Bor-Tsuen Wang	2000
Ewart A. Wetherill	2000

Allan D. Pierce, Editor-in-Chief, *ex officio*
Elaine Moran, ASA Office Manager, *ex officio*
Charles E. Schmid, Executive Director, *ex officio*
Thomas D. Rossing, Echoes Editor, *ex officio*

Publication Policy

	Term to
Floyd Dunn, *Chair* to 2000	2000
Robert C. Bilger	2002
James F. Lynch	2002
James H. Miller	2002
George S. K. Wong	2002
Allan J. Zuckerwar	2002
Sigfrid D. Soli	2001
Richard Stern	2001
Michael R. Stinson	2001
A. Simmons	2000

Katherine S. Harris, President-Elect, *ex officio*
Allan D. Pierce, Editor-in-Chief, *ex officio*

Regional Chapters

Roger T. Richards, *Chair* to 2000

Elmer L. Hixson	Austin
Angelo J. Campanella	Central Ohio
Courtney B. Burroughs	Central Pennsylvania
John W. Kopec	Chicago
Ernest M. Weiler	Cincinnati
Edwin H. Toothman	Delaware Valley
Gary W. Siebein	Florida
Yves H. Berthelot	Georgia
Timothy J. Foulkes	Greater Boston
Robert M. Hoover	Houston
Michael J. Anderson	Inland Northwest
Neil A. Shaw	Los Angeles
Hari S. Paul	Madras, India
Marehalli G. Prasad	Metropolitan New York
Elizabeth A. McLaughlin	Narragansett
Larry H. Royster	North Carolina
Peter F. Assmann	North Texas
James R. Angerer	Northwest
R. Dean Ayers	Orange County
Paul Baxley	San Diego
David Braslau	Upper Midwest
VACANT	Washington, D.C.
Thomas M. Disch	Wisconsin

Victor W. Sparrow, Chair, Education in Acoustics, *ex officio*

Rules and Governance

	Term to
Tony F. W. Embleton, *Chair* to 2002	2002
William M. Hartmann	2002
Richard H. Lyon	2002
William J. Cavanaugh	2001
Floyd Dunn	2001
Elaine Moran	2000
Charles E. Schmid	2000

Special Fellowships

	Term to
Wayne M. Wright, *Chair* to 2000	2000
Uwe J. Hansen	2002
John Erdreich	2001
Constantine Trahiotis	2001
James D. Miller	2000
Gary W. Elko	2000

Standards

Executive Committee
Daniel L. Johnson, Chair (Standards Director)
Paul D. Schomer, Vice Chair
Avril Brenig, Standards Manager, *ex officio*

S1 Representation
John P. Seiler, Chair S1 and ASA representative on S1
George S. K. Wong, Vice Chair S1 and ASA alternate representative on S1

S2 Representation
David J. Evans, Chair S2
Richard F. Taddeo, Vice Chair
Sabih I. Hayek, ASA representative on S2
Bruce E. Douglas, ASA alternate representative on S2

S3 Representation
R. F. Burkard, Chair S3 and ASA representative on S3
J. Franks, Vice Chair S3 and ASA alternate representative on S3

S12 Representation
Paul D. Schomer, Chair S2
Robert D. Hellweg, Vice Chair S12
Bennett M. Brooks, ASA representative on S12
William J. Galloway, ASA alternate representative on S12

International TAGs (ex officio)
Paul D. Schomer, Chair, U.S. TAG for ISO/TC 43 and ISO/TC 43/SC1
David J. Evans, Chair, U.S. TAG for ISO/TC 108
Victor A. Nedzelnitsky, U.S. Technical Advisor for IEC/TC 29

ASA Technical Committee Representatives
Mauro Pierucci, Chair of ASA Technical Council, *ex officio*
Er Chang Shang, Acoustical Oceanography
Ann E. Bowles, Animal Bioacoustics
George E. Winzer, Architectural Acoustics
Robin O. Cleveland, Biomedical Ultrasound/Bioresponse to Vibration
Mahlon D. Burkhard, Engineering Acoustics
Ian M. Lindevald, Musical Acoustics
Richard J. Peppin, Noise
Sameer I. Madanshetty, Physical Acoustics
Christine M. Rankovic, Psychological and Physiological Acoustics
Diane Kewley-Port, Speech Communication
Louis A. Herstein, Structural Acoustics and Vibration
Arnie L. Van Buren, Underwater Acoustics

ASA Officers
Paul B. Ostergaard, Treasurer, *ex officio*
Charles E. Schmid, Executive Director, *ex officio*

Past Chair of ASACOS (*ex officio*)
Tony F. W. Embleton

Associate Editors for Standards News—JASA (ex officio)
Avril Brenig
George S. K. Wong

Tutorials

	Term to
Joseph Pope, *Chair* to 2000	2000
Yves H. Berthelot	2002
Uwe J. Hansen	2002
Ralph T. Muehleisen	2002
Ann E. Bowles	2001
Douglas D. O'Shaughnessy	2001
Gary W. Siebein	2001
Beverly A. Wright	2001
Kenneth E. Gilbert	2000
Kenneth J. Plotkin	2000

Charles E. Schmid, Executive Director, *ex officio*

Women in Acoustics

	Term to
Nancy S. Timmerman, Cochair	2000
Zoi-Heleni Michalpoulou, Cochair	2000
Uwe J. Hansen	2002
Mardi C. Hastings	2002
Michelle Langlais	2002
Peggy B. Nelson	2002
Linda M. Carroll	2001
Lawrence A. Crum	2001
Corinne M. Darvennes	2001
Laura K. Smith	2001
Arlene E. Carney	2000
Chi-Fang Chen	2000
Peter H. Dahl	2000
Helen M. Hanson	2000
Elizabeth S. Ivey	2000
Bozena Kostek	2000

OBITUARIES

This section of the Journal publishes obituaries concerning the death of Fellows of the Society and other acousticians eminent in the world of acoustics. When notified, the Editor-in-Chief solicits a summary of the person's life and contributions from an ASA member thoroughly familiar with the details, if possible. If a promised obituary is never received, a brief obituary notice may be published later.

Sir James Lighthill • 1924–1998

Sir James Lighthill died at age 74 as he was just completing—for the seventh time—a swim round the Channel Island of Sark, having 25 years ago been the first person to manage this marathon in which the swimmer has to cope with massive waves and treacherous currents. No doubt he was putting on a finishing spurt, but sadly that exertion at the end of 10 h in unusually heavy seas revealed a mitral valve condition leading to his sudden death in the water; but no doubt this was also one of the ways in which this so-much-larger-than-life figure would have wished to go.

Lighthill was one of the greatest scientists of this century in fluid mechanics, acoustics, and applied mathematics, initiating whole new fields, contributing very many novel and seminal ideas, and inspiring scientific communities by his personal example and commitment. Born in 1924 in Paris, he went to school in England at Winchester (a contemporary there of Freeman Dyson), excelled there across the board, and at age 17 went up to Trinity College Cambridge for the then-standard two-year wartime mathematics course, during which he met Nancy Dumaresq, also a mathematician, whom he married in 1945. He then went to the National Physical Laboratory at Teddington, to work on supersonic aerodynamics. His earliest papers have all the hallmarks of the roughly 160 to follow: the clearest imaginable presentation of the scientific issue at stake, the most cogent arguments behind the mathematical model and the necessary approximations (always motivated by physical reasoning rather than by formal asymptotics), the most elegant and powerful mathematical analysis, and the most compelling interpretation—in scientific and practical terms—of the outcome.

From NPL Lighthill went quickly to Manchester, and at age 26 succeeded Sydney Goldstein as Beyer Professor of Applied Mathematics. Lighthill's time (1946–1959) at Manchester was one of extraordinary creativity and productivity. He worked ingeniously and prolifically on a huge range of topics in fluid mechanics and applied mathematics, and supervised many Ph. D. students while discharging a heavy administrative load, teaching with great verve and gusto (his beautiful 1958 book *An Introduction to Fourier Analysis and Generalised Functions* was prepared for Manchester undergraduate courses), and assuming major national and international responsibilities. Leaving Manchester in 1959, he went directly to the Royal Aircraft Establishment, Farnborough, in the exciting days of Concorde design. He managed a staff of 8000, of whom 1400 were professional scientists and engineers. All were astonished to find each and every report and paper emanating from RAE subjected to the eagle eye of the Director, a Director, moreover, whose personal scientific output was sufficiently voluminous and distinctive to warrant a special series of RAE technical reports. Rarely indeed has there been a major applied scientific institution of this scale in which the Director had such complete technical mastery of its scientific output—although naturally Lighthill did not confine himself simply to that, but immersed himself in all aspects of RAE finance, administration, and management.

In 1964 Lighthill took a Royal Society Research Professorship at Imperial College London (he had been elected FRS at the very early age of 29), and left the world of aerodynamics for those of waves, geophysics, biomechanics, and mathematical biology. My colleague Julian Hunt conjectures, perceptively, that Lighthill was so disillusioned by the massive cuts inflicted on all civil and military aerospace funding by the 1964 Wilson government that his interests in all things aeronautical largely ceased (and this reaction to the Wilson government appears to have been punished by a delay until 1971 in the award of a knighthood to Lighthill). Among those interests, regrettably, was that in *aeroacoustics*, the subject which he founded with his famous 1952 paper ''On sound generated aerodynamically. Part I. General theory,'' a quite remarkable paper neither containing nor needing a single reference to any prior work.

From Imperial College Lighthill went to take the Lucasian Professorship of Mathematics in Cambridge in 1969, succeeding Dirac and, three centuries earlier, Newton, relinquishing the Chair to Stephen Hawking in 1979 for the office of Provost of University College London. Retiring ten years later, after impressive development and expansion of UCL in the number and quality of its students and academic staff no less than in its physical facilities, Lighthill devoted his last nine years to biological mechanics, to geophysics, and to Chairmanship of the Special Committee on the International Decade for Natural Disaster Reduction. His achievements were recognized by innumerable awards and medals, by 24 Honorary Degrees, and by membership of the most prestigious academies worldwide.

His launching of the field of *aeroacoustics* in 1952 is well known. To be emphasized is the fact that his theory has underpinned *all subsequent work*, analytical, experimental, and computational, in this field; and that his theory is so versatile and all-embracing that it has been used to predict not only the noise from jets, propellers, and turbomachinery, but also that from the convective turbulence in the atmosphere of the sun, and that from the splashing of spray, the splitting of bubbles and the interaction of surface water waves. And although J. S. Mendousse [J. Acoust. Soc. Am. **25**, 51–54 (1953)] had already, unknown to Lighthill, made the connection between Burgers' equation and finite-amplitude effects in sound, and Z. A. Gol'dberg [Sov. Phys. Acoust. **3**, 340–347 (1957)] had recognized the competition between nonlinear steepening and linear diffusion effects, it was Lighthill's 100-page 1956 article ''Viscosity effects in sound waves of finite amplitude,'' in honor of the 70th birthday of G. I. Taylor, that really launched the field of *nonlinear acoustics*, dealing, as it did, with shock formation, propagation, and decay in plane and other quasi-one-dimensional geometries, and with thermoviscous and relaxation effects on shocks, pulses, and periodic waves in the ''nonlinear acoustics approximation.'' Lighthill's 1997 work on ''shock bunching'' in randomly irregular shocked flows provides a seamless continuation of the 1956 paper.

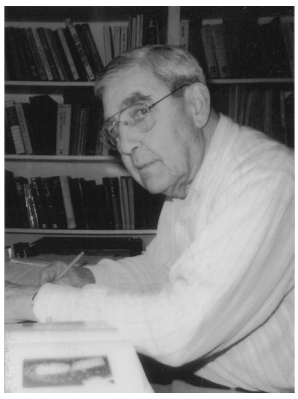
But Lighthill's contributions to acoustics go further than these two major efforts. He supervised, among many others, N. S. Curle (solid surface effects in aeroacoustics), G. B. Whitham (shock dynamics and shock diffraction), M. A. Swinbanks (pioneering work on active control of sound in ducts), and M. S. Howe (whose notable contributions to acoustics are simply too wide-ranging and numerous to itemize here); and he himself worked on acoustic streaming, streaming in the ear itself, on cochlear mechanics, on biochemical feedback at low auditory levels, and on the mathematical theory of wave generation and propagation, linear and nonlinear, in dispersive and possibly anisotropic, inhomogeneous and moving media.

He served as founding President of the Institute of Mathematics and its Applications (1965–67), as President of the International Union for Theoretical and Applied Mechanics (1984–88), and as founding President, from 1995 to the time of his death, of the International Institute of Acoustics and Vibration.

His contributions to acoustics are to be set alongside those of Rayleigh, and while his contributions to fluid mechanics and applied mathematics have similar permanence and status, he would have been content with this accolade.

D. G. CRIGHTON

David E. Goldman • 1910–1998



Dr. David E. Goldman, head of the Biophysics Division of the Naval Medical Research Institute in Bethesda MD (1947–1967), and professor of Physiology and Biophysics at the Medical College of Pennsylvania, Philadelphia (1969–1976), died on 22 July 1998 at his Falmouth, Cape Cod home after an extended illness. Dave was a member of the Acoustical Society of America since 1951. He was the author of over 100 scientific publications on the effects of high intensity noise vibration and shock on people and on the electrical and mechanical phenomena in cells and tissues.

Born in Boston in 1910, Dave earned his bachelors degree in physics at Harvard and his doctorate degree in physiology at Columbia University in 1943. His doctorate work produced the well-known Goldman equation relating ion permeabilities to cell membrane potentials. In early 1952, after joining the Navy, Dave and E. S. Mendelson conducted a shipboard study on noise levels associated with flight operations aboard the U.S.S. Coral Sea. The results of the survey indicated the acute noise problems on aircraft carriers and the problems anticipated with more powerful jet engines and jet engines with afterburners. When the Navy turned to the National Research Council for advice, this event initiated (in the same year, 1952) the Triservice interest in the formulation of the NRC ''Committee on Hearing and Bioacoustics'' (CHABA). The committee, ''Biomechanics,'' was later added to its responsibilities and served the U.S. Government and academic communities admirably for over four decades. CHABA, at its first council meeting, established a working group, with Hallowell Davis as chair, to survey the whole spectrum of problems connected with high intensity noise from aircraft operations. The project resulted in the recommendations of the BENOX report (Biological Effects of Noise, Exploratory, 1953), which for years to come determined protection and research approaches. Dave Goldman prepared the section in this report on ''Bioacoustics of Tissue, Vibration and Ultrasonics.'' In the years to follow, he served the emerging biophysical disciplines well as consultant to the National Academy of Sciences, the National Research Council, the National Institutes of Health, and the President's Office of Science and Technology. From 1955 to 1957 he served as Scientific Liaison Officer to the Office of Naval Research in London, U.K., and expanded his scientific contacts and influence to many international friends in bioacoustics and biophysics. In 1958 he was member of the NIH Biophysics and Biophysical Chemistry Study Section, which also included Richard Bolt, and which organized the study program on Biophysical Science under F. O. Schmitts' Chairmanship. This committee promoted the advances and increasingly productive interactions of the physical and biological sciences.

In 1973 Dave received the Kenneth S. Cole award from the Biophysical Society. He was a member of many scientific societies.

Dave was a musician and music lover; he loved to travel in the United States and Europe with his wife Jeanne of 60 years (and good friend for 70 years). His son Dr. James Eliot Goldman and two grandchildren also survive him.

Dave was a modest and quiet man, full of humor, and with broad interests. It was a pleasure to collaborate with him, which I had the privilege to do for over 30 years. Through his early reviews and publications he contributed significantly to the recognition of vibration, shock and noise problems. He promoted the protection of personnel against such hazards through prevention and research. We will miss his knowledgeable objective and fair voice and above all, his wonderful personality. But we are thankful for his lifelong contributions.

HENNING E. VON GIERKE

Scattering of a plane acoustic wave from a transversely isotropic cylinder encased in a solid elastic medium

Y. Fan and A. N. Sinclair[a)]
Department of Mechanical and Industrial Engineering, University of Toronto, 5 Kings College Road, Toronto, Ontario M5S 3G8, Canada

F. Honarvar
Department of Mechanical Engineering, K. N. Toosi University of Technology, P.O. Box 16765-3881, Tehran 16579, Iran

(Received 3 April 1999; accepted for publication 28 May 1999)

A mathematical model is developed to describe the scattering of a plane wave incident at an arbitrary angle on a transversely isotropic cylinder embedded in a solid elastic matrix. The model is based on the normal-mode expansion method, but is complicated by the coupling between the potential functions representing compressional and axially polarized shear waves. The solid matrix around the cylinder precludes the appearance of the leaky Rayleigh modes that dominate the spectrum of a cylinder immersed in a fluid. Instead, interfacial modes contribute to the scattered spectrum—these modes could be instrumental in the development of a nondestructive evaluation technique for the matrix-to-fiber bonds in a fiber-reinforced composite material. The presence of a solid matrix that supports shearing action leads to scattered compressional waves, and shear waves with polarization components in both the axial direction and r-θ planes. The sensitivity of resonances in each of these scattered wave components to perturbations in the cylinder's elastic constants is explored; results indicate that a judicious selection of resonance modes allows characterization of the cylinder's elastic properties.
[S0001-4966(99)01609-4]

PACS numbers: 43.20.Bi, 43.20.Fn, 43.20.Ks [AN]

INTRODUCTION

The ultrasonic nondestructive evaluation (NDE) of fiber-reinforced composite materials is hampered by the highly complicated response function expected from even an undamaged specimen. The fibers are typically cylindrical in shape, each containing up to several transversely isotropic layers. The interface between the fiber and the surrounding elastic matrix must be strong, as a low pull-out strength will compromise the material's integrity under stress. In this paper, we determine the diffraction spectrum resulting from oblique incidence of a plane compression or shear wave on a single homogeneous cylinder with transversely isotropic symmetry, encased in a solid elastic matrix.

This effort is built upon a wealth of earlier work dealing with compression wave scattering from homogeneous, isotropic cylinders and shells immersed in a fluid. Following the pioneering work in this field by Faran[1] and White,[2] further studies by Flax and Neubauer,[3] Nagl *et al.*,[4] and Veksler[5] led to a broader understanding of the phenomenon. This included the development of Resonance Scattering Theory (RST) that explained the resonances in the scattered wave spectrum as the product of constructive interference among surface waves encircling the cylinder. Identification of the various wave modes, and their associated shapes, was assisted by formulating the scattered spectrum as an expansion of normal modes. Conoir *et al.*[6] focused in particular on the wave modes and resonant frequencies resulting from compression waves incident over a range of oblique angles α. Comprehensive reviews of these topics and extensive bibliographies can be found in the works of Uberall,[7] Pao and Mow,[8] Gaunaurd,[9] Uberall,[10] and Veksler.[11] Relatively few studies have been conducted on scattering from cylinders encased in a solid elastic matrix. Lewis and Kraft applied a normal-mode decomposition technique to this problem, although their solution is valid only in the long wavelength limit.[12] Flax and Uberall[13] conducted numerical studies on the interfacial waves that arise when compression waves are incident on spherical iron inclusions inside a solid aluminum matrix; experimental verification of their results was obtained only recently,[14] and included a brief investigation on Stoneley waves on the matrix/inclusion interface. Hsieh *et al.* noted that the propagation of Stoneley waves along an interface is sensitive to several boundary parameters, such as bond integrity and degree of contact.[15] Beattie *et al.* obtained reasonable agreement between theory and experiment for plane waves normally incident on isotropic rods embedded in epoxy.[16]

Possible application of this line of investigation was pursued by Sinclair and Addison, who compared the experimental and theoretical diffraction spectrum of a SiC fiber in a titanium matrix;[17] similar studies were later conducted for steel wires in plastic and included the cases of both good and bad interfacial adhesion.[18] In both sets of studies, only normally incident compression waves were employed, such that the transversely isotropic nature of the scattering cylinder did not play a role. Huang *et al.* suggested the incorporation of a spring boundary condition to represent the condition of the interface in numerical modeling.[19,20] Honarvar and Sinclair

[a)]Author to whom correspondence should be addressed; electronic mail: sinclair@mie.utoronto.ca

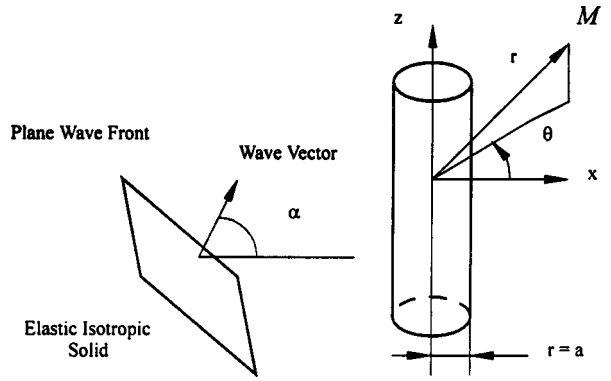


FIG. 1. Infinite plane wave incident at an angle α on a transversely isotropic cylinder embedded in a solid isotropic matrix. The observation point is $M(r,\theta,z)$.

proposed a systematic method to extract elastic constants from the diffracted spectrum resonances, called Material Characterization by Resonance Acoustic Spectroscopy (MCRAS).[21]

Many manufactured rods, fibers, and other cylindrical structures have transverse isotropy. There has been considerable study of acoustic wave-guide propagation along such structures, e.g., Refs. 22 and 23, but relatively little work on acoustic scattering from transversely isotropic cylindrical structures. Experimental measurements on scattering spectra from anisotropic cylinders in water were reported by de Billy.[24,25] More recently, Honarvar and Sinclair developed an exact normal-mode expansion for the scattering of a compression wave from an immersed, transversely isotropic cylinder,[26] and compared theoretical to measured spectra for the case of a fiber-reinforced rod with high axial stiffness.[21] The extension to be pursued here to a transversely isotropic cylinder encased in an elastic matrix adds considerable complication; now shear components of various polarizations must be included in both the incident and scattered waveform.

I. THEORY

A. Displacement potential functions

Consider a plane wave traveling through an infinite, isotropic medium, incident at an angle α on a long, transversely isotropic cylinder of radius $r=a$ (Fig. 1). The cylinder may be either a stiff rod, or a flexible reinforcing fiber encased in a matrix. The displacement field $\mathbf{u}(r,\theta,z,t)$ in both the matrix and cylinder can be expressed in terms of three scalar potentials ϕ, ψ, and χ,[26]

$$\mathbf{u}=\nabla\phi+\nabla\times(\chi\mathbf{e}_z)+a\nabla\times\nabla\times(\psi\mathbf{e}_z). \tag{1}$$

For the case $\alpha=0$, the three potential functions in Eq. (1) correspond to compressional, transversely (r-θ plane) polarized shear, and axially polarized shear components, respectively; such a tidy decomposition of Eq. (1) into constituent parts is not possible for nonzero values of incident angle α.

In the matrix material, each of the three wave potentials can have two components: an ''incident'' plane wave component, and a ''scattered'' component corresponding to the field originating from the cylinder. The method of solution is to first develop general expressions for the wave potentials in both the cylinder (material 1) and matrix (material 2), by substituting Eq. (1) into the equation of motion. Values for the unknown coefficients in the potentials are then determined by applying the appropriate boundary conditions at the surface of the embedded cylinder.

For the case where the incident wave is compressional with circular frequency ω, the associated normalized potential function in the matrix (material 2) has the general form,[17]

$$\phi_{2,\text{incident}}=\sum_{n=0}^{\infty}\epsilon_n(i)^n J_n(K_{2\perp}r)\cos(n\theta)\exp(i[K_{2z}z-\omega t]), \tag{2}$$

where $K_{2\perp}=K_2\cos\alpha$, $K_{2z}=K_2\sin\alpha$, and K_2 is the compressional wave number in the matrix material. The symbol ϵ_n is the Neumann factor ($\epsilon_n=1$ for $n=0$, and $\epsilon_n=2$ for $n>0$). The J_n terms are Bessel functions of the first kind. For a shear wave polarized in the r-θ plane, the associated normalized potential is

$$\chi_{2,\text{incident}}=\sum_{n=0}^{\infty}\epsilon_n(i)^n J_n(k_{2\perp}r)\sin(n\theta)\exp(i[k_{2z}z-\omega t]), \tag{3}$$

where $k_{2\perp}=k_2\cos\alpha$, $k_{2z}=k_2\sin\alpha$, and k_2 is the shear wave number in the matrix. Last, for the case of an incident shear wave polarized in the r-z plane, the associated potential function is[16]

$$\psi_{2,\text{incident}}=\sum_{n=0}^{\infty}\epsilon_n(i)^n J_n(k_{2\perp}r)\cos(n\theta)\exp(i[k_{2z}z-\omega t]). \tag{4}$$

Substitution of the appropriate combination of Eqs. (2), (3), and (4) into Eq. (1) then yields the total displacement field $\mathbf{u}_{2,\text{incident}}$ associated with the incident wave.

Consider now the transversely isotropic cylinder (material 1) of density ρ_1, where the constitutive relationship is characterized by five independent elastic constants,

$$\begin{bmatrix}\sigma_{rr}\\ \sigma_{\theta\theta}\\ \sigma_{zz}\\ \sigma_{\theta z}\\ \sigma_{rz}\\ \sigma_{r\theta}\end{bmatrix}=\begin{bmatrix}c_{11}&c_{12}&c_{13}&0&0&0\\ c_{12}&c_{11}&c_{13}&0&0&0\\ c_{13}&c_{13}&c_{33}&0&0&0\\ 0&0&0&c_{44}&0&0\\ 0&0&0&0&c_{44}&0\\ 0&0&0&0&0&\dfrac{(c_{11}-c_{12})}{2}\end{bmatrix}\times\begin{bmatrix}\epsilon_{rr}\\ \epsilon_{\theta\theta}\\ \epsilon_{zz}\\ 2\epsilon_{\theta z}\\ 2\epsilon_{rz}\\ 2\epsilon_{r\theta}\end{bmatrix}. \tag{5}$$

The displacement field $\mathbf{u}_1$ for $r<a$ can also be expressed in the format of Eq. (1). The associated potential functions required to give a nontrivial solution are given by[26]

$$\phi_1=\sum_{n=0}^{\infty}\left[A_nJ_n(s_1r)+q_2B_nJ_n(s_2r)\right]\times\cos(n\theta)\exp(i[\kappa_z z-\omega t]), \tag{6a}$$

$$\psi_1=\sum_{n=0}^{\infty}\left[q_1A_nJ_n(s_1r)+B_nJ_n(s_2r)\right]\times\cos(n\theta)\exp(i[\kappa_z z-\omega t]), \tag{6b}$$

$$\chi_1=\sum_{n=0}^{\infty}C_nJ_n(s_3r)\sin(n\theta)\exp(i[\kappa_z z-\omega t]), \tag{6c}$$

where the unknown coefficients A_n, B_n, and C_n have yet to be determined. Continuity of stress and displacement at the rod/matrix interface dictate that the axial wave vector κ_z inside the rod must be equal to the axial wave vector outside of the rod: K_{2z} for the case of an incident compression wave, or k_{2z} in the case of an incident shear wave. It is noted that Eqs. (6a) and (6b) are coupled together, while Eq. (6c) is uncoupled. (This is analogous to the coupling between a compression wave and vertically polarized shear wave inside a plate, while the horizontally polarized shear wave is uncoupled.) The wave numbers s_1, s_2, and s_3 are given by

$$s_1^2=\frac{\xi-\sqrt{\xi^2-4\zeta c_{11}c_{44}}}{2c_{11}c_{44}}, \tag{7a}$$

$$s_2^2=\frac{\xi+\sqrt{\xi^2-4\zeta c_{11}c_{44}}}{2c_{11}c_{44}}, \tag{7b}$$

$$s_3^2=\frac{2(\rho_1\omega^2-c_{44}\kappa_z^2)}{c_{11}-c_{12}}, \tag{7c}$$

where

$$\xi=(c_{13}+c_{44})^2\kappa_z^2+c_{11}(\rho_1\omega^2-c_{33}\kappa_z^2)+c_{44}(\rho_1\omega^2-c_{44}\kappa_z^2) \tag{8}$$

and

$$\zeta=(\rho_1\omega^2-c_{44}\kappa_z^2)(\rho_1\omega^2-c_{33}\kappa_z^2). \tag{9}$$

The parameters q_1 and q_2 appearing in Eq. (6) are given by

$$q_1=-\frac{-c_{11}s_1^2-(c_{13}+2c_{44})\kappa_z^2+\rho_1\omega^2}{ai\kappa_z[-(c_{11}-c_{13}-c_{44})s_1^2-c_{44}\kappa_z^2+\rho_1\omega^2]}, \tag{10}$$

$$q_2=-\frac{ai\kappa_z[-(c_{11}-c_{13}-c_{44})s_2^2-c_{44}\kappa_z^2+\rho_1\omega^2]}{-c_{11}s_2^2-(c_{13}+c_{44}\kappa_z^2+\rho_1\omega^2)}. \tag{11}$$

Last, one must consider the component of the displacement field **u** in the matrix material originating from waves scattered by the cylinder, and the associated potential functions,[17]

$$\phi_{2,\text{scattered}}=\sum_{n=0}^{\infty}D_nH_n(K_{2\perp}r)\cos(n\theta)\exp(i[\kappa_z z-\omega t]), \tag{12a}$$

$$\psi_{2,\text{scattered}}=\sum_{n=0}^{\infty}E_nH_n(k_{2\perp}r)\cos(n\theta)\exp(i[\kappa_z z-\omega t]), \tag{12b}$$

$$\chi_{2,\text{scattered}}=\sum_{n=0}^{\infty}F_nH_n(k_{2\perp}r)\sin(n\theta)\exp(i[\kappa_z z-\omega t]), \tag{12c}$$

where H_n represents the Hankel function of the first kind. [Although Hankel functions of both the first and second kind satisfy the wave equation in cylindrical coordinates, only the former are consistent with the condition of an outwardly propagating wave, scattered from the rod.[27] It is noted that the expressions of Eq. (6) for the wave potentials *inside* the rod were given in terms of Bessel functions of the first kind, J_n, because Hankel functions become infinite as their argument approaches zero.] The total displacement in material 2 can now be obtained by summing the incident and scattered components,

$$u_2=u_{2,\text{incident}}+u_{2,\text{scattered}}, \tag{13}$$

and making the appropriate substitutions from Eqs. (1)–(4) and (12).

B. Boundary conditions

There are six boundary conditions related to the continuity of stress and displacement at the cylinder/matrix interface,

$$[u_r,u_\theta,u_z,\sigma_{rr},\sigma_{r\theta},\sigma_{rz}]_1=[u_r,u_\theta,u_z,\sigma_{rr},\sigma_{r\theta},\sigma_{rz}]_2 \quad \text{at } r=a, \tag{14}$$

where expressions for the stress at any point can be derived from the displacement field determined according to Eq. (1), plus the strain-displacement relationships and Hooke's Law. In matrix format, Eq. (14) can be rewritten as

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16}\\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26}\\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36}\\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46}\\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56}\\ a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & a_{66} \end{bmatrix}\begin{bmatrix} A_n\\ B_n\\ C_n\\ D_n\\ E_n\\ F_n \end{bmatrix}=\begin{bmatrix} b_1\\ b_2\\ b_3\\ b_4\\ b_5\\ b_6 \end{bmatrix}, \tag{15}$$

where the a_{ij} and b_i coefficients are given in the Appendix. Solution of Eq. (15) followed by substitution into Eqs. (1), (5), and (12), allows the calculation of the normalized spectrum of scattered waves, also known as the backscattered form function f_∞. For the case of incident compression waves, it is given by

$$f_\infty(K_2a)=\sum_n\left(2/\sqrt{i\pi K_2a}\right)(-1)^n\epsilon_nQ_n, \tag{16a}$$

and for incident shear waves (of either r-z or r-θ polarization) by

$$f_\infty(k_2a)=\sum_n\left(2/\sqrt{i\pi k_2a}\right)(-1)^n\epsilon_{sn}Q_n. \tag{16b}$$

The parameter Q_n depends on the type of scattered wave of interest, as defined in Eq. (12): D_n, E_n, or F_n, corresponding to scattered compression waves, r-z polarized shear waves, or r-θ polarized shear waves, respectively.

TABLE I. Properties of isotropic materials.

Material	Density (kg/m^3)	Lame constants λ (GPa)	Lame constants μ (GPa)	Bulk modulus (GPa)
Water	1000		0	2.2
Stainless steel[a]	9449	140.2	93.8	
Epoxy[b]	1129	3.9	2.0	

[a]From Ref. 21.
[b]From Ref. 16.

II. NUMERICAL SIMULATION

The MATLAB software library was selected to solve Eq. (15) for any user-specified array of input parameters: elastic constants of matrix and cylinder, cylinder radius, incidence angle α, frequency span, and frequency step. One may also specify the mode and polarization of both the incident wave and scattered wave of interest: compression, shear polarized in the r-z plane, or shear polarized in the transverse (r-θ) plane. A graphical user interface (GUI) is used for managing the data input and results. The program is implemented on a Pentium II processor operating at 350 MHz. Partial verification of this new model is possible by comparing the results of a restricted class of test cases with those published in the literature; two such examples are indicated in the following paragraphs. A rigorous verification of the new formulation by this route is not possible, as there are limited published results corresponding to the selection of wave modes, incidence angles, and material properties encompassed by Eq. (15).

The first test case is that of an axially polarized shear wave normally incident on a 0.37-mm radius stainless steel wire embedded in an isotropic epoxy matrix. Elastic properties of the materials are given in Table I. The form function f_∞ of the scattered axially polarized shear wave, indicated by the solid line in Fig. 2, is identical to that obtained by Beattie *et al.*,[16] although their formulation did not include the options of a nonzero angle of incidence, nor that of a transversely isotropic rod. The major resonances correspond to an integral number $n=1,2,3$ of wavelengths of the axially polarized waves encircling the boundary of the cylinder; the integer n for each resonance was identified via a decomposition of the normal-mode expansion series of Eq. (12) into its constituent parts. A higher order ''breathing mode,'' corresponding to $n=0$, is also observed near $ka=9$.

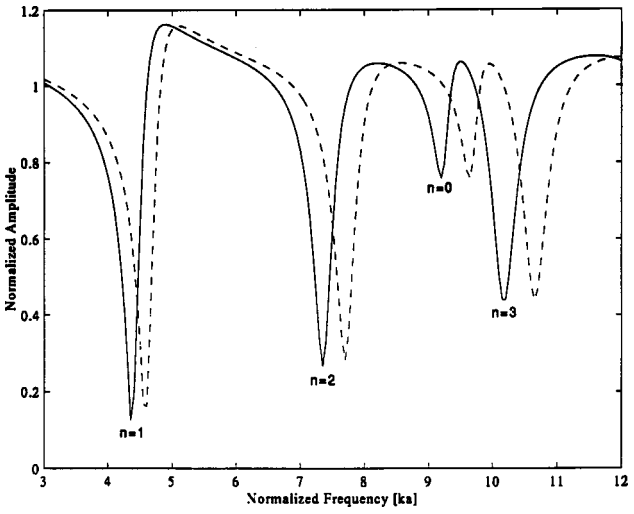


FIG. 2. Form function of the backscattered axially polarized shear wave, resulting from an axially polarized shear wave incident at $\alpha=0$ deg on a cylinder of radius 0.37 mm. The cylinder is embedded in an isotropic epoxy matrix. Solid line: isotropic stainless steel cylinder; dashed line: stainless steel cylinder, but with a 10% increase in $c_{44}(=c_{55})$, with all other stiffness constants left unchanged.

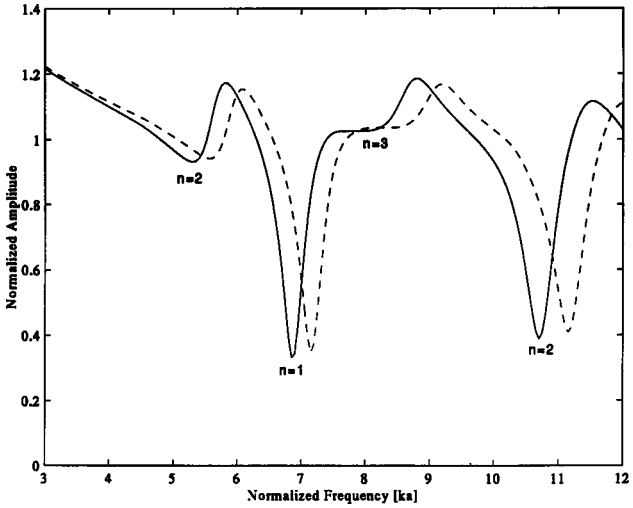


FIG. 3. Form function of backscattered shear waves polarized in the r-θ plane, resulting from r-θ polarized shear wave incident at $\alpha=0$ deg. The cylinder is embedded in an isotropic epoxy matrix. Solid line: Isotropic stainless steel cylinder; dashed line: Stainless steel cylinder, but with a 10% increase in c_{66}. [A corresponding adjustment was also made to the value of c_{11}, according to the relation $c_{66}=(c_{11}-c_{12})/2$.]

Also shown in Fig. 2 is the effect on f_∞ of switching from an isotropic to a transversely isotropic rod by increasing the shear modulus $c_{44}(=c_{55})$ by 10%, while leaving all other stiffness constants unchanged. This would have the effect of increasing the velocity of axially polarized bulk shear waves by about 5%; a corresponding shift in the resonance frequencies is noted in Fig. 2.

Figure 3 show the form function for the identical rod/matrix system of Fig. 2, but corresponding to incident and scattered shear waves polarized in the r-θ plane, with $\alpha=0$. Mode conversion leads to the presence of both compressional and shear components inside the rod, although axially polarized shear components (corresponding to the potential ψ) are absent. Within the frequency range spanned in Fig. 3, two pronounced resonances are evident ($n=1,2$), corresponding to waves encircling the rod along the rod-epoxy interface. The resonance frequencies f_R are closely linked to the transverse shear modulus c_{66}; this relationship leads to

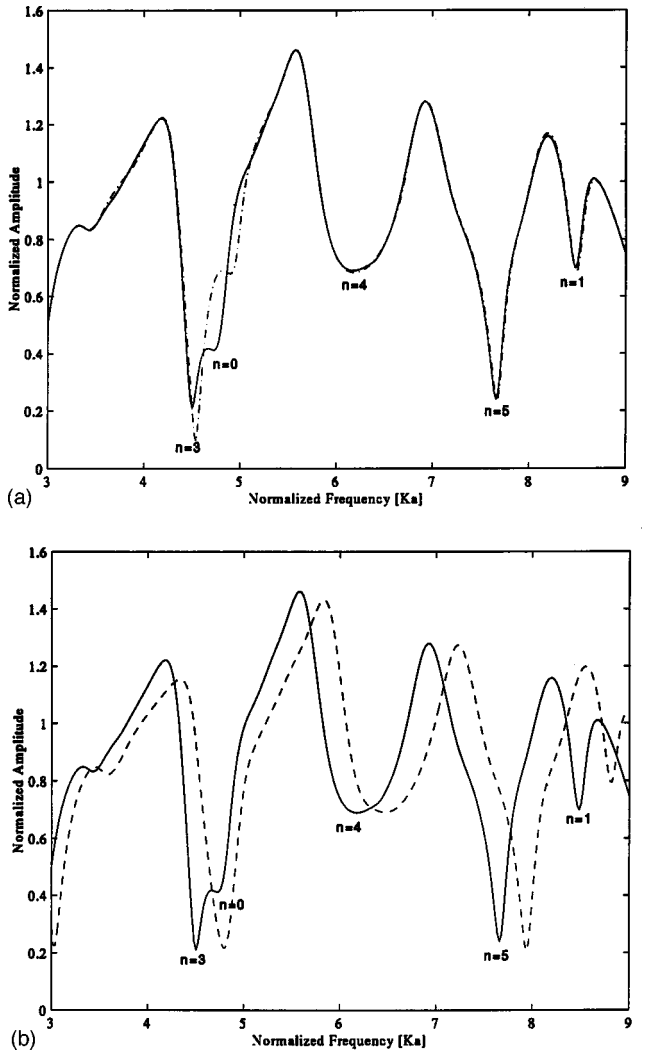


FIG. 4. Form function of the compression wave scattered at $\theta=180$ deg, resulting from a plane compression wave incident at an angle $\alpha=3$ deg. The cylinder is embedded in an isotropic epoxy matrix. Only one of the resonances, corresponding to $n=0$ corresponds to a guided mode. (a) Compression wave scattered at $\theta=180$ deg; Solid line: isotropic stainless steel cylinder; dashed line: stainless steel cylinder with a 10% increase in c_{44} $(=c_{55})$, with all other stiffness constants left unchanged. (b) Compression wave scattered at $\theta=180$ deg: Solid line: isotropic stainless steel cylinder; dashed line: stainless steel cylinder with a 10% increase in c_{66}. [A corresponding adjustment was also made to the value of c_{11}, according to the relation $c_{66}=(c_{11}-c_{12})/2$.]

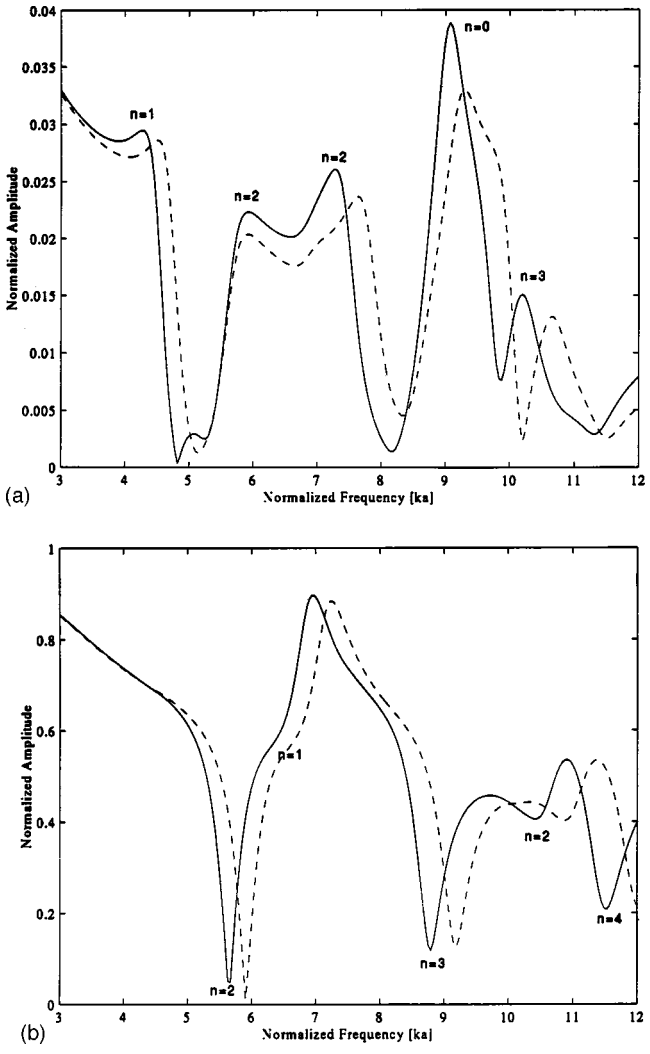


FIG. 5. Form function corresponding to scattered shear waves, resulting from a plane compression wave incident at an angle $\alpha=3$ deg. The cylinder is embedded in an isotropic epoxy matrix. (a) Axially polarized shear waves scattered at $\theta=180$ deg: Solid line: isotropic stainless steel cylinder; dashed line: stainless steel cylinder with a 10% increase in $c_{44}(=c_{55})$, with all other stiffness constants left unchanged. (b) Transversely polarized shear wave scattered at $\theta=180$ deg: Solid line: isotropic stainless steel cylinder; dashed line: stainless steel cylinder with a 10% increase in c_{66}. [A corresponding adjustment was also made to the value of c_{11}, according to the relation $c_{66}=(c_{11}-c_{12})/2$.]

the expected positive correlation between f_R and c_{66} for both resonances, as shown in Fig. 3.

Figure 4 features the same isotropic rod/matrix system as in Figs. 2 and 3, but now for a plane compression wave incident at an angle $\alpha=3°$. This incident wave leads to scattered compression waves, and shear waves with components polarized along both the z-axis and in the r-θ planes. Figure 4(a) shows f_∞ corresponding to compression waves backscattered at $\theta=180$ deg. Due to the small value of the incidence angle α, the resonances are primarily associated with waves encircling the rod, as was the case in Figs. 2 and 3. Only one axially guided mode is observed, corresponding to $n=0$, at $ka\approx4.8$. The differing sensitivities of these two types of resonances to perturbations in the elastic constants of the rod are quite striking: a 10% increase in $c_{44}(=c_{55})$, shown by the dashed line in Fig. 4(a) leads to an upward shift in f_R for the guided mode resonance, which is dominated by shearing motion in the θ-z and r-z planes. However, there is no shift in f_R corresponding to any of the other resonances, which are dominated by strain in the r-θ plane.

In Fig. 4(b), it is seen that a 10% increase in c_{66} for the rod leads to an increase in f_R for all resonances, *except* for the guided mode. In fact, the guided mode resonance is no longer visible in the form function when c_{66} is increased, as the resonance dip is masked by the strong interfacial mode resonance corresponding to $n=3$.

Figure 5(a) shows the form function corresponding to axially polarized shear waves scattered from the cylinder, as a result of compression waves incident at an angle $\alpha=3$ deg. The small amplitude of f_∞ in Fig. 5(a) is noteworthy—only a small component of the molecular mo-

tion in the incident compression wave is polarized in the axial direction due to the small angle of incidence α. (There are considerable experimental difficulties in the measurement of form functions for large α values.) A second interesting feature of this graph is that the resonances appear as peaks, instead of the dips seen in Figs. 2–4; this is due to a phase matching between the energy leaked from the resonant mode and the main echo backscattered from the cylinder-matrix interface.[28,29]

It has been reported that the shearing components of the cylinder/matrix boundary conditions [Eq. (14)] may be particularly sensitive to imperfections in the interface.[30,31] A defective interface would therefore be most apparent in the form functions corresponding to modes featuring shearing action at the boundaries, and transmission of shear waves through the interface into the matrix. These considerations have formed the basis for NDE techniques to characterize fiber-reinforced composite materials, utilizing selected modes of axially guided waves. Consistent with this premise, it is seen in Fig. 5(a) that the form function is sensitive to $c_{44}(=c_{55})$, in a manner similar to that for the axially polarized shear wave normally incident on the stainless steel rod seen in Fig. 2. However, the relatively weak signal, as indicated by the vertical scale of Fig. 5(a), would greatly hinder the use of this particular setup in a NDE scheme for such a small incidence angle α.

The form function of Fig. 5(b) corresponding to scattered shear waves polarized in the r-θ plane has a higher amplitude than that of Fig. 5(a). The resonance frequencies are seen to have good sensitivity to perturbations in the value of c_{66} for the cylinder; experimental work is required to determine whether a weak interface correlates more closely with perturbations in c_{66} or in c_{44}. The contrast between Figs. 5(b) and 3 is striking: both feature the form function corresponding to scattered shear wave components in the r-θ plane. Although the shapes of the two form functions appear markedly different, closer examination reveals that the resonant frequencies in the two figures are very similar. It is primarily the oscillatory background components of the two figures that are so different; this background is particularly sensitive to the incidence angle α.

Figure 6(a) shows the form function corresponding to scattered compression waves, resulting from compression waves obliquely incident at an angle $\alpha=3°$ on a transversely isotropic rod immersed in water. [The horizontal axis is expressed in units of MHz, in order to facilitate comparison with Fig. 6(b)–(d).] The rod is a continuous-fiber aluminum matrix composite (CF-AMC) manufactured by 3M, with elastic properties given in Table II. The form function is identical to that reported earlier by Honarvar and Sinclair.[21] The spectrum features leaky pseudo-Rayleigh (R) resonances at $f=0.95$ and 1.45 MHz; whispering gallery (WG) modes at $f=0.8$ and 1.3 MHz; and a leaky axially guided (GD) mode at $f=1.2$ MHz. These are all superimposed on an oscillating background originating from Franz and Scholte-Stoneley waves in the fluid encircling the rod. It should be noted that the water imparts minimal loading on the rod, such that the resonances appearing in Fig. 6(a) represent close to free modes of vibration.

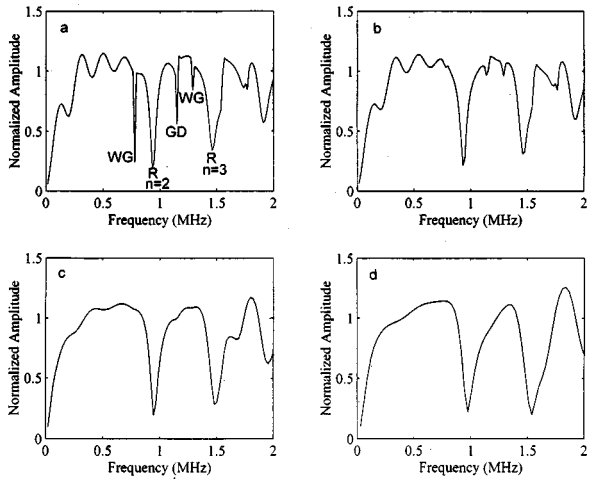


FIG. 6. Form functions corresponding to backscattered compression waves, resulting from a compression wave obliquely incident at an angle $\alpha=3$ deg on a transversely isotropic CF-AMC rod. The moduli of the surrounding medium are increased in steps from graph (a) to graph (d), where: For graph (a): Bulk modulus=2.2 GPa; $\mu=0$ (corresponding to water); for graph (d): $\lambda=3.9$ GPa; $\mu=2.0$ GPa (corresponding to epoxy).

Figure 6(b) to (d) show the effects of step increases in the elastic constants of the bulk medium surrounding the rod: from values corresponding to water in Fig. 6(a), to those describing epoxy in Fig. 6(d). The oscillatory background, whispering gallery modes, and guided mode are all suppressed as the elastic constants of the matrix are increased in the sequence of simulations. The pseudo-Rayleigh resonances become associated with interfacial modes, although their resonant frequencies are shifted only to a marginal extent from their values in Fig. 6(a). The small size of this shift is believed to be due to the relatively low acoustic impedance of the epoxy compared to the CF-AMC. In the limit of a large radius-to-wavelength ratio, these interfacial modes are associated with leaky pseudo-Stoneley waves,[32] and have been recommended for NDE of interfacial properties of adhesive bonds or interference-fit fasteners.[30,31]

The interfacial waves existing on the curved rod-matrix interface are dispersive. Lee and Corbly[31] have determined dispersion curves for axially propagating Stoneley waves on a cylindrical interface; however, the experimental constraint of relatively small incidence angles α restricts our interest to wave vectors oriented primarily in the r-θ plane. It would be possible to construct dispersion curves for these interfacial waves by considering the family of resonances over a large frequency range, in a manner similar to that used by Flax and Uberall for spherical inclusions embedded in a solid matrix.[13]

III. CONCLUSION

A normal-mode expansion solution has been presented for the wave scattered by a transversely isotropic cylinder embedded in a solid elastic medium; the formulation accommodates incident plane waves of compression mode, axially polarized shear, or r-θ polarized shear, at any angle of incidence α with respect to the axis of the rod. The geometry is a good model for a composite material featuring a reinforcing fiber encased in an elastic matrix. Results from the new

TABLE II. Properties of CF-AMC rod.[a]

Density (kg/m^3)	c_{11} (GPa)	c_{33} (GPa)	c_{44} (GPa)	c_{13} (GPa)	c_{12} (GPa)	c_{66} (GPa) $=(c_{11}-c_{12})/2$
3220	179	309	50	95	84	47.7

[a]Elastic constants estimated to within ±3%.

formulation indicate several resonances that are sensitive to perturbations in the elastic constants of the cylinder. Leaky pseudo-Stoneley waves that travel on the matrix-cylinder interface lead to resonances in the scattered wave spectrum that would be good indicators of deficiencies in the integrity of the interface. A Regge-pole plot of these resonances would lead to dispersion curves for these waves, as a function of the wavelength-to-radius ratio.

To adapt the formulation to the nondestructive evaluation of fiber-reinforced composites, it would need to be expanded to accommodate multiple fibers in the matrix. Preliminary work in this direction has been conducted by other researchers, e.g., Ref. 33, for waves normally incident on a random distribution of fibers in a solid matrix. In addition, it would be advantageous to address the experimental difficulties that preclude the use of large incidence angles for the planer wave impinging on the embedded cylinder.

ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support of the Natural Sciences and Engineering Research Council (NSERC) of Canada.

APPENDIX

Elements of the $\{a_{ij}\}$ matrix of Eq. (15) are as follows:

$$a_{11}=(1+iaq_1\kappa_z)[nJ_n(s_1a)-s_1aJ_{n+1}(s_1a)],$$

$$a_{12}=(q_2+ia\kappa_z)[nJ_n(s_2a)-s_2aJ_{n+1}(s_2a)],$$

$$a_{13}=nJ_n(s_3a),$$

$$a_{14}=-nH_n(K_{2\perp}a)+K_{2\perp}aH_{n+1}(K_{2\perp}a),$$

$$a_{15}=-ia\kappa_z[nH_n(k_{2\perp}a)-k_{2\perp}aH_{n+1}(k_{2\perp}a)],$$

$$a_{16}=-nH_n(k_{2\perp}a),$$

$$a_{21}=(-n-ina\kappa_zq_1)J_n(s_1a),$$

$$a_{22}=(-nq_2-ina\kappa_z)J_n(s_2a),$$

$$a_{23}=-nJ_n(s_3a)+s_3aJ_{n+1}(s_3a),$$

$$a_{24}=nH_n(K_{2\perp}a),$$

$$a_{25}=ina\kappa_zH_n(k_{2\perp}r),$$

$$a_{26}=nH_n(k_{2\perp}a)-k_{2\perp}aH_{n+1}(k_{2\perp}a),$$

$$a_{31}=(as_1^2q_1+i\kappa_z)J_n(s_1a),$$

$$a_{32}=(as_2^2+i\kappa_zq_2)J_n(s_2a),\quad a_{33}=0,$$

$$a_{34}=-i\kappa_zH_n(K_{2\perp}a),$$

$$a_{35}=-ak_{2\perp}^2H_n(k_{2\perp}a),\quad a_{36}=0,$$

$$a_{41}=[c_{11}^{(1)}+ia\kappa_zq_1(c_{11}^{(1)}-c_{13}^{(1)})][(n^2-n-s_1^2a^2)J_n(s_1a)+s_1aJ_{n+1}(s_1a)]+[c_{12}^{(1)}+(c_{12}^{(1)}-c_{13}^{(1)})ia\kappa_zq_1]\times[nJ_n(s_1a)-s_1aJ_{n+1}(s_1a)]+[-c_{13}^{(1)}\kappa_z^2a^2-c_{12}^{(1)}n^2+(c_{13}^{(1)}-c_{12}^{(1)})ian^2q_1\kappa_z]J_n(s_1a),$$

$$a_{42}=[c_{11}^{(1)}q_2+ia\kappa_z(c_{11}^{(1)}-c_{13}^{(1)})][(n^2-n-s_2^2a^2)J_n(s_2a)+s_2aJ_{n+1}(s_2a)]+[c_{12}^{(1)}q_2+(c_{12}^{(1)}-c_{13}^{(1)})ia\kappa_z]\times[nJ_n(s_2a)-s_2aJ_{n+1}(s_2a)]+[-c_{13}^{(1)}\kappa_z^2q_2a^2-c_{12}^{(1)}n^2q_2+(c_{13}^{(1)}-c_{12}^{(1)})ian^2\kappa_z]J_n(s_2a),$$

$$a_{43}=(c_{11}^{(1)}-c_{12}^{(1)})n[(n-1)J_n(s_3a)-s_3aJ_{n+1}(s_3a)],$$

$$a_{44}=-[c_{11}^{(2)}(n^2-n-K_{2\perp}^2a^2)+c_{12}^{(2)}n(1-n)-c_{13}^{(2)}\kappa_z^2a^2]H_n(K_{2\perp}a)+(c_{12}^{(2)}-c_{11}^{(2)})K_{2\perp}aH_{n+1}(K_{2\perp}a),$$

$$a_{45}=-ia\kappa_z(c_{11}^{(2)}-c_{13}^{(2)})[(n^2-n-k_{2\perp}^2a^2)H_n(k_{2\perp}a)+k_{2\perp}aH_{n+1}(k_{2\perp}a)]-(c_{12}^{(2)}-c_{13}^{(2)})ia\kappa_z[nH_n(k_{2\perp}a)-k_{2\perp}aH_{n+1}(k_{2\perp}a)]-(c_{13}^2-c_{12}^2)ian^2\kappa_zH_n(k_{2\perp}a),$$

$$a_{46}=-(c_{11}^{(2)}-c_{12}^{(2)})n[(n-1)H_n(k_{2\perp}a)-k_{2\perp}aH_{n+1}(k_{2\perp}a)],$$

$$a_{51}=n(c_{11}^{(1)}-c_{12}^{(1)})(1+iaq_1\kappa_z)[(1-n)J_n(s_1a)+s_1aJ_{n+1}(s_1a)],$$

$$a_{52}=n(c_{11}^{(1)}-c_{12}^{(1)})(q_2+ia\kappa_z)[(1-n)J_n(s_2a)+s_2aJ_{n+1}(s_2a)],$$

$$a_{53}=\frac{(c_{11}^{(1)}-c_{12}^{(1)})}{2}\{[s_3^2a^2-2n(n-1)]J_n(s_3a)-2s_3aJ_{n+1}(s_3a)\},$$

$$a_{54}=(c_{11}^{(2)}-c_{12}^{(2)})[H_n(K_{2\perp}a)n(n-1)-nK_{2\perp}aH_{n+1}(K_{2\perp}a)],$$

$$a_{55}=(c_{11}^{(2)}-c_{12}^{(2)})[H_n(k_{2\perp}a)ia\kappa_zn(n-1)-in\kappa_zk_{2\perp}a^2H_{n+1}(k_{2\perp}a)],$$

$$a_{56}=\frac{(c_{11}^{(2)}-c_{12}^{(2)})}{2}[H_n(k_{2\perp}a)(2n(n-1)-k_{2\perp}^2a^2)$$
$$+2k_{2\perp}aH_{n+1}(k_{2\perp}a)],$$
$$a_{61}=c_{44}^{(1)}[q_1(s_1^2a-\kappa_z^2a)+2i\kappa_z]$$
$$\times[nJ_n(s_1a)-s_1aJ_{n+1}(s_1a)],$$
$$a_{62}=c_{44}^{(1)}(s_2^2a-a\kappa_z^2+2i\kappa_zq_2)$$
$$\times[nJ_n(s_2a)-s_2aJ_{n+1}(s_2a)],$$
$$a_{63}=c_{44}^{(1)}in\kappa_zJ_n(s_3a),$$
$$a_{64}=c_{44}^{(2)}[-2in\kappa_zH_n(K_{2\perp}a)$$
$$+2i\kappa_zK_{2\perp}aH_{n+1}(K_{2\perp}a)],$$
$$a_{65}=c_{44}^{(2)}(\kappa_z^2-k_{2\perp}^2)[anH_n(k_{2\perp}a)$$
$$-k_{2\perp}a^2H_{n+1}(k_{2\perp}a)],$$
$$a_{66}=-c_{44}^{(2)}in\kappa_zH_n(k_{2\perp}a).$$

Elements of the $\{b_i\}$ vector of Eq. (15) depend on the type of wave incident on the cylinder. For an incident compression wave, they are as follows:

$$b_1=\epsilon_n(i)^n[nJ_n(K_{2\perp}a)-K_{2\perp}aJ_{n+1}(K_{2\perp}a)],$$
$$b_2=-n\epsilon_n(i)^nJ_n(K_{2\perp}a),$$
$$b_3=\epsilon_n(i)^{n+1}\kappa_zJ_n(K_{2\perp}a),$$
$$b_4=\epsilon_n(i)^n[c_{11}^{(2)}(n^2-n-K_{2\perp}^2a^2)+c_{12}^{(2)}n(1-n)$$
$$-c_{13}^{(2)}K_{2\perp}^2a^2]J_n(K_{2\perp}a)$$
$$-\epsilon_n(i)^n(c_{12}^{(2)}-c_{11}^{(2)})K_{2\perp}aJ_{n+1}(K_{2\perp}a),$$
$$b_5=\epsilon_n(i)^n(c_{11}^{(2)}-c_{12}^{(2)})[J_n(K_{2\perp}a)(n-n^2)$$
$$+nK_{2\perp}aJ_{n+1}(K_{2\perp}a)],$$
$$b_6=c_{44}^{(2)}2\epsilon_n(i)^{n+1}\kappa_z[nJ_n(K_{2\perp}a)-K_{2\perp}aJ_{n+1}(K_{2\perp}a)].$$

For an incident shear wave, polarized in the r-θ plane,

$$b_1=\epsilon_n(i)^nnJ_n(k_{2\perp}a),$$
$$b_2=-\epsilon_n(i)^n(nJ_n(k_{2\perp}a)-k_{2\perp}aJ_{n+1}(k_{2\perp}a)),$$
$$b^3=0,$$
$$b_4=\epsilon_n(i)^n(c_{11}^{(2)}-c_{12}^{(2)})n[(n-1)J_n(k_{2\perp}a)$$
$$-k_{2\perp}aJ_{n+1}(k_{2\perp}a)],$$
$$b_5=\epsilon_n(i)^n\frac{(c_{12}^{(2)}-c_{11}^{(2)})}{2}[J_n(k_{2\perp}a)(2n(n-1)-k_{2\perp}^2a^2)$$
$$+2k_{2\perp}aJ_{n+1}(k_{2\perp}a)],$$
$$b_6=\epsilon_n(i)^nc_{44}^{(2)}in\kappa_zJ_n(k_{2\perp}a).$$

For an incident shear wave, polarized in the r-z plane,

$$b_1=\epsilon_n(i)^{n+1}a\kappa_z[nJ_n(k_{2\perp}a)-k_{2\perp}aJ_{n+1}(k_{2\perp}a)],$$
$$b_2=-\epsilon_n(i)^{n+1}na\kappa_zJ_n(k_{2\perp}a),$$
$$b_3=\epsilon_n(i)^nak_{2\perp}^2J_n(k_{2\perp}a),$$
$$b_4=\epsilon_n(i)^{n+1}a\kappa_z(c_{11}^{(2)}-c_{13}^{(2)})[(n^2-n-k_{2\perp}^2a^2)J_n(k_{2\perp}a)$$
$$+k_{2\perp}aJ_{n+1}(k_{2\perp}a)]-(c_{13}^{(2)}-c_{12}^{(2)})\epsilon_n(i)^{n+1}$$
$$\times a\kappa_z[nJ_n(k_{2\perp}a)-k_{2\perp}aJ_{n+1}(k_{2\perp}a)]$$
$$-(c_{12}^{(2)}-c_{13}^{(2)})\epsilon_n(i)^{n+1}an^2\kappa_zJ_n(k_{2\perp}a),$$
$$b_5=(c_{12}^{(2)}-c_{11}^{(2)})\epsilon_n(i)^{n+1}[J_n(k_{2\perp}a)a\kappa_zn(n-1)$$
$$-n\kappa_zk_{2\perp}a^2J_{n+1}(k_{2\perp}a)],$$
$$b_6=c_{44}^{(2)}(k_{2\perp}^2-\kappa_z^2)\epsilon_n(i)^n[anJ_n(k_{2\perp}a)$$
$$-k_{2\perp}a^2J_{n+1}(k_{2\perp}a)].$$

[1] J. J. Faran, Jr., J. Acoust. Soc. Am. **23**, 405 (1951).
[2] R. M. White, J. Acoust. Soc. Am. **30**, 771 (1958).
[3] L. Flax and W. G. Neubauer, J. Acoust. Soc. Am. **64**, 675 (1978).
[4] A. Nagl, H. Uberall, P. P. Delsanto, J. D. Almar, and E. Rosario, Wave Motion **5**, 235 (1983).
[5] N. D. Veksler, Acustica **71**, 111 (1990).
[6] J. M. Conoir, P. Rembert, O. Lenoir, and J. L. Izbicki, J. Acoust. Soc. Am. **93**, 1300 (1993).
[7] H. Uberall, ''Surface waves in acoustics,'' in *Physical Acoustics*, edited by W. P. Mason and R. N. Thurston (Academic, New York, 1973), Vol. 10, Chap. 1.
[8] Y. H. Pao and C. C. Mow, *Diffraction of Elastic Waves and Dynamic Stress Concentrations* (Crane, Russack, New York, 1973).
[9] G. C. Gaunaurd, Appl. Mech. Rev. **42**, 143 (1989).
[10] H. Uberall, *Acoustic Resonance Scattering* (Gordon and Breach Science, Philadelphia, 1992).
[11] N. D. Veksler, *Resonance Acoustic Spectroscopy*, Springer Series on Wave Phenomena (Springer-Verlag, Berlin, 1993).
[12] T. S. Lewis and D. W. Kraft, J. Appl. Phys. **47**, 1265 (1976).
[13] L. Flax and H. Uberall, J. Acoust. Soc. Am. **67**, 1432 (1980).
[14] J. P. Sessarego, J. Sageloli, R. Guillermin, and H. Uberall, J. Acoust. Soc. Am. **104**, 2836 (1998).
[15] T. M. Hsieh, E. A. Lindgren, and M. Rosen, Ultrasonics **29**, 38 (1991).
[16] P. Beattie, R. C. Chivers, and L. W. Anson, J. Acoust. Soc. Am. **94**, No. 6, 3421 (1993).
[17] A. N. Sinclair and R. C. Addison, Jr., J. Acoust. Soc. Am. **94**, No. 2; Pt. 1, 1126 (1993).
[18] A. N. Sinclair, F. Honarvar, M. R. Piggott, and M. V. Ravichandran, *Proc. Intl. Conf. on Composite Materials and Energy* (Technomic, Lancaster, 1995).
[19] W. Huang, S. Brisuda, and S. I. Rokhlin, J. Acoust. Soc. Am. **97**, 807 (1995).
[20] W. Huang, Y. J. Wang, and S. I. Rokhlin, J. Acoust. Soc. Am. **99**, 2742 (1996).
[21] F. Honarvar and A. N. Sinclair, Ultrasonics **36**, 845 (1998).
[22] I. Mirsky, J. Acoust. Soc. Am. **37**, 1016 (1965).
[23] I. Mirsky, J. Acoust. Soc. Am. **37**, 1022 (1965).
[24] M. de Billy, J. Acoust. Soc. Am. **97**, 852 (1995).
[25] M. de Billy, Acustica **81**, 281 (1995).
[26] F. Honarvar and A. N. Sinclair, J. Acoust. Soc. Am. **100**, 57 (1996).
[27] F. Honarvar, ''Nondestructive evaluation of cylindrical components by resonance acoustic spectroscopy,'' Ph.D. thesis, University of Toronto, 1997.
[28] S. K. Numrich, W. E. Howell, J. V. Subrahmanyam, and H. Uberall, J. Acoust. Soc. Am. **80**, 1161 (1986).
[29] C. C. H. Guyott and P. Cawley, J. Acoust. Soc. Am. **83**, 632 (1988).
[30] T. M. Hsieh, E. S. Lindgren, and M. Rosen, Ultrasonics **29**, 38 (1991).
[31] S. A. Lee and D. M. Corbly, IEEE Trans. Sonics Ultrason. **SU-24**, 206 (1977).
[32] W. Maurice Ewing, W. S. Jardetzky, and F. Press, *Elastic Waves in Layered Media* (McGraw–Hill, New York, 1957).
[33] S. K. Bose and A. K. Mal, J. Met. Phys. Sol. **22**, 217 (1974).

The radiating near field of a circular normal transducer of arbitrary apodization on an elastic half-space

Dmitri Gridin[a)]
Centre for Waves and Fields, School of EEIE, South Bank University, 103 Borough Road, London SE1 0AA, United Kingdom

(Received 10 October 1998; accepted for publication 18 May 1999)

A new fast method is developed for simulating the propagation of pulses radiated by a circular normal transducer of arbitrary apodization into an isotropic and homogeneous elastic half-space. First, the model proposed in Fradkin, Kiselev and Krylova for a time-harmonic uniform transducer [J. Acoust. Soc. Am. **104**, 1178–1187 (1998)] is extended to the case of nonuniform load to obtain high-frequency asymptotics of the time-harmonic field. Then, the transient field is described by means of harmonic synthesis. The asymptotics elucidate the physics of the problem and give explicit dependence of the radiated waves on model parameters. The formulas are applicable in the radiating near field, that is the near field with the evanescent wave zone excluded. They involve in geometrical regions elementary and inside boundary layers, well-known special functions (Fresnel integral and Bessel functions). Unlike the uniform load case, the direct shear wave is present and the edge waves may be practically eliminated. The asymptotic code has been tested against an exact numerical solution when apodization is parabolic. It has proved to be hundreds of times faster but in many realistic cases the accuracy does not suffer much.
[S0001-4966(99)00709-2]

PACS numbers: 43.20.Bi, 43.20.Dk, 43.35.Zc [DEC]

INTRODUCTION

Modeling the transient fields radiated by surface loads into an elastic half-space is a first step towards the system model of ultrasonic nondestructive evaluation (NDE) of industrial materials. For the past several decades the propagation of ultrasonic pulses radiated by circular piston-type transducers into elastic solids has been intensively studied by various full numerical schemes and approximate/asymptotic methods (see, e.g., Lhémery, 1994; Gridin and Fradkin, 1998; and references therein). The problem formulation traditionally involves the prescribed stresses on the loaded surface. These boundary conditions are reasonable, because the stiffnesses of the transducer and underlying solid are comparable, and have been confirmed by experiments (Weight, 1987; Djelouah and Baboux, 1992). Usually assumed *uniform*, the actual stress distribution is not known, and it might be useful to have an idea of the outgoing wavefield when the distribution is *nonuniform*. Of course, the real boundary conditions might be even more involved. However, in this paper we will use the traditional prescribed stresses.

A uniform circular normal transducer radiates a wavefield of a complicated structure consisting of the direct compressional and edge compressional and shear waves as well as head waves (e.g., Weight, 1987). To simplify the structure, nonuniform transducers have been designed and developed (e.g., Weight, 1984). These include plane wave only (PWO) and edge wave only (EWO) transducers. PWO transducers are designed to reduce the diffraction effects, that is, the amplitude of the edge waves. The reduction is achieved by introducing the nonuniform pressure profile which is strongest in the center and falls away to zero at the edge. By contrast, the pressure exerted by an EWO transducer is strongest at the edge and negligible elsewhere. This way most of the plane-wave component is eliminated.

It is well known that the *far-field* pattern of a nonuniform transducer is determined by the spatial Fourier (or, in the case of axial symmetry, Hankel) transform of the apodization (e.g., Morse and Ingard, 1968). In the past decade several researchers have studied the transient *near* field radiated by a nonuniform circular load into a solid. Harris (1988) has considered the Gaussian distribution of normal tractions. He used the parabolic approximation to the wave equation to find asymptotic expressions for the plane waves. Bresse and Hutchins (1989) have applied the Cagniard–de Hoop method and considered the case of pressure distribution falling off parabolically to zero at the transducer edge. Their impulse response has been expressed in terms of double integrals, and, thus, their code for modelling pulse response requires triple integration. Moreover, the Cagniard-de Hoop method can be applied to some but not all apodizations. McNab *et al.* (1989) have developed a discrete numerical procedure based on integrating point load solutions, and thus it is suitable for dealing with arbitrary apodized loads. The wavefields radiated by PWO and EWO transducers have been studied by Stacey and Weight (1994), both experimentally and using a finite-difference code. Also, the approximate Lhémery's model (1994) has been extended to deal with nonuniform transducers by Lhémery and Stacey (1995). In their computations these authors used a superposition of 20 annular transducers in the region where the profile is nonuniform. Ushakov and Danilov (1997) have studied the directivity patterns of nonuniform rectangular transducers both experimentally and by applying the saddle point

a)Electronic mail: gridind@sbu.ac.uk

method to the integral representation of the displacement. All researchers have noticed that compared to the uniform case a nonuniform profile can lead to significant changes in the structure of the outgoing field.

Recently, Fradkin *et al.* (1998) have developed the so-called two-tier approach to description of the radiating near field of a time-harmonic uniform circular transducer based on modern diffraction theory and the asymptotic approximation of integrals: First, they had obtained the far-field asymptotics of a point source acting on the surface of an elastic half-space. Then they have integrated these asymptotics over the transducer surface and found the radiating near-field asymptotics of the transducer, inside both the geometrical regions and the transition zones (penumbral and axial regions). The high-frequency asymptotic method combined with the fast Fourier transform has been shown to be a hundred to a thousand times faster but practically as accurate as an exact numerical solution (Gridin and Fradkin, 1998). In the present paper, the method is extended to circular transducers of arbitrary apodization. The paper is organized as follows: First, the high-frequency asymptotics for the elastic field of a time-harmonic nonuniform transducer are obtained. Then, for the case of parabolic apodization the asymptotic formulas are tested against an exact numerical solution. At the end of the paper, the outputs of the asymptotic and exact numerical codes for simulating the pulse propagation are compared.

I. TIME-HARMONIC NONUNIFORM LOAD

A. Formulation

Let us consider the time-harmonic motion of an isotropic and homogeneous elastic half-space caused by a normal circular load of radius l. The displacement field $\mathbf{u}$ inside the medium is described by the reduced elastodynamic equation

$$c_S^2\nabla\times(\nabla\times\mathbf{u})-c_P^2\nabla(\nabla\cdot\mathbf{u})-\omega^2\mathbf{u}=\mathbf{0}, \tag{1}$$

where c_P is a speed of P (also known as compressional or longitudinal) wave; c_S is a speed of S (also known as shear or transverse) wave, and ω is a circular frequency. The cylindrical coordinates (r,φ,z), with the origin at the center of the load and the z axis directed into the solid, are used below. The idealized boundary conditions caused by a normal circular load are

$$\sigma_{zr}|_{z=0}=0,$$
$$\sigma_{zz}|_{z=0}=\begin{cases}-P_0q(r)e^{-i\omega t}, & r\leqslant l;\\ 0, & r>l,\end{cases} \tag{2}$$

where P_0 has a dimension of pressure and $q(r)$ is dimensionless apodization. Here σ_{zr} is a tangential and σ_{zz}, a normal traction. The usual radiation condition at infinity is assumed. The common factor $\exp(-i\omega t)$ will be suppressed throughout the paper. Below, we consider the two classes of general apodization which are of most importance:

Class I: $q(r)$ is an arbitrary function, such that

$$q(l)\geqslant\varepsilon>0, \tag{3}$$

i.e., the pressure is not small at the transducer edge.

Class II: $q(r)$ is an arbitrary function, such that

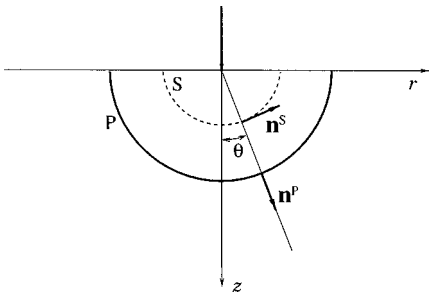


FIG. 1. Point source acting normally on the surface of an elastic half-space (head waves neglected). Solid line is the front of the *P*, and dashed line, of the *S* wave.

$$q(l)=0,\quad |q'(l)|\geqslant\varepsilon>0, \tag{4}$$

i.e., the pressure is zero of the first order at the transducer edge.

In both classes it is assumed that both $q(r)$ and $q'(r)$ are continuous, and $q'(0)=0$. The latter condition means that the apodization does not have a discontinuity of the first derivative at the transducer center. Note that the uniform and EWO transducers are instances of Class I, and the PWO transducers belong to Class II.

B. The two-tier approach

The two-tier approach introduced in Fradkin *et al.* (1998) is based on the Green's theorem. The first tier involves finding the far-field asymptotics of the Green's function of Lamb's problem, that is, of a point source acting normally on the surface of an isotropic and homogeneous elastic half-space. It is well known that in the far field the Green's function $\mathbf{u}^{(\text{point})}$ can be decomposed into P and S waves,

$$\mathbf{u}^{(\text{point})}\simeq\mathbf{u}^{P(\text{point})}+\mathbf{u}^{S(\text{point})}. \tag{5}$$

The leading terms of the corresponding far-field asymptotics are obtained under condition

$$k_\alpha z\gg 1,\quad \alpha=P,S, \tag{6}$$

and given by

$$\mathbf{u}_0^{\alpha\ (\text{point})}\simeq-\frac{1}{2\pi\varrho c_S^2}\frac{A^\alpha(\theta)}{s}e^{ik_\alpha s}\mathbf{n}^\alpha(\theta) \tag{7}$$

(e.g., Miller and Pursey, 1954). Here ϱ is material density, $k_\alpha=\omega/c_\alpha$ is a wavenumber and we have

$$s=\sqrt{r^2+z^2},\quad \sin\theta=r/s \tag{8}$$

(see Fig. 1). The directivities A^P and A^S and the unit displacement vectors are given in Appendix A.

The second tier involves integrating the above asymptotics over the loaded surface so that we have

$$\mathbf{u}\simeq P_0\int_0^{2\pi}\int_0^l q(r')\mathbf{u}^{(\text{point})}r'\,dr'\,d\varphi'. \tag{9}$$

Then the high-frequency asymptotics of this integral are evaluated under the near zone condition

$$k_P z\ll(k_P l)^2, \tag{10}$$

which implies that the integrand in (9) contains a rapidly oscillating factor.

Conditions (6) and (10) may be combined and relaxed to give

$$\lambda_P < z < l^2/\lambda_P, \tag{11}$$

where λ_P is the P wavelength. Condition (11) defines the so-called radiating near zone of a circular transducer, that is, the near zone with the evanescent wave region excluded. The fact that the frequencies have to be only *relatively* high, so that the above relaxation is allowed, has been confirmed by numerical experiments.

C. The high-frequency asymptotics of the *P* field

The P field is described by the integral

$$\mathbf{u}^P \simeq -\frac{P_0}{2\pi\varrho c_S^2}\int_0^{2\pi}\int_0^l q(r')\frac{A^P(\theta')}{s'}\,\tilde{\mathbf{n}}^P e^{ik_P s'} r'\,dr'\,d\varphi', \tag{12}$$

where $(r',\varphi',0)$ and (r,φ,z) are the coordinates of the point source and the observation point, respectively. The distance s' between these points and the corresponding angle θ' are given by

$$\begin{aligned} s' &= [r^2 + r'^2 - 2rr'\cos(\varphi-\varphi') + z^2]^{1/2}, \\ \sin\theta' &= [s'^2 - z^2]^{1/2}/s'. \end{aligned} \tag{13}$$

The unit vector $\tilde{\mathbf{n}}^P$ can be written as

$$\tilde{\mathbf{n}}^P = \frac{r'\sin(\varphi-\varphi')}{s'}\mathbf{e}_r + \frac{r - r'\cos(\varphi-\varphi')}{s'}\mathbf{e}_\varphi + \cos\theta'\,\mathbf{e}_z, \tag{14}$$

where the vectors $\mathbf{e}_r$ and $\mathbf{e}_\varphi$ are taken at the observation point.

The integrand in (12) contains a slowly varying amplitude and a rapidly oscillating exponential factor. It is well known that the main contributions to an integral of this type come from critical points. These include stationary points of the phase function, various types of critical boundary points and singular points of the amplitude function (e.g., Borovikov, 1994).

The amplitude function in (12) contains no singular points. Its phase function s' has one stationary point,

$$\begin{aligned} &(r',\varphi') = (r,\varphi), \quad 0 < r < l; \\ &r' = r, \quad r = 0, \end{aligned} \tag{15}$$

where it is at a minimum. For $r/l > 0$ there are two boundary critical points,

$$(r',\varphi') = (l,\varphi), \quad \text{and} \quad (r',\varphi') = (l,\varphi+\pi), \tag{16}$$

which are the nearest and the farthest edge points, respectively. Each type of a critical point gives rise to a particular type of wave. In practice, the discontinuities of the derivatives of $q(r)$ which are higher than the first order can be neglected, since they give asymptotic contributions of higher order than the phase stationary and boundary critical points.

According to the modern diffraction theory (e.g., Babich and Buldyrev, 1991; Borovikov and Kinber, 1994) a wave field of high frequency may be described in terms of the so-called geometrical regions separated by boundary layers. The geometrical regions are so called because they may be modeled efficiently using the concept of a ray. The field inside such regions may be represented in the form of a finite number of waves, each described by a *ray series*, that is, an asymptotic series in inverse powers of dimensionless wavenumber k,

$$\mathbf{u}(k,\mathbf{s}) = e^{ik\tau(\mathbf{s})}\sum_{n=0}^{\infty}\left(\frac{i}{k}\right)^{n+\nu} A_n(\mathbf{s}), \quad \nu = \text{const} \tag{17}$$

(e.g., Babich and Buldyrev, 1991). In (17) the eikonal (or phase function) $\tau(\mathbf{s})$ and the amplitude functions $A_n(\mathbf{s})$ all depend on the coordinates $\mathbf{s}$ of the observation point only. The ray-series representation breaks down inside boundary layers where the fields are described in terms of asymptotic series involving special functions rather than just inverse powers of dimensionless wavenumber. The wave fronts and boundary layers generated by a nonuniform circular transducer acting on the surface of an elastic half-space are shown in Fig. 2. Their description follows.

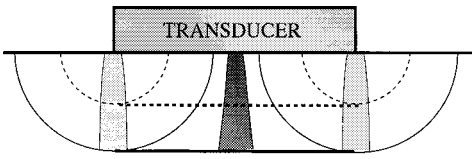


FIG. 2. Wave fronts and boundary layers underneath a circular nonuniform transducer acting normally on the surface of an elastic half-space (head waves neglected). Thick solid line—direct P, thin solid lines—edge P, thick dashed line—direct S, and thin dashed lines—edge S wave. Light-shaded areas—the penumbra, and dark-shaded area—the axial region.

1. Geometrical regions

The contribution of the stationary point (15) to the integral in (12) gives the direct P wave. Considering (12) as a repeated integral and using the standard formulas of the one-dimensional stationary phase method twice [e.g., Borovikov, 1994, Eqs. (1.10) and (1.11a)], we obtain the leading (or first order) term

$$\mathbf{u}_1^{P(\text{direct})} \simeq \overline{\mathbf{u}}^{P(\text{direct})} q(r), \quad r < l, \tag{18}$$

where $\overline{\mathbf{u}}^{P(\text{direct})}$ is the leading term for the uniform case, $q(r) = 1$, as given by Eq. (B1). It is important to note that in the latter case the leading term represents the exact solution. This point is crucial in allowing us to find the *second* order term for the direct P wave from Eq. (12) despite the fact that the integrand involves only the *leading* term of the Green's function. The trick is that when using formulas for the second order contribution of the stationary point [e.g., Borovikov, 1994, Eqs. (1.10) and (1.11b)], we should take into account only terms involving derivatives of $q(r)$. Only these terms are not compensated by the leading order contribution of the stationary point to the integral of the second term of the far-field asymptotics of the Green's function over the transducer surface. Thus, we obtain the components of the second order term as

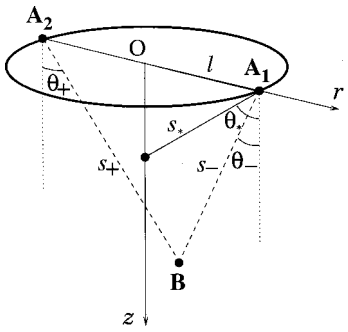


FIG. 3. The shortest ($\mathbf{A}_1\mathbf{B}$) and longest ($\mathbf{A}_2\mathbf{B}$) ray path from the transducer edge.

$$u_{z,2}^{P(\text{direct})} \simeq \overline{u}^{P(\text{direct})} \frac{i}{k_P} \frac{z}{2} \Delta_\perp q(r), \quad r<l, \tag{19}$$

and

$$u_{r,2}^{P(\text{direct})} \simeq \frac{P_0}{\varrho c_P^2 k_P^2} q'(r) e^{ik_P z} = -\frac{i}{k_P} \frac{\partial(u_1^{P(\text{direct})})}{\partial r}, \quad r<l, \tag{20}$$

where we have

$$\Delta_\perp = \frac{\partial^2}{\partial r'^2} + \frac{1}{r'} \frac{\partial}{\partial r'}. \tag{21}$$

Note that the amplitude in Eq. (19) satisfies the transport equation

$$2(\nabla\tau, \nabla A_1) + \Delta\tau A_1 = \Delta A_0 \tag{22}$$

[e.g., Borovikov and Kinber, 1994, Eq. (3.5)] which for $\tau = z$ reduces to

$$\frac{\partial A_1}{\partial z} = \frac{\Delta_\perp A_0}{2}. \tag{23}$$

It follows from Eq. (19) that the second order term of the ray series grows with the distance from the transducer. Such behavior is common in plane waves with the amplitude nonuniform on the initial wave-front (e.g., Borovikov and Kinber, 1994, Sec. 3.4.6). It is due to the interaction between the adjacent ray tubes which is otherwise known as the amplitude diffusion. Equation (20) gives the leading term of the transverse component of *P* wave as the transverse gradient of the leading term which is aligned with the direction of propagation [cf. Fradkin and Kiselev, 1997, Eq. (17)].

The contributions of the boundary critical points (16) to the integral in (12) give two arrivals of the edge *P* wave, one along the shortest and the other along the longest ray path (see Fig. 3). Using formulas for the contribution of a boundary point [e.g., Borovikov, 1994, Eqs. (1.3) and (1.6)], we obtain the corresponding leading terms as

$$\text{Class I:} \quad \mathbf{u}_\pm^{P(\text{edge})} \simeq \overline{\mathbf{u}}_\pm^{P(\text{edge})} q(l), \tag{24}$$

$$\text{Class II:} \quad \mathbf{u}_\pm^{P(\text{edge})} \simeq \overline{\mathbf{u}}_\pm^{P(\text{edge})} \frac{i}{k_P} \frac{s_\pm}{l \pm r} q'(l), \tag{25}$$

where $\overline{\mathbf{u}}_\pm^{P(\text{edge})}$ are the leading terms for the uniform case given by Eq. (B3). The plus sign relates to the edge wave arrival along the longest, the minus, along the shortest ray path. Note that for Class II the edge wave is an order smaller than for Class I.

2. Boundary layers

The edge wave radiated by a circular transducer focuses along the transducer axis. The surrounding region is called axial. From the physical point of view, inside this region the phase difference between the edge wave arrivals from any two edge points is small. From the mathematical point of view, this corresponds to the case of critical boundary points close to being degenerate (Fradkin *et al.*, 1998). Inside the axial region we obtain the leading term of the edge *P* wave as

$$\text{Class I:} \quad \mathbf{u}^{P(\text{axial})} \simeq \overline{\mathbf{u}}^{P(\text{axial})} q(l), \tag{26}$$

$$\text{Class II:} \quad \mathbf{u}^{P(\text{axial})} \simeq \overline{\mathbf{u}}^{P(\text{axial})} \frac{i}{k_P} \frac{s_*}{l} q'(l), \tag{27}$$

where $\overline{\mathbf{u}}^{P(\text{axial})}$ is the leading term for the uniform case given by Eq. (B5). The shape of the axial region is determined by the condition that the phase difference between the edge arrivals from the nearest and farthest points is approximately π, which is approximately equivalent to

$$k_P r \sin\theta_* \approx \frac{\pi}{2}. \tag{28}$$

The transition zone between the main beam and the geometrical shadow is known as the penumbral boundary layer. Inside this zone the phase difference between the direct and edge *P* arrival along the shortest ray path is small. From the mathematical point of view, this corresponds to the case of the stationary point close to the edge. For the *P* field inside the penumbral boundary layer we obtain

$$\text{Class I:} \quad u_z^{P(\text{penumbra})} \simeq F[\text{sgn}(l-r)\sqrt{k_P(s_- - z)}] \times [u_1^{P(\text{direct})} + u_{z,2}^{P(\text{direct})}], \tag{29}$$

$$\text{Class II:} \quad u_z^{P(\text{penumbra})} \simeq F[\text{sgn}(l-r)\sqrt{k_P(s_- - z)}] \times [u_1^{P(\text{direct})} + u_{z,2}^{P(\text{direct})}] + u_2^*, \tag{30}$$

where

$$u_2^* \simeq -\frac{P_0}{2\sqrt{\pi}\varrho c_P^2 k_P} \frac{q(r)}{\text{sgn}(l-r)\sqrt{k_P(s_- - z)}} e^{i(k_P s_- - \pi/4)}, \tag{31}$$

and

$$\text{Class I:} \quad u_r^{P(\text{penumbra})} \simeq u_{r,-}^{P(\text{edge})} + F[\text{sgn}(l-r) \times \sqrt{k_P(s_- - z)}] u_{r,2}^{P(\text{direct})}, \tag{32}$$

$$\text{Class II:} \quad u_r^{P(\text{penumbra})} \simeq F[\text{sgn}(l-r) \times \sqrt{k_P(s_- - z)}] u_{r,2}^{P(\text{direct})}. \tag{33}$$

In the above formulas F is the Fresnel integral [e.g., Borovikov, 1994, Eq. (6.1a)]. Note that for the z-component we have included term $u_{z,2}^{P(\text{direct})}$, because it grows with the distance from the transducer, and thus can be significant for large z. In Class II we have also added an additional term u_2^* (see Borovikov and Kinber, 1994, p. 107). This term is important, because the leading term $u_1^{P(\text{direct})}$ is zero straight underneath the transducer rim, and, although both the numerator and denominator of the second factor in (31) have zeros there, their ratio has a finite value. The shape of the penumbral region is determined by the condition that the phase difference between the direct P and first edge P arrival is approximately π, that is,

$$k_P(s_- - z) \approx \pi. \tag{34}$$

Therefore, it is parabolic.

D. High-frequency asymptotics of the *S* field

The S field is described by the integral

$$\mathbf{u}^S \simeq -\frac{P_0}{2\pi\varrho c_S^2}\int_0^{2\pi}\int_0^l q(r')\frac{A^S(\theta')}{s'}\tilde{\mathbf{n}}^S e^{ik_S s'} r'\, dr'\, d\varphi', \tag{35}$$

where the unit vector $\tilde{\mathbf{n}}^S$ is given by

$$\tilde{\mathbf{n}}^S = \cos\theta'\left[\frac{r'\sin(\varphi-\varphi')}{s'\sin\theta'}\mathbf{e}_r + \frac{r - r'\cos(\varphi-\varphi')}{s'\sin\theta'}\mathbf{e}_\varphi\right] - \sin\theta'\,\mathbf{e}_z. \tag{36}$$

The integral in (35) possesses the same critical points as the integral in (12). We describe their contributions similarly to the P field.

1. Geometrical regions

The contribution of the stationary point gives the direct S wave and the corresponding leading term is zero

$$\mathbf{u}_1^{S(\text{direct})} \simeq 0, \quad r < l. \tag{37}$$

This is due to the fact that $A^S(0)=0$ [see (A2)]. For the uniform case all other terms of the ray series vanish as well (see Fradkin *et al.*, 1998). Therefore, using the same argument as for the direct P wave, we obtain the second order term as

$$\mathbf{u}_2^{S(\text{direct})} \simeq -\frac{2P_0}{\varrho c_P c_S k_S^2} e^{ik_S z} q'(r)\mathbf{e}_r, \quad r < l. \tag{38}$$

The leading terms of the edge S wave in the geometrical regions are obtained similarly to the edge P wave and are given by

$$\text{Class I:} \quad \mathbf{u}_\pm^{S(\text{edge})} \simeq \overline{\mathbf{u}}_\pm^{S(\text{edge})} q(l), \tag{39}$$

$$\text{Class II:} \quad \mathbf{u}_\pm^{S(\text{edge})} \simeq \overline{\mathbf{u}}_\pm^{S(\text{edge})} \frac{i}{k_S}\frac{s_\pm}{l \pm r} q'(l). \tag{40}$$

2. Boundary layers

The leading terms of the edge S wave inside the axial region are obtained similarly to the edge P wave and given by

$$\text{Class I:} \quad \mathbf{u}^{S(\text{axial})} \simeq \overline{\mathbf{u}}^{S(\text{axial})} q(l), \tag{41}$$

$$\text{Class II:} \quad \mathbf{u}^{S(\text{axial})} \simeq \overline{\mathbf{u}}^{S(\text{axial})} \frac{i}{k_S}\frac{s_*}{l} q'(l). \tag{42}$$

In the above equations $\overline{\mathbf{u}}_\pm^{S(\text{edge})}$ and $\overline{\mathbf{u}}^{S(\text{axial})}$ are the corresponding leading terms for the uniform case given by Eqs. (B4) and (B6), respectively.

The r-component of the penumbral S field may be obtained similarly to the P field:

$$\text{Class I:} \quad u_r^{S(\text{penumbra})} \simeq u_{r,-}^{S(\text{edge})} + F[\operatorname{sgn}(l-r) \times \sqrt{k_S(s_- - z)}]\, u_2^{S(\text{direct})}, \tag{43}$$

$$\text{Class II:} \quad u_r^{S(\text{penumbra})} \simeq F[\operatorname{sgn}(l-r) \times \sqrt{k_S(s_- - z)}]\, u_2^{S(\text{direct})}. \tag{44}$$

The z-component of the penumbral asymptotic series coincides with the z-component of the edge S wave, since the direct S wave has neither first nor second order term in this direction.

E. Discussion and numerical results

In order to conduct numerical tests we consider a particular type of apodization

$$q(r) = 1 - (r/l)^2, \tag{45}$$

so that the pressure falls parabolically to zero at the transducer edge (Class II). Then we have

$$q'(r) = -2r/l^2, \quad \Delta_\perp q(r) = -4/l^2, \tag{46}$$

and the second order term of the z-component of the direct P wave turns out to be

$$u_{z,2}^{P(\text{direct})} \simeq \frac{P_0}{\varrho c_P^2 k_P}\frac{2z}{k_P l^2} e^{ik_P z} = \frac{P_0}{\varrho c_P^2 k_P}\frac{z}{\pi d_{\text{NF}}} e^{ik_P z}, \quad r < l, \tag{47}$$

where $d_{\text{NF}} = l^2/\lambda_P$ is the near-field length. It follows that at the boundary of the near zone the amplitude of the second order term is π times smaller than the leading term on the axis. The leading term of the r-component of the direct P wave becomes

$$u_{r,2}^{P(\text{direct})} \simeq -\frac{2P_0}{\varrho c_P^2 k_P^2}\frac{r}{l^2} e^{ik_P z}, \quad r < l. \tag{48}$$

For the parabolic apodization the leading term of the direct S wave is

$$\mathbf{u}_2^{S(\text{direct})} \simeq \frac{4P_0}{\varrho c_P c_S k_S^2}\frac{r}{l^2} e^{ik_S z}\mathbf{e}_r, \quad r < l. \tag{49}$$

The numerical implementation of the asymptotic formulas is quite straighforward. For instance, the z-component of

the total P field inside the main beam is a sum of the direct and edge P waves,

$$u_{z,1}^{P(\text{direct})}+u_{z,2}^{P(\text{direct})}+u_{z,+}^{P(\text{edge})}+u_{z,-}^{P(\text{edge})}, \tag{50}$$

while outside the main beam the first two terms in the last equation are zero. Inside the axial region the total P field is a sum of the direct P wave and the axial boundary-layer asymptotics of the edge P wave,

$$u_{z,1}^{P(\text{direct})}+u_{z,2}^{P(\text{direct})}+u_{z}^{P(\text{axial})}. \tag{51}$$

The field inside the penumbral boundary layer is a sum of the penumbral asymptotics and the edge P arrival from the farthest edge point,

$$u_{z}^{P(\text{penumbra})}+u_{z,+}^{P(\text{edge})}. \tag{52}$$

The sizes of the axial and penumbral layers are determined by Eqs. (28) and (34), respectively. The layers are narrower the higher the frequency. There might be a small mismatch at the transition point from the geometrical region to either boundary layer. It may be removed by using the uniform asymptotics valid both inside and outside the layer, or by numerical smoothing. However, no smoothing is required when dealing with pulses.

Below, we compare our asymptotic results with those obtained by the direct numerical integration of the integral representation of the displacement. The latter is given in Appendix C. The test parameters are chosen as follows: the wavespeeds are $c_P=5840$ m/s and $c_S=3170$ m/s; the density is $\varrho=7770$ kg/m^3; the transducer radius is $l=10$ mm; the applied pressure is $P_0=10$ MPa, and the frequency is $f=\omega/2\pi=5$ MHz. For these parameters the radiating near zone extends from about 1.2 mm to about 86 mm. Note that as the frequency f increases the radiating near zone gets larger, and the high-frequency asymptotics become more accurate.

The z-component of the total P field [Fig. 4(a)] makes the most significant contribution to the field of a normal transducer. This component is strongest on the axis, diminishes towards the edge and practically vanishes outside the main beam. Due to the small amplitude of the edge wave there is practically no oscillation. The r-component of the direct S wave is much weaker, and, therefore, the amplitude oscillation of the total S field is noticeable [Fig. 4(b)]. This component grows towards the edge, in accordance with Eq. (49). Both figures show a good agreement between the asymptotic and exact solutions.

In Fig. 5(a) and (b) we indicate the limits of applicability of the radiating near- and far-field asymptotics for the z-component of the total P field. The far-field asymptotics for the parabolic apodization are given in Appendix C. It can be seen that on the transducer axis [see Fig. 5(a)] the exact and radiating near field asymptotic solutions practically coincide, so that the solid and dashed lines are indistinguishable. In Fig. 5(b) the comparison is made at $r=10$ mm, that is straight underneath the transducer edge. In this case the asymptotics give good agreement with the exact numerical solution within the radiating near zone and become inapplicable at distances of about two near-field lengths. Both figures demonstrate that at these distances the far-field asymptotics may give a good approximation.

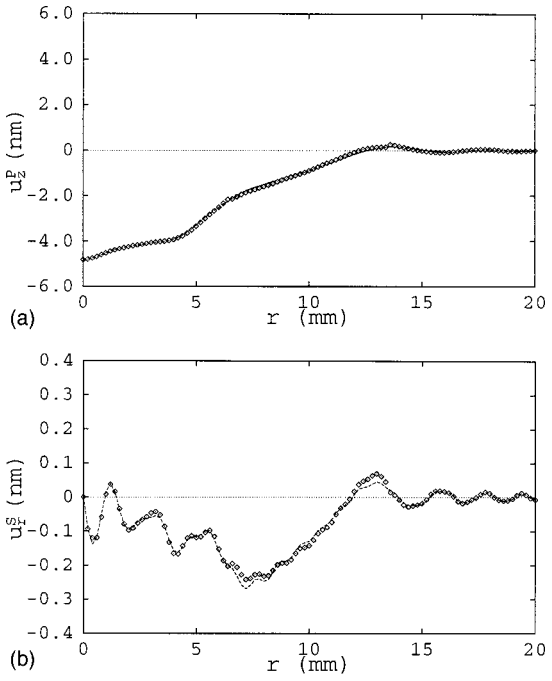


FIG. 4. Comparison between the asymptotic (boxes) and exact (dashed line) displacement at $z=20$ mm: (a) z-component of the total P field; (b) r-component of the total S field.

II. PROPAGATION OF PULSES

The asymptotic description of the radiating near field of a time-harmonic circular normal transducer of arbitrary

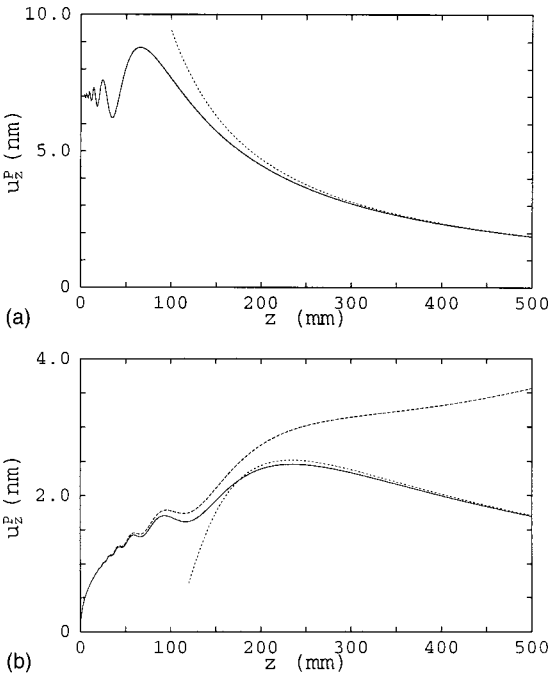


FIG. 5. The amplitude of the z-component of the direct P wave: (a) on the axis; (b) straight underneath the transducer rim. The solid line is the exact solution, the dashed line is the asymptotic solution, and the dotted line is the far-field asymptotic solution, respectively. The solid and dashed lines are indistinguishable from each other in (a).

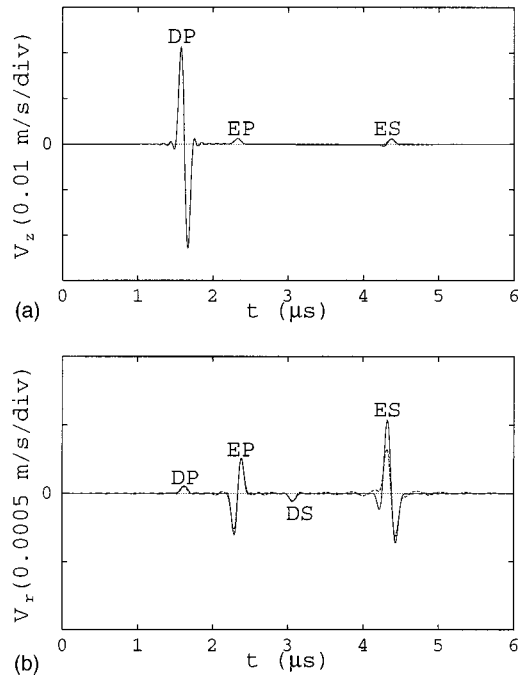


FIG. 6. The near velocity field in the axial region at $z=10$ mm: (a) the z-component on the axis; (b) the r-component at $r=0.3$ mm. Solid and dashed lines are the exact and asymptotic solutions, respectively. DP—direct P, EP—edge P, DS—direct S, and ES—edge S pulse.

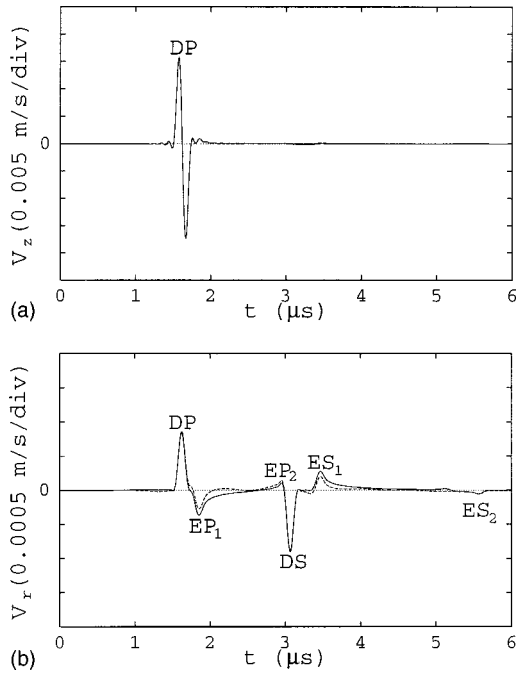


FIG. 7. The near velocity field in the geometrical region at $z=10$ mm and $r=5$ mm: (a) the z-component; (b) the r-component. Subscripts ''$_1$'' and ''$_2$'' relate to edge pulses arriving from the nearest and farthest edge points respectively.

apodization which has been derived in the previous section can be used to model the propagation of pulses by means of harmonic synthesis, that is by performing the inverse Fourier transform with respect to time. The displacement $\mathbf{U}(t)$ is given by

$$\mathbf{U}(t)=\mathrm{Re}\frac{1}{\pi}\int_0^\infty \mathbf{u}(\omega)p(\omega)e^{-i\omega t}\,d\omega, \tag{53}$$

where $\mathbf{u}(\omega)$ is the time-harmonic displacement considered above, and $p(\omega)$ is the Fourier transform of the initial pulse shape $P(t)$. Below, we consider the particle velocity $\mathbf{V}(t)$, because this is an important quantity measured by receiving probes (e.g., Weight, 1987).

In computations described below the same parameters have been used as in the previous section. A wide band pulse has been considered containing not only high, but intermediate and low frequencies as well. It has been chosen to be one cycle of $\sin(2\pi ft)$ with $f=5$ MHz and, consequently, the duration 0.2 μs. Similarly to the previous section, we compare the asymptotic and exact numerical solutions, both fast Fourier transformed. In the asymptotic code the bandwidth is taken to be 0.1–10 MHz, and in the exact, it is extended down to 0 MHz.

A. Velocity field inside the axial region

First, let us choose the observation point to lie on the transducer axis at depth $z=10$ mm. The corresponding waveform of the z-component of the particle velocity is presented in Fig. 6(a). The first pulse in the train is due to the direct P wave (DP). The second and third pulses are due to the edge P (EP) and S (ES) arrivals, respectively. The amplitudes of these pulses are much smaller than the direct P pulse. This is in contrast with the uniform transducer where the three pulses have a comparable amplitude and the edge S may be larger than the direct P pulse [e.g., McNab *et al.*, 1989, Fig. 5(c); Gridin and Fradkin, 1998, Fig. 6(a)]. The waveform of the r-component [Fig. 6(b)] possesses both direct P and S pulses which are not present in the uniform case [e.g., Gridin and Fradkin, 1998, Fig. 6(c)]. The other pulses are due to the edge P and S waves, respectively. All pulses have much smaller amplitudes than the z-component of the direct P pulse. There is a good agreement between the asymptotic and exact solutions, except for the edge S pulse. The discrepancy is due to the head wave effects being neglected. It may be present for observation points lying inside the narrow boundary layers surrounding the critical rays (see Gridin, 1998).

B. Velocity field outside the axial region

Let us consider the case of the observation point lying outside the axial region but well inside the main beam. The respective waveforms of the z- and r-components at $z=10$ mm and $r=5$ mm are presented in Fig. 7(a) and (b). The z-component of the direct P pulse dominates. The r-components of both direct P and S pulses are about 15 times smaller, and the edge P and S wave arrivals are smaller still. This means that the edge P and S waves are practically eliminated.

Straight underneath the transducer rim (Fig. 8) the first and second pulses are due to the first P and S arrivals, respectively. The other discernible pulse is the edge P arrival from the farthest edge point. For the z-component of the direct P pulse the agreement between the asymptotic and exact numerical solution is very good. For the r-components

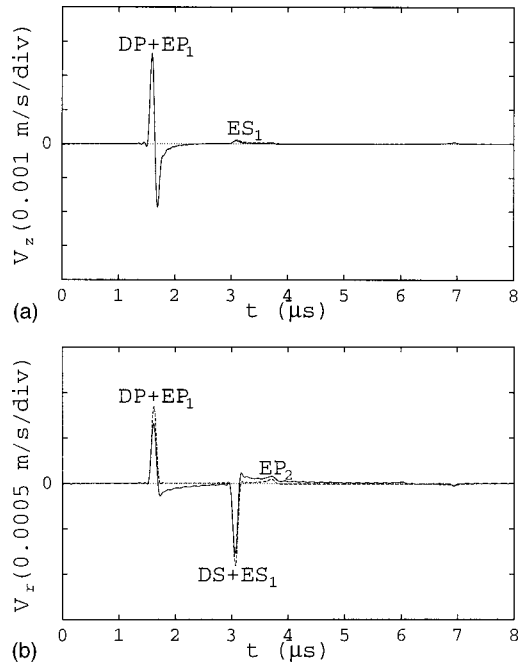


FIG. 8. The near velocity field straight underneath the transducer rim ($r=l=10$ mm) at $z=10$ mm: (a) the z-component; (b) the r-component. The key as in Fig. 7.

of the direct P and S waves it is slightly worse but might be improved by taking into account the next terms of the ray series. Since the r-components are about 30 times smaller than the direct P pulse on the axis, there is no need to strive for such accuracy.

The asymptotic code is hundreds of times faster than the direct numerical one. For example, the computation time of the asymptotic waveform in Fig. 8(a) is about 140 times smaller than that of the exact waveform. Note that in Fig. 8(a) the distance from the transducer to the observation point is just 10 mm. As this distance increases the integral in the direct numerical code becomes more rapidly oscillating. Thus, the direct numerical code may become several times slower, while the speed of the asymptotic code does not change. The same conclusion holds for a transducer of arbitrary apodization of either Class I or II.

III. CONCLUSIONS

A method for simulating the propagation of pulses radiated by a circular normal transducer of arbitrary apodization into an isotropic and homogeneous elastic half-space has been developed. First, the model proposed in Fradkin *et al.* (1998) for the uniform transducer has been extended to the case of nonuniform load to obtain the high-frequency asymptotics of the time-harmonic field. Then, the transient field has been described by means of harmonic synthesis. The asymptotics involving elementary or well-known special functions elucidate the physics and give explicit dependence of the radiated waves on model parameters. The formulas are applicable in the radiating near field. Unlike the uniform case, both direct P and S waves are present, with the latter being smaller by an order of magnitude. Moreover, when the apodization falls to zero at the transducer edge both edge P and S waves may be practically negligible. The code based on the asymptotics has been tested against the exact numerical solution for the case of parabolic apodization. It has proved to be hundreds of times faster, and in many realistic cases the accuracy is quite acceptable.

ACKNOWLEDGMENTS

The author gratefully acknowledges a South Bank Research Scholarship. He would like to thank Dr. L. J. Fradkin for the careful reading of the manuscript and useful discussions. Comments of unknown reviewers are also appreciated.

APPENDIX A: DIRECTIVITY PATTERNS FOR A NORMAL POINT LOAD

The directivity patterns for the P and S spherical waves in the far field of a point source acting normally on the surface of a homogeneous and isotropic elastic half-space are as follows:

$$A^P(\theta)=\frac{2\sin^2\theta-\gamma^{-2}}{\mathrm{R}^P(\sin\theta)}\cos\theta \tag{A1}$$

and

$$A^S(\theta)=\frac{2\sin\theta(\gamma^2-\sin^2\theta)^{1/2}}{\mathrm{R}^S(\sin\theta)}\cos\theta, \tag{A2}$$

with $\gamma=c_S/c_P$, and the Rayleigh-type functions

$$\mathrm{R}^P(\xi)=(2\xi^2-\gamma^{-2})^2+4\xi^2(1-\xi^2)^{1/2}(\gamma^{-2}-\xi^2)^{1/2} \tag{A3}$$

and

$$\mathrm{R}^S(\xi)=(2\xi^2-1)^2-4\xi^2(\xi^2-1)^{1/2}(\xi^2-\gamma^2)^{1/2}. \tag{A4}$$

The unit displacement vectors are given by

$$\begin{aligned}\mathbf{n}^P(\theta)&=\sin\theta\mathbf{e}_r+\cos\theta\mathbf{e}_z,\\ \mathbf{n}^S(\theta)&=\cos\theta\mathbf{e}_r-\sin\theta\mathbf{e}_z.\end{aligned} \tag{A5}$$

APPENDIX B: THE UNIFORM LOAD

The leading terms of the high-frequency asymptotics for a uniform circular normal transducer acting on the surface of an isotropic and homogeneous elastic half-space have been derived in Fradkin *et al.* (1998). Some of their formulas contain misprints, and for this reason we summarize their main results below:

1. Geometrical regions

The direct P and S waves inside the main beam:

$$\overline{\mathbf{u}}^{P(\mathrm{direct})}=\frac{P_0}{\varrho c_P^2 k_P}e^{i(k_P z+\pi/2)}\mathbf{e}_z, \tag{B1}$$

$$\overline{\mathbf{u}}^{S(\mathrm{direct})}=0. \tag{B2}$$

The edge P and S waves:

$$\overline{\mathbf{u}}_{\pm}^{P(\text{edge})} \simeq \frac{P_0}{(2\pi)^{1/2}\varrho c_S^2 k_P} A^P(\theta_{\pm}) \times \left[\frac{s_{\pm}}{k_P(l\pm r)^2}\right]^{1/2} \left(\frac{l}{r}\right)^{1/2} e^{i(k_P s_{\pm} + \pi/2 \mp \pi/4)} \times [\pm \sin\theta_{\pm}\mathbf{e}_r + \text{sgn}(l\pm r)\cos\theta_{\pm}\mathbf{e}_z] \tag{B3}$$

and

$$\overline{\mathbf{u}}_{\pm}^{S(\text{edge})} \simeq \frac{P_0}{(2\pi)^{1/2}\varrho c_S^2 k_S} A^S(\theta_{\pm}) \times \left[\frac{s_{\pm}}{k_S(l\pm r)^2}\right]^{1/2} \left(\frac{l}{r}\right)^{1/2} e^{i(k_S s_{\pm} + \pi/2 \mp \pi/4)} \times [\pm \cos\theta_{\pm}\mathbf{e}_r - \text{sgn}(l\pm r)\sin\theta_{\pm}\mathbf{e}_z]. \tag{B4}$$

2. Boundary layers

The edge P and S waves inside the axial region:

$$\overline{\mathbf{u}}^{P(\text{axial})} \simeq -\frac{P_0}{\varrho c_S^2 k_P} A^P(\theta_*) e^{ik_P s_*} [\sin\theta_* J_1(k_P r \sin\theta_*)\mathbf{e}_r - i\cos\theta_* J_0(k_P r\sin\theta_*)\mathbf{e}_z] \tag{B5}$$

and

$$\overline{\mathbf{u}}^{S(\text{axial})} \simeq -\frac{P_0}{\varrho c_S^2 k_S} A^S(\theta_*) e^{ik_S s_*} [\cos\theta_* J_1(k_S r \sin\theta_*)\mathbf{e}_r - i\sin\theta_* J_0(k_S r\sin\theta_*)\mathbf{e}_z], \tag{B6}$$

where J_0 and J_1 are the Bessel functions of the zeroth and first order, respectively.

The P field inside the penumbral boundary layer:

$$u_z^{P(\text{penumbra})} \simeq F[\text{sgn}(l-r)\sqrt{k_P(s_- - z)}]\overline{u}^{P(\text{direct})}. \tag{B7}$$

The distances and angles used in the above formulas are as follows:

$$s_{\pm} = [z^2 + (l\pm r)^2]^{1/2}, \quad \sin\theta_{\pm} = |l\pm r|/s_{\pm}, \tag{B8}$$

and

$$s_* = (z^2 + l^2)^{1/2}, \quad \sin\theta_* = l/s_*. \tag{B9}$$

APPENDIX C: THE INTEGRAL REPRESENTATION OF THE DISPLACEMENT AND THE FAR-FIELD ASYMPTOTICS FOR THE PARABOLIC APODIZATION

Applying the Hankel transform technique to the problem formulated in Sec. I A and the parabolic apodization (45) we obtain the integral representation of the displacement as

$$\begin{aligned} u_r &= \frac{2P_0}{\varrho c_S^2}\int_0^{\infty} [(2\xi^2 - k_S^2)e^{-\alpha z} - 2\alpha\beta e^{-\beta z}]\frac{1}{R} \times J_1(\xi r)J_2(\xi l)\, d\xi, \\ u_z &= \frac{2P_0}{\varrho c_S^2}\int_0^{\infty} [(2\xi^2 - k_S^2)e^{-\alpha z} - 2\xi^2 e^{-\beta z}]\frac{\alpha}{R\xi} \times J_0(\xi r)J_2(\xi l)\, d\xi, \end{aligned} \tag{C1}$$

where

$$\alpha = (\xi^2 - k_P^2)^{1/2}, \quad \beta = (\xi^2 - k_S^2)^{1/2}, \tag{C2}$$

and Rayleigh's function R is given by

$$R = (2\xi^2 - k_S^2)^2 - 4\alpha\beta\xi^2. \tag{C3}$$

The branches of α and β and the branch cuts have been chosen in such a way that the radicals have non-negative real parts on the real ξ-axis, and the contour of integration in Eq. (C1) passes below the branch points.

The far-field asymptotics can be obtained from Eq. (C1), and their leading terms are

$$\mathbf{u}^{P(\text{far field})} \simeq \mathbf{u}_0^{P(\text{point})} Q_0 \frac{4J_2(k_P l\sin\theta)}{(k_P l\sin\theta)^2} \tag{C4}$$

and

$$\mathbf{u}^{S(\text{far field})} \simeq \mathbf{u}_0^{S(\text{point})} Q_0 \frac{4J_2(k_S l\sin\theta)}{(k_S l\sin\theta)^2}, \tag{C5}$$

where $\mathbf{u}_0^{P(\text{point})}$ and $\mathbf{u}_0^{S(\text{point})}$ are the far-field asymptotics of a normal point source [Eq. (7)], and $Q_0 = \pi l^2 P_0$.

Babich, V. M., and Buldyrev, V. S. (**1991**). *Short-wavelength Diffraction Theory: Asymptotic Methods* (Springer-Verlag, Heidelberg).

Borovikov, V. A. **(1994)**. *Uniform Stationary Phase Method* (IEE, London).

Borovikov, V. A., and Kinber, B. Ye. (**1994**). *Geometrical Theory of Diffraction* (IEE, London).

Bresse, L. F., and Hutchins, D. A. (**1989**). ''Transient generation of elastic waves in solids by a disk-shaped normal force source,'' J. Acoust. Soc. Am. **86**, 810–817.

Djelouah, H., and Baboux, J. C. (**1992**). ''Transient ultrasonic field radiated by a circular transducer in a solid medium,'' J. Acoust. Soc. Am. **92**, 2932–2941.

Fradkin, L. J., and Kiselev, A. P. (**1997**). ''The two-component representation of time-harmonic elastic body waves in the high- , and intermediate-frequency regimes,'' J. Acoust. Soc. Am. **101**, 52–65.

Fradkin, L. J., Kiselev, A. P., and Krylova, E. (**1998**). ''The radiating near field asymptotics of a time-harmonic circular normal ultrasonic transducer in an elastic half-space,'' J. Acoust. Soc. Am. **104**, 1178–1187.

Gridin, D. (**1998**). ''High-frequency asymptotic description of head waves and boundary layers surrounding critical rays in an elastic half-space,'' J. Acoust. Soc. Am. **104**, 1188–1197.

Gridin, D., and Fradkin, L. J. (**1998**). ''High-frequency asymptotic description of pulses radiated by a circular normal transducer into an elastic half-space,'' J. Acoust. Soc. Am. **104**, 3190–3198.

Harris, J. G. (**1988**). ''The wavefield radiated into an elastic half-space by a transducer of large aperture,'' J. Appl. Mech. **55**, 398–404.

Lhémery, A. (**1994**). ''A model for the transient ultrasonic field radiated by an arbitrary loading in a solid,'' J. Acoust. Soc. Am. **96**, 3776–3786.

Lhémery, A., and Stacye, R. (**1995**). ''Transient field radiated by nonuniform transducers in a solid calculated by an approximate model and the finite-difference method,'' Rev. Prog. Quant. Nondestr. Eval. **14**, 997–1004.
Miller, G. F., and Pursey, H. (**1954**). ''The field , and radiation impedance of mechanical radiators on the free surface of a semi-infinite isotropic solid,'' Proc. R. Soc. London, Ser. A **223**, 521–541.
Morse, P. M., and Ingard, K. U. (**1968**). *Theoretical Acoustics* (McGraw–Hill, New York).
McNab, A., Cochran, A., and Campbell, M. A. (**1989**). ''The calculation of acoustic fields in solids for transient normal surface force sources of arbitrary geometry, and apodization,'' J. Acoust. Soc. Am. **87**, 1455–1465.
Stacey, R., and Weight, J. P. (**1994**). ''Pulse-echo scattering in solids with nonuniform transducers,'' IEE Proc., Part A: Phys. Sci., Meas. Instrum., Manage. Educ. **141**(5), 363–368.
Ushakov, V. M., and Danilov, V. N. (**1997**). ''Formation of the directivity pattern of transducers with a nonuniform pressure distribution over the radiating surface of the piezoelectric plates,'' Russian J. NDT **33**, 291–301.
Weight, J. P. (**1984**). ''New transducers for high-resolution ultrasonic testing,'' NDT Int. **17**, 3–8.
Weight, J. P. (**1987**). ''A model for the propagation of short pulses of ultrasound in a solid,'' J. Acoust. Soc. Am. **81**(4), 815–826.

A unified boundary element method for the analysis of sound and shell-like structure interactions. I. Formulation and verification

Shaohai Chen and Yijun Liu[a)]
Department of Mechanical, Industrial and Nuclear Engineering, P.O. Box 210072, University of Cincinnati, Cincinnati, Ohio 45221-0072

(Received 2 February 1999; revised 12 May 1999; accepted 27 May 1999)

A unified boundary element method (BEM) is developed in this paper to model both the exterior acoustic field and the elastic shell-like structure in a coupled analysis. The conventional boundary integral equation (BIE) for three-dimensional (3D) elastodynamics is applied to thin shell-like structures which can have arbitrary shapes and small thicknesses. The nearly singular integrals existing in the BIE when applied to thin bodies are transformed to nonsingular line integrals and are evaluated accurately and efficiently. For the exterior 3D acoustic domain, the Burton and Miller composite BIE formulation is employed to overcome the fictitious eigenfrequency difficulty (FED) and the thin-shape breakdown (TSB). Conforming C^0 quadratic elements are employed in the discretization of the two sets of BIEs. The developed BIE formulations are valid for both radiation and scattering problems and for all wave numbers. Numerical examples using spherical and cylindrical shells, including nonuniform thickness and nondimensional wave numbers up to 12, clearly demonstrate the effectiveness and accuracy of the developed BEM approach. © *1999 Acoustical Society of America.* [S0001-4966(99)01709-9]

PACS numbers: 43.20.Fn, 43.20.Rz, 43.40.Rj [CBB]

INTRODUCTION

The effective control of noise and vibration in a structural acoustic system depends largely on the accurate evaluation of the sound–structure interaction which is characterized by the energy transferring back and forth between the acoustic field and the elastic structure. When the structural impedance is comparable to the acoustic impedance, both of the responses of the structure and the sound field can be significantly affected by this sound–structure interaction. Many numerical techniques have been developed for the analysis of the sound–structure interaction problems, since analytical approaches are limited to simple geometries and loading conditions. For a review of the subject on sound and elastic structure interactions, refer to the classical work in Refs. 1–3 and the references therein.

For the numerical analysis of the acoustic wave, the finite element method (FEM),[4] infinite element method (IEM),[5,6] and boundary element method (BEM)[7–15] have been investigated intensively, among others. Detailed reviews and more references for the three major techniques can be found in Refs. 4, 5, and 15. The FEM uses 3D elements to model the 3D acoustic field. When the infinite acoustic field is encountered, the finite-element model has to be truncated at an artificial outer boundary at which an approximate nonreflecting boundary condition is applied. The Sommerfeld radiation condition is in general not satisfied in this early FEM approach. In the early versions of the IEM, a variety of shape functions were used to approximate the spatial decay of the acoustic pressure outside the finite-element model. Recently, a new infinite element approach using a multipole expansion of the acoustic pressure in the field exterior to a spheroid surrounding the structure was developed.[5,6] This multipole expansion in spheroidal coordinates satisfies the Sommerfeld radiation condition automatically at infinity and can converge to the exact solution with only a few layers of 3D acoustic elements outside the structure.

The BEM has long been considered as a rigorous approach to exterior acoustic problems. The Sommerfeld radiation condition is satisfied exactly by the boundary integral equation (BIE) and only the interior boundary (i.e., the outer surface of a structure) needs to be discretized. Therefore, the analysis of structures with simple or complicated geometries, or multiple scatterers, can be performed conveniently by the BEM. Since near field solutions may be sensitive to small features on the surface of the elastic structure, the ability of modeling these small features without additional efforts also makes the BEM attractive. The possible drawbacks in the BEM approach include the nonuniqueness problem which arises when the conventional BIE (CBIE) is applied to an exterior acoustic domain. This problem is also referred to as the fictitious eigenfrequency difficulty (FED), since nonunique solutions arise at the eigenfrequencies of the associated interior problems.[7,8] However, this FED can be circumvented by either the CHIEF method[7] or the Burton and Miller composite formulation.[8] It has been shown in Refs. 12–15 and many others that the Burton and Miller composite BIE formulation, employing a linear combination of the CBIE and the hypersingular BIE (HBIE), is the most effective method to overcome the fictitious eigenfrequency difficulty for exterior acoustic problem and elastodynamic problems.[16] It has also been demonstrated[17–19] that the composite BIE formulation can overcome the thin-shape breakdown (TSB)[18,20] existing in the CBIE when it is applied to domains surrounding

[a)]Electronic mail: Yijun.Liu@uc.edu

thin shell-like structures. The hypersingular integral in the composite BIE presents no difficulty at all, since it can be readily transformed into weakly singular integrals and accurately evaluated by ordinary numerical quadrature.[14,15]

For the modeling of elastic structures, the FEM with its various formulations for beams, plates, shells, and solids, is the mostly accepted method in structural acoustic analysis. However, there are many assumptions involved in the beam, plate, or shell elements. Solid elements with proper aspect ratios should be used when high accuracy is demanded. Very large FEM models may result when solid elements are applied to thin shell-like structures. It may also be difficult to generate the FEM mesh for thin structures if the geometry is complicated. On the other hand, the BEM has established itself as a viable alternative or complement to the FEM for both elastostatic and elastodynamic problems (see, e.g., Refs. 16, 19, and 21–29) after its accuracy and efficiency have been demonstrated and the numerical difficulties have been eliminated. For structural acoustic analysis, the BEM based on 3D elastodynamics and using surface discretization is also advantageous, since the proper coupling of the elastic structure and the acoustic field can be ensured using the same surface mesh. However, there have been two major difficulties or concerns when the CBIEs are applied directly to thin bodies (including thin voids or open cracks, thin shell-like structures, and thin layered structures), where two parts of the boundary become close to each other. One difficulty is the possible degeneracy of the CBIE for thin bodies.[18] The other is the difficulty of the nearly singular integrals[18,29,30] which arise when the integration is conducted on a surface with the source point being very close to the surface. Because of these two difficulties, the BEM has been considered unsuitable for thin-body problems for a long time. It has been shown in Refs. 18 and 29 that these two difficulties can be overcome readily with some analytical efforts. The degeneracy, which happens when CBIEs are applied to the two surfaces of a thin void or crack in an exterior-type problem, can be overcome by employing the Burton and Miller composite formulation.[17–19] For an interior-type problem (thin shells, etc.), it has been shown[29] that no degeneracy will happen when CBIEs are applied on both sides of a thin shell. Accurate results for both 2D and 3D thin structures have been obtained after the nearly singular integrals are handled correctly using the line integral approach.[29,31]

For the coupled structural acoustic problem, the most commonly used approach is the FEM/BEM approach,[32–34] which employs the finite elements for the elastic structure and boundary elements for the exterior acoustic field. This approach combines the advantages of both the FEM and BEM. The drawbacks in this approach include the mismatch of the desirable mesh sizes on the interface between the acoustic field and the elastic field. Since the FEM mesh density required for the elastic structure is usually higher than the BEM mesh density for the acoustic field, and a common mesh should be used for the two domains to ensure proper interface conditions, the selected mesh could be unrealistically dense for the acoustic field and the efficiency can suffer. The FEM/FEM (or FEM/IEM) approach, where the structure is modeled by FEM and the exterior acoustic field by the IEM, is regaining popularity in computational structural acoustics, due to recent success.[4–6] The least explored approach is the BEM/BEM approach, which has been studied for bulky elastic structures in Refs. 35 and 36. In Ref. 35, the BEM/BEM approach is applied to the acoustic wave interacting with bulky elastic bodies in the context of the nondestructive evaluation. In Ref. 36, the BEM/BEM approach was tested on both bulky solid and hollow sphere (thick shell with thickness to radius ratio $=0.5$) using isoparametric elements. The effectiveness and efficiency of the BEM/BEM approach were demonstrated clearly with accurate results obtained for both radiation and scattering problems.[35,36] However, the interaction of the acoustic wave with thin shells was not analyzed in Refs. 35 and 36, probably due to concerns of the degeneracy of CBIEs for thin bodies and the difficulty of dealing with nearly singular integrals which were still troublesome to compute a decade ago.

The present paper extends the BEM/BEM approach to the structural acoustic problem for thin shell-like structures which can have arbitrarily small and nonuniform thickness. Two sets of BIEs, one for the elastic structure (finite interior domain) and the other for the acoustic field (infinite exterior domain), are presented. The conventional BIE based on 3D elastodynamics is used for the elastic structure. The nearly singular integrals are transformed into line integrals which are computed very accurately and efficiently, based on the recent development of the BEM for thin structures.[18,19,29–31] The fictitious eigenfrequency difficulty and the thin-shape breakdown in the CBIE are removed by using the composite BIE formulation using a linear combination of the CBIE and HBIE in the acoustic domain. The weakly singular form of the hypersingular BIE[15] is employed, which can be readily evaluated by the usual numerical quadrature. The two sets of BIEs are coupled at the outer (wet) surface of the structure by the interface conditions. Quadratic conforming elements (with C^0 continuity) are used for the discretization of the surfaces of the elastic structure. For the verification of the coupled BIE formulations, spherical shells of different thickness (including nonuniform thickness) and materials, and a cylindrical shell structure, are tested for radiation and scattering problems. Very satisfactory results are obtained, which clearly demonstrate the effectiveness and accuracy of the developed BEM/BEM approach to the structural acoustic problems for thin shell-like structures. Efforts are underway to further improve the computational efficiency of the developed BEM/BEM approach and to study the multidomain BEM for slender structures. Results will be reported in subsequent papers.

I. BOUNDARY INTEGRAL EQUATION FORMULATION

Consider a 3D elastic thin structure (V) of an arbitrary shape and immersed in an acoustic media (E) with its outer surface denoted by S_a and inner surface by S_b (Fig. 1). The normal on either surface is defined as pointing away from the elastic domain. We consider only time-harmonic wave motion. The acoustic field is assumed to be inviscid and the elastic structure is assumed to be homogeneous, isotropic, and linearly elastic. Body forces are assumed to be negli-

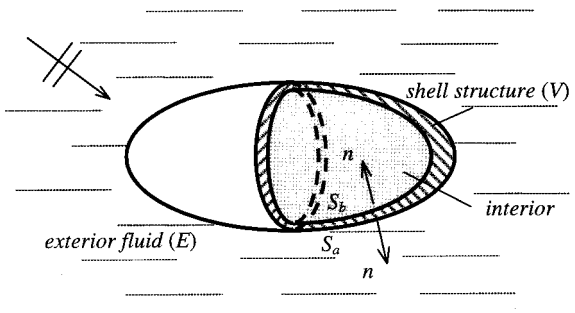


FIG. 1. A 3D shell structure immersed in fluid.

gible. Under these conditions, the wave equation governing the elastic domain (V) can be written as (index notation is used in this paper)

$$(c_1^2-c_2^2)u_{k,ki}(P)+c_2^2u_{i,kk}(P)+\omega^2u_i(P)=0,\ \ \forall P\in V, \tag{1}$$

in which c_1 and c_2 are the wave speeds of the pressure wave (P-wave) and shear wave (S-wave), respectively; u_i the displacement at a point $P\in V$; ω the angular frequency of oscillation. The dependence of u_i on ω has been suppressed. The governing equation for the acoustic domain is the Helmholtz equation,

$$\nabla^2\phi(P)+k^2\phi(P)=0,\ \ \forall P\in E, \tag{2}$$

where $\phi=\phi^S+\phi^I$ is the total disturbed acoustic pressure at a point P, ϕ^S the scattered wave, ϕ^I the incident wave for a scattering problem, $k=\omega/c$ the wave number, and c the speed of sound in the fluid.

On the two surfaces S_a and S_b of the structure, boundary or interface conditions need to be specified. On surface S_b, harmonic excitations in the form of surface displacement or surface traction can be applied, corresponding to a well-posed boundary value problem. On S_a, where the two domains are in contact, the following interface conditions are specified:

(a) The normal derivative of the acoustic pressure is related to the displacement in the normal direction as

$$\frac{\partial\phi}{\partial n}=\rho_f\omega^2u_n, \tag{3a}$$

where ρ_f is the mean density of the fluid, and u_n the normal component of the displacement.

(b) The normal stress is equal to the acoustic pressure such that

$$t_i=-\phi n_i, \tag{3b}$$

where t_i is the traction and n_i the components of the normal ($i=1,2,3$) in the global coordinates. In addition to these conditions, the acoustic pressure field must satisfy the Sommerfeld radiation condition at infinity, which is automatically satisfied by the BIE.

For the 3D elastodynamic problem, the integral representation of Eq. (1) may be written in the following form:

$$C_{ij}(P_0)u_j(P_0)=\int_S U_{ij}(P,P_0)t_j(P)dS(P)-\int_S T_{ij}(P,P_0)u_j(P)dS(P), \tag{4}$$

in which U_{ij} and T_{ij} are the dynamic displacement and traction kernels, respectively; P the field point and P_0 the source point, S the boundary of the elastic structure ($S=S_a\cup S_b$), and the coefficient $C_{ij}(P_0)=\delta_{ij}$, $1/2\delta_{ij}$, or 0 when the source point P_0 is in the interior region V, on the boundary S (if it is smooth) or in the exterior region E, respectively (δ_{ij} is the Kronecker delta). The second integral in Eq. (4), when P_0 is on the boundary S, is of the Cauchy principle value (CPV) type, which requires delicate numerical quadrature in general. This CPV integral can be avoided by recasting Eq. (4) into a weakly singular form[26] (with P_0 on S) as

$$\int_S[T_{ij}(P,P_0)-\overline{T}_{ij}(P,P_0)]u_j(P)dS(P)+\int_S\overline{T}_{ij}(P,P_0)\times[u_j(P)-u_j(P_0)]dS(P)=\int_S U_{ij}(P,P_0)t_j(P)dS(P),\ \ \forall P_0\in S,\ (S=S_a\cup S_b), \tag{5}$$

where $\overline{T}_{ij}$ is the static traction kernel. In Eq. (5) every integral is at most weakly singular and can be computed using the conventional quadrature.

For thin shell-like structures, the second integral in Eq. (5) becomes nearly singular when the source point is at one surface and the integration is performed on the nearby elements on the other surface. This nearly singular integral can be transformed into line integrals which are not singular at all.[29,30] With the help of these line integrals, Eq. (5) can be applied to shell-like structures and will not break down even when the thickness of the shell is very small.[29] Certainly, one can simply increase the number of integration points or use subdivisions on the element to deal with the nearly singular integrals in the BEM as applied for thin bodies. However, this approach has been found inefficient and prohibitively expensive for computing such integrals.[30]

For the acoustic domain embracing the elastic structure, the conventional boundary integral representation of Eq. (2) is the Helmholtz integral (note the direction of the normal n, Fig. 1),

$$C(P_0)\phi(P_0)=\int_{S_a}\left[\frac{\partial G(P,P_0)}{\partial n}\phi(P)-G(P,P_0)\frac{\partial\phi(P)}{\partial n}\right]dS(P)+\phi^I(P_0), \tag{6}$$

where $G(P,P_0)$ ($=e^{ikr}/4\pi r$, with $r=|\mathbf{P}_0\mathbf{P}|$) is the full space Green's functions, and the coefficient $C(P_0)=1$, 1/2, or 0 when the source point P_0 is in E, on the boundary S_a (if it is smooth) or in V, respectively. When the source point P_0 is on the boundary S_a, the integral for both integrands is weakly singular, contrary to the case of Eq. (4). However,

when the conventional BIE (6) is applied to exterior domains, nonunique solutions will arise at the frequencies corresponding to the eigenfrequencies of the interior domain. It was shown in Refs. 12–15 that with the use of the well-known Burton and Miller formulation, that is, a linear combination of CBIE and HBIE as shown symbolically by

$$\text{CBIE}+\beta\text{HBIE}=0 \quad (\beta=\text{constant}), \tag{7}$$

the fictitious eigenfrequency difficulty can be overcome effectively. It was also found in Refs. 18 and 19 that the thin-shape breakdown of the CBIE can be solved as well by using this composite BIE formulation.

The HBIE in Eq. (7) is readily obtained by taking the directional derivative of Eq. (6) in the direction n_0,

$$\frac{\partial\phi(P_0)}{\partial n_0}=\int_{S_a}\left[\frac{\partial^2 G(P,P_0)}{\partial n\,\partial n_0}\phi(P)-\frac{\partial G(P,P_0)}{\partial n_0}\frac{\partial\phi(P)}{\partial n}\right]dS(P)+\frac{\partial\phi^I(P_0)}{\partial n_0},\quad \forall P_0\in E. \tag{8}$$

Equation (8) can be written in the weakly singular form[15] (with the source point P_0 on the boundary S_a) as

$$\begin{aligned}&\frac{\partial\phi(P_0)}{\partial n_0}-\int_{S_a}\frac{\partial^2\bar{G}(P,P_0)}{\partial n\,\partial n_0}\left[\phi(P)-\phi(P_0)-\frac{\partial\phi(P_0)}{\partial\xi_\alpha}(\xi_\alpha-\xi_{0\alpha})\right]dS(P)-\int_{S_a}\frac{\partial^2}{\partial n\,\partial n_0}[G(P,P_0)-\bar{G}(P,P_0)]\phi(P)dS(P)\\&\quad-e_{\alpha k}\frac{\partial\phi(P_0)}{\partial\xi_\alpha}\int_{S_a}\left[\frac{\partial\bar{G}(P,P_0)}{\partial n_0}n_k(P)+\frac{\partial\bar{G}(P,P_0)}{\partial n}n_k(P_0)\right]dS(P)\\&=-\int_{S_a}\left[\frac{\partial G(P,P_0)}{\partial n_0}+\frac{\partial\bar{G}(P,P_0)}{\partial n}\right]\frac{\partial\phi(P)}{\partial n}dS(P)+\int_{S_a}\frac{\partial\bar{G}(P,P_0)}{\partial n}\left[\frac{\partial\phi(P)}{\partial n}-\frac{\partial\phi(P_0)}{\partial n}\right]dS(P)+\frac{\partial\phi^I(P_0)}{\partial n_0},\quad \forall P_0\in S_a,\end{aligned} \tag{9}$$

where $\bar{G}$ is the static kernel, ξ_α and $\xi_{0\alpha}$ $(\alpha=1,2)$ the two tangential coordinates of the points P and P_0, in a local coordinate $O\xi_1\xi_2\xi_3$ $(\xi_3=n)$, respectively, and $e_{\alpha k}=\partial\xi_\alpha/\partial x_k$ $(k=1,2,3)$ are the first two column vectors of the inverse of the Jacobian matrix.[15]

For the hypersingular integral in (8) to exist as the source point P_0 approaches the boundary or for the weakly singular forms in (9) to work, the density function $\phi(P)$ is required, in theory, to have continuous tangential derivatives ($C^{1,\alpha}$ continuity) in the neighborhood of the source point P_0. This smoothness requirement imposes severe limitations to the applications of HBIEs. For example, this smoothness requirement will exclude, theoretically, the use of C^0 boundary elements, such as the conforming quadratic elements, in the discretizations of HBIEs. Relaxation of this smoothness requirement for HBIEs has been attempted by several authors (see, e.g., Refs. 12–14, 37 and 38). The validation of this relaxation has also been provided in Refs. 15, 39, and 40. It has been postulated in Ref. 15 that the original $C^{1,\alpha}$ continuity requirement on the density function in the HBIE formulations can be relaxed to piecewise $C^{1,\alpha}$ continuity in the numerical implementation of the HBIEs as in the form of Eq. (9), so that conforming quadratic elements can be applied. Converged and very good numerical results have been obtained by adopting this strategy for the acoustic problems.[15] However, for domains with edges and corners, the use of conforming quadratic elements for Eq. (9) is not straightforward. Techniques, such as using coincident nodes, to deal with these situations, need to be tested.

II. DISCRETIZATION OF THE BIE

To obtain the numerical solution of Eqs. (5) and (7), surfaces S_a and S_b are discretized using isoparametric quadratic elements (Fig. 2). The discretized form of Eq. (5) can be expressed in matrix form as

$$\begin{bmatrix}\mathbf{U_{aa}} & \mathbf{U_{ab}}\\ \mathbf{U_{ba}} & \mathbf{U_{bb}}\end{bmatrix}\begin{Bmatrix}\mathbf{t_a}\\ \mathbf{t_b}\end{Bmatrix}+\begin{bmatrix}\mathbf{T_{aa}} & \mathbf{T_{ab}}\\ \mathbf{T_{ba}} & \mathbf{T_{bb}}\end{bmatrix}\begin{Bmatrix}\mathbf{u_a}\\ \mathbf{u_b}\end{Bmatrix}=\begin{Bmatrix}\mathbf{0}\\ \mathbf{0}\end{Bmatrix}, \tag{10}$$

in which subscripts a and b refer to the outer surface S_a and inner surface S_b, respectively; matrices **U** and **T** are from the displacement and traction kernels, respectively; **u** and **t** are the displacement and traction vectors, respectively. The total number of elements and nodes on S_a are denoted by M_a and N_a, respectively, and similarly M_b and N_b for the number of elements and nodes on S_b. The two coefficient matrices in Eq. (10) are square matrices of the dimension $3N$ by $3N(N=N_a+N_b)$. Since $\mathbf{t_a}$, $\mathbf{u_a}$, and $\mathbf{t_b}$ (or $\mathbf{u_b}$) are unknowns, additional information will be needed from the acoustic domain in order to solve the coupled structural acoustic problem.

The model for the acoustic field shares the same mesh with the model for the elastic field on the interface surface.

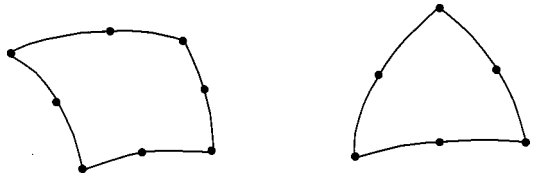


FIG. 2. Conforming quadratic boundary elements.

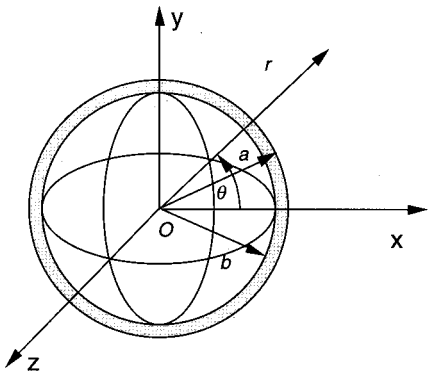


FIG. 3. A spherical shell with uniform thickness (outer radius$=a$, inner radius$=b$).

The linear system of equations obtained after discretization of the acoustic BIE (7) can be written in the following matrix form:

$$\mathbf{G}\frac{\partial \boldsymbol{\phi}}{\partial n}+\mathbf{H}\boldsymbol{\phi}=\boldsymbol{\phi}^{\mathbf{I}}, \tag{11}$$

where $\mathbf{G}$ and $\mathbf{H}$ are both N_a by N_a matrices and $\boldsymbol{\phi}^{\mathbf{I}}$ is the known vector from the incident wave. For a coupled structural acoustic problem, both $(\partial\boldsymbol{\phi}/\partial n)$ and $\boldsymbol{\phi}$ on the surface (S_a) are unknowns but are related to the unknowns of the elastic domain ($\mathbf{t_a}$ and $\mathbf{u_a}$) through the interface conditions [Eqs. (3a) and (3b)]. By using the interface conditions, two sets of these unknowns can be eliminated and the resulting linear system of equations can be expressed in matrix form as $(\partial\boldsymbol{\phi}/\partial n)$ and $\mathbf{t_a}$ are eliminated, $\mathbf{t_b}$ is assumed known),

$$\begin{bmatrix} \mathbf{H} & \mathbf{D} & \mathbf{0} \\ \mathbf{E_a} & \mathbf{T_{aa}} & \mathbf{T_{ab}} \\ \mathbf{E_b} & \mathbf{T_{ba}} & \mathbf{T_{bb}} \end{bmatrix} \begin{Bmatrix} \boldsymbol{\phi} \\ \mathbf{u_a} \\ \mathbf{u_b} \end{Bmatrix} = \begin{Bmatrix} \boldsymbol{\phi}^{1} \\ -\mathbf{U_{ab}t_b} \\ -\mathbf{U_{bb}t_b} \end{Bmatrix}, \tag{12}$$

in which $\mathbf{D}=\rho_f\omega^2\mathbf{G}\boldsymbol{\nu}$ is an N by $3N$ matrix and $\boldsymbol{\nu}$ is an N by $3N$ matrix in the following form:

$$\boldsymbol{\nu}=\begin{bmatrix} \mathbf{n}_1^T & \mathbf{0} & \mathbf{0} & & \mathbf{0} \\ \mathbf{0} & \mathbf{n}_2^T & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{n}_N^T \end{bmatrix}, \tag{13}$$

where $\mathbf{n}_\alpha^T$ is the surface normal vector at node α ($\alpha=1,2,...,N$). Also in Eq. (12), $\mathbf{E_a}=-\mathbf{U_{aa}}\boldsymbol{\nu}^T$ and $\mathbf{E_b}=-\mathbf{U_{ab}}\boldsymbol{\nu}^T$ are matrices of the dimension $3N$ by N.

III. NUMERICAL EXAMPLES

First, the developed BEM/BEM approach was tested on radiation and scattering problems using a spherical shell (Fig. 3) with outer radius $a=1$ and thickness $h=0.5$, 0.05, and 0.01 m, respectively. Quadratic elements were used (Fig. 2) on both S_a and S_b. Four BEM meshes with increasing total numbers of elements (64, 112, 160, and 306) were used. Three typical materials, steel (Young's modulus $E=2.07\times10^{11}$ Pa, Poisson's ratio $\nu=0.3$, and density $\rho=7810\,\mathrm{kg/m^3}$), aluminum ($E=7.10\times10^{10}$ Pa, $\nu=0.33$, $\rho=2700\,\mathrm{kg/m^3}$), and hard rubber ($E=2.30\times10^{9}$ Pa, $\nu=0.4$, $\rho=2117\,\mathrm{kg/m^3}$), were used for the shell structures. The surrounding acoustic media were assumed to be seawater in all the cases; the density and speed of sound used are 1026 $\mathrm{kg/m^3}$ and 1500 m/s, respectively.

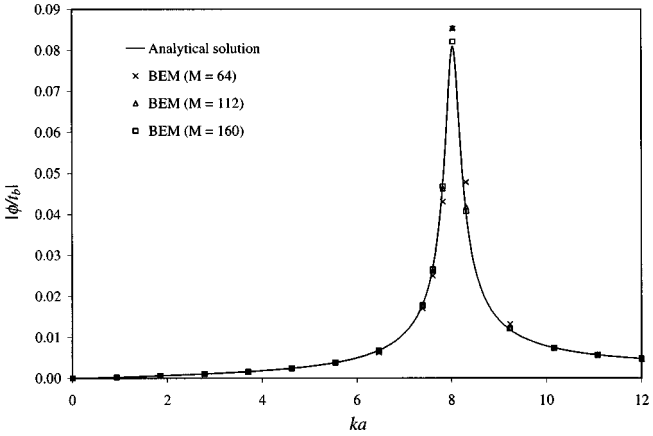


FIG. 4. Normalized radiated sound pressure from a steel spherical shell ($r=5a$, $h/a=0.5$).

For radiation problems, uniform time-harmonic pressure was applied on S_b with amplitude $t_b=1\times10^6$ N/m and the angular frequency being given. The radiated wave by a steel spherical shell with the thickness to radius ratio $h/a=0.5$ was studied first. The normalized sound pressure calculated at a distance $r=5a$ from the center of the shell is plotted in Fig. 4 versus ka for three different meshes. Very fast convergence of the BEM solution is observed as compared to the analytical solution (given in Ref. 36 where typographical errors have been corrected), although the convergence at the resonant frequency is slower than those at other frequencies. Since the mesh with $M=112$ already gives very good results except at the resonant frequency (Fig. 4), we used this mesh for the next two test cases on radiation problem. Figure 5 shows the BEM solution of the normalized radiated sound pressure plotted versus ka from a thin spherical shell ($h/a=0.01$) for three different materials. The BEM solutions match the analytical solution as expected. It is noted that the resonance occurs at lower frequency when the material of the shell is softer. The effect of the thickness of a steel spherical shell on the radiated sound field is shown in Fig. 6, where the

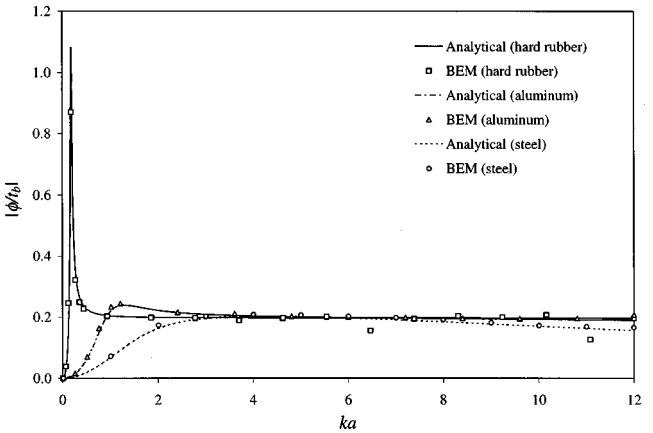


FIG. 5. Normalized radiated sound pressure from a thin spherical shell of different material ($r=5a$, $h/a=0.01$, $M=112$).

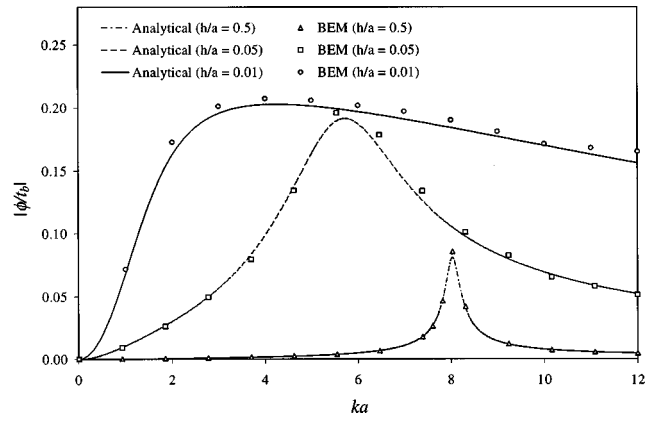


FIG. 6. Normalized radiated sound pressure from steel spherical shells with different thickness ($r=5a$, $M=112$).

normalized radiated sound pressure at $r=5a$ is plotted versus ka.

For scattering problems, an elastic spherical shell impinged upon by an incident wave ϕ^I traveling in the positive x direction was considered. The inner surface of the spherical shell is assumed traction-free. The BEM solution using the mesh with $M=112$ is found to be a good approximation. It is therefore used in the test cases for scattering problem. First, the effect of thickness on the backscattering by the spherical shell is studied. A steel spherical shell with $h/a=0.05$ and $h/a=0.01$ was tested. Figure 7 shows the BEM solutions of the normalized backscattered sound pressure at $r=5a$ plotted versus ka. The membrane solutions[2] are also shown in the figure for comparison. The membrane model assumes that flexural stresses are negligible as compared to membrane stresses. This means that the membrane solution is a good approximation only when the shell is thin enough and the frequency is low. It can be seen that the BEM solution agrees with the membrane solution at low frequencies even when resonant frequencies are involved. At higher frequencies, large deviation can be observed near the resonant frequencies. Since the BEM solution is based on rigorous 3D elastodynamics, it is considered more accurate than the membrane solution.

The effect of different composition of the shell was studied next using a hard rubber spherical shell and a steel

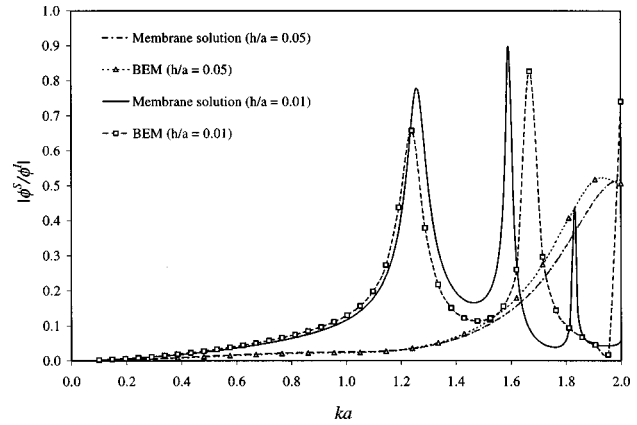


FIG. 7. Normalized backscattered sound pressure by a steel spherical shell with varying thickness ($r=5a$, $M=112$).

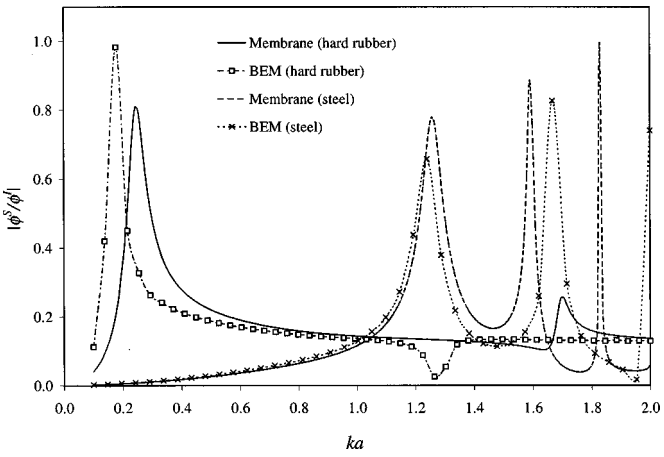


FIG. 8. Normalized backscattered sound pressure by a thin spherical shell of different materials ($r=5a$, $h/a=0.01$, $M=112$).

spherical shell with $h/a=0.01$. Figure 8 shows the normalized backscattered sound pressure at $r=5a$ plotted versus ka. It can be seen that the first resonant frequency for the hard rubber shell occurs at a lower frequency than that of the steel shell.

A test case on spherical shell with nonuniform thickness (Fig. 9) was performed with no additional modeling effort involved. With its outer surface and inner surface forming a sphere and a spheroid, respectively, the thickness of the spherical shell varies. The three axes of the spheroid were denoted by b_x, b_y, and b_z, respectively (Fig. 9). The radiated wave from a steel shell with $b_x=0.95a$ and $b_y=b_z=0.99a$ was calculated with uniform pressure t_b applied on the inner surface. Figure 10 shows the normalized radiated pressure at $r=5a$ in the direction of $\theta=0$ and 90 deg at frequencies up to $ka=12$. Two meshes with $M=112$ and $M=160$ are used in the calculation. The result calculated with 112 elements is shown since the mesh with $M=160$ provided only little improvement. The results from uniform thickness spherical shell with $h/a=0.05$ and 0.01 are shown in the same figure for comparison. It can be seen that the result for the nonuniform thickness shell in both directions approximates the average of the results from the two uniform thickness shells at low frequencies ($ka<3$). At higher frequencies, however, significant differences between the results for the two directions can be observed.

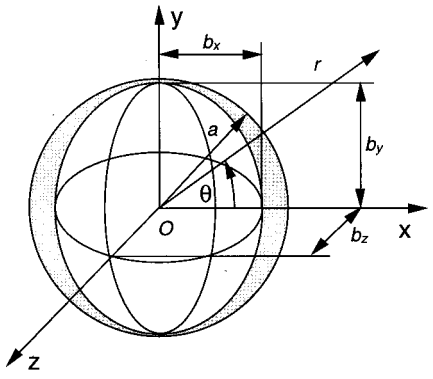


FIG. 9. A spherical shell with nonuniform thickness.

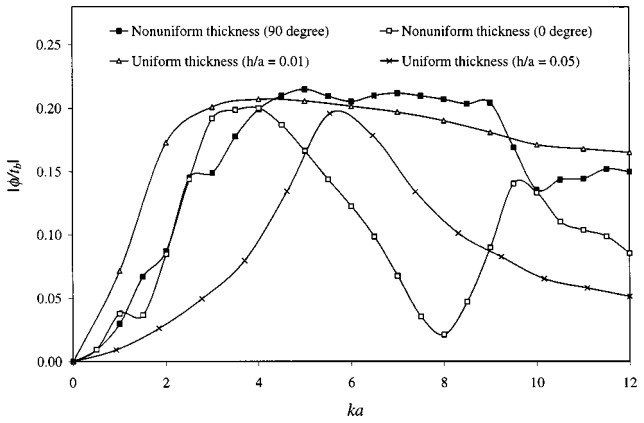


FIG. 10. Normalized radiated sound pressure from a steel spherical shell with nonuniform thickness ($r=5a$, $M=112$, $b_x=0.95a$, $b_y=b_z=0.99a$).

Finally, scattering from a cylindrical (capsulelike) thin shell (Fig. 11) made of steel is studied with a total of 216 elements (626 nodes). The incident wave is from the x-direction and the sound pressures from backscattering (at $x=-10\,\mathrm{m}$, $y=z=0$) and forward scattering (at $x=10\,\mathrm{m}$, $y=z=0$) from the elastic shell are plotted in Fig. 12. For comparisons, the results from the acoustics BEM[15] (assuming the shell to be rigid) and the current coupled BEM with large Young's modulus and density for the shell are also plotted in the figure. These latter two results agree very well for the rigid case, as expected. For forward scattering, the result for steel shell using the coupled BEM oscillates around the data for the rigid shell case. However, the result for backscattering from the steel shell using the coupled BEM differs significantly from the corresponding rigid shell case, especially when the frequencies are higher. This further signifies the necessity of a coupled analysis for structural acoustic problems.

IV. DISCUSSIONS

A unified BEM/BEM approach to sound–structure interaction problems in the frequency domain for shell-like structures is developed in this paper. The formulation is valid for general loading conditions and for all frequencies. The Burton and Miller composite BIE formulation is employed to overcome the fictitious eigenfrequency difficulty and the thin-shape breakdown for the acoustic domain. The hypersingular integrals involved are transformed into the weakly singular forms and evaluated by ordinary quadrature. The nearly singular integrals existing in the 3D elastodynamic BIE for thin-shell-like structures are treated by semianalytical methods and evaluated accurately. Numerical examples on radiation and scattering problems from bulky and thin spherical and cylindrical shells (including ones with nonuniform thickness) clearly demonstrate the effectiveness and accuracy of the developed approach.

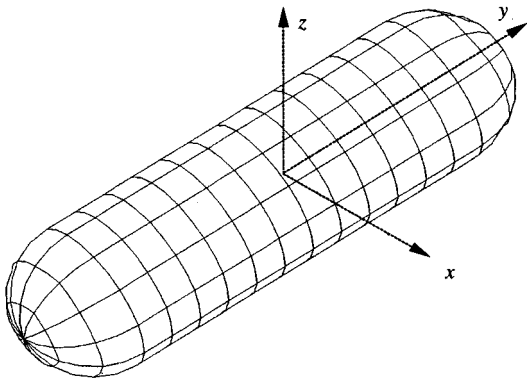


FIG. 11. A cylindrical (capsulelike) shell structure with thickness=0.01, radius=1.0, and total length=7.0 m.

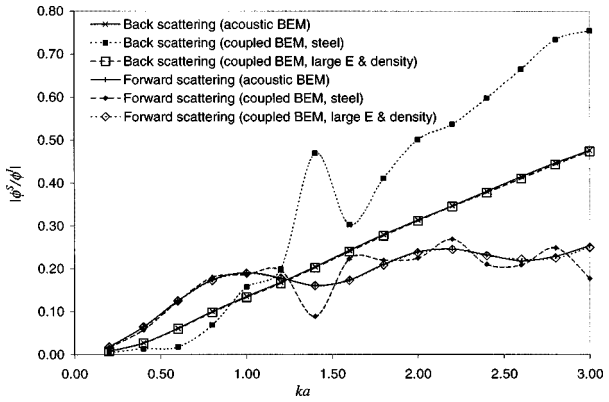


FIG. 12. Scattered sound pressure from the capsulelike shell structure.

There are many advantages in the developed BEM/BEM approach to the structural acoustic analysis. First of all, this approach renders high accuracy for both acoustic and elastic domains due to the semianalytical nature of the BEM. Second, the coupling effect is modeled effectively and efficiently by sharing the same surface mesh on the interface between the elastic domain and the acoustic domain. Moreover, the unified BEM approach is much easier in modeling than the other methods for structures with complicated features in either the interior (e.g., stiffeners) or the exterior (e.g., rudders, turbo blades of a submarine). Multiple scatterers (e.g., an array of shell structures) can also be modeled readily by the BEM with the multidomain technique. Finally, shell structures with nonuniform thickness or coatings (layered shell structures) can be handled accurately by the BIE formulations. The remaining major issue is the computational efficiency, as the BEM usually generates fully populated matrices, although of smaller size than those of the matrices from the FEM/IEM approach. With the recent development of iterative solvers for asymmetric and dense complex systems (see, e.g., Ref. 41), which can dramatically increase the speed of solving large linear systems, this efficiency concern may be eased in the near future.

Studies on efficient solution techniques, including iterative solvers, and more complicated shell structures using the developed BEM/BEM approach to structural acoustics are underway and the results will be reported in subsequent papers. Multiple scatterers and layered shell structures (shells with coatings) will be interesting and challenging future research topics.

ACKNOWLEDGMENTS

The second author (Y.J.L.) would like to thank Professor Frank J. Rizzo at Iowa State University for many helpful

discussions on the BIE/BEM for acoustics and elastodynamics. Both authors would like to acknowledge the partial support by the University Research Council of the University of Cincinnati.

[1] M. C. Junger, ''Sound scattering by thin elastic shells,'' J. Acoust. Soc. Am. **24**, 366–373 (1952).
[2] M. C. Junger and D. Feit, *Sound, Structure, and Their Interaction* (MIT, Cambridge, MA, 1972).
[3] M. C. Junger, ''Acoustic fluid-elastic structure interactions: Basic concepts,'' Comput. Struct. **65**, 287–293 (1997).
[4] G. C. Everstine, ''Finite element formulations of structural acoustics problems,'' Comput. Struct. **65**, 307–321 (1997).
[5] D. S. Burnett, ''A three-dimensional acoustic infinite element based on a prolate spheroidal multipole expansion,'' J. Acoust. Soc. Am. **96**, 2798–2816 (1994).
[6] D. S. Burnett and R. L. Holford, ''Prolate and oblate spheroidal acoustic infinite elements,'' Comput. Methods Appl. Mech. Eng. **158**, 117–141 (1998).
[7] H. A. Schenck, ''Improved integral formulation for acoustic radiation problems,'' J. Acoust. Soc. Am. **44**, 41–58 (1968).
[8] A. J. Burton and G. F. Miller, ''The application of integral equation methods to the numerical solution of some exterior boundary-value problems,'' Proc. R. Soc. London, Ser. A **323**, 201–210 (1971).
[9] W. L. Meyer, W. A. Bell, B. T. Zinn, and M. P. Stallybrass, ''Boundary integral solutions of three-dimensional acoustic radiation problems,'' J. Sound Vib. **59**, 245–262 (1978).
[10] T. Terai, ''On calculation of sound fields around three-dimensional objects by integral equation methods,'' J. Sound Vib. **69**, 71–100 (1980).
[11] A. F. Seybert, B. Soenarko, F. J. Rizzo, and D. J. Shippy, ''An advanced computational method for radiation and scattering of acoustic waves in three dimensions,'' J. Acoust. Soc. Am. **77**, 362–368 (1985).
[12] C. C. Chien, H. Rajiyah, and S. N. Atluri, ''An effective method for solving the hypersingular integral equations in 3-D acoustics,'' J. Acoust. Soc. Am. **88**, 918–937 (1990).
[13] T. W. Wu, A. F. Seybert, and G. C. Wan, ''On the numerical implementation of a Cauchy principle value integral to insure a unique solution for acoustic radiation and scattering,'' J. Acoust. Soc. Am. **90**, 554–560 (1991).
[14] Y. J. Liu and F. J. Rizzo, ''A weakly-singular form of the hypersingular boundary integral equation applied to 3-D acoustic wave problems,'' Comput. Methods Appl. Mech. Eng. **96**, 271–287 (1992).
[15] Y. J. Liu and S. Chen, ''A new form of the hypersingular boundary integral equation for 3-D acoustics and its implementation with C^0 boundary elements,'' Comput. Methods Appl. Mech. Eng. **173**, 375–386 (1999).
[16] Y. J. Liu and F. J. Rizzo, ''Hypersingular boundary integral equations for radiation and scattering of elastic waves in three dimensions,'' Comput. Methods Appl. Mech. Eng. **107**, 131–144 (1993).
[17] G. Krishnasamy, F. J. Rizzo, and Y. J. Liu, ''Scattering of acoustic and elastic waves by crack-like objects: The role of hypersingular integral equations,'' in *Review of Progress in Quantitative Nondestructive Evaluation* (Plenum, Brunswick, Maine, 1991).
[18] G. Krishnasamy, F. J. Rizzo, and Y. J. Liu, ''Boundary integral equations for thin bodies,'' Int. J. Numer. Methods Eng. **37**, 107–121 (1994).
[19] Y. J. Liu and F. J. Rizzo, ''Scattering of elastic waves from thin shapes in three dimensions using the composite boundary integral equation formulation,'' J. Acoust. Soc. Am. **102**(2), 926–932 (1997).
[20] R. Martinez, ''The thin-shape breakdown (TSB) of the Helmholtz integral equation,'' J. Acoust. Soc. Am. **90**, 2728–2738 (1991).
[21] F. J. Rizzo, ''An integral equation approach to boundary value problems of classical elastostatics,'' Q. Appl. Math. **25**, 83–95 (1967).
[22] T. A. Cruse and F. J. Rizzo, ''A direct formulation and numerical solution of the general transient elastodynamic problem, I,'' J. Math. Anal. Appl. **22**, 244–259 (1968).
[23] T. A. Cruse, ''A direct formulation and numerical solution of the general transient elastodynamic problem, II,'' J. Math. Anal. Appl. **22**, 341–355 (1968).
[24] T. A. Cruse, ''An improved boundary-integral equation method for three-dimensional elastic stress analysis,'' Comput. Struct. **4**, 741–754 (1974).
[25] F. J. Rizzo and D. J. Shippy, ''An advanced boundary integral equation method for three-dimensional thermoelasticity,'' Int. J. Numer. Methods Eng. **11**, 1753–1768 (1977).
[26] F. J. Rizzo, D. J. Shippy, and M. Rezayat, ''A boundary integral equation method for radiation and scattering of elastic waves in three dimensions,'' Int. J. Numer. Methods Eng. **21**, 115–129 (1985).
[27] M. Rezayat, D. J. Shippy, and F. J. Rizzo, ''On time-harmonic elastic-wave analysis by the boundary element method for moderate to high frequencies,'' Comput. Methods Appl. Mech. Eng. **55**, 349–367 (1986).
[28] T. A. Cruse, *Boundary Element Analysis in Computational Fracture Mechanics* (Kluwer Academic, Dordrecht, 1988).
[29] Y. J. Liu, ''Analysis of shell-like structures by the boundary element method based on 3-D elasticity: Formulation and verification,'' Int. J. Numer. Methods Eng. **41**, 541–558 (1998).
[30] Y. J. Liu, D. Zhang, and F. J. Rizzo, ''Nearly singular and hypersingular integrals in the boundary element method,'' in *Boundary Elements XV* (Computational Mechanics, Worcester, MA, 1993), pp. 453–468.
[31] J. F. Luo, Y. J. Liu, and E. J. Berger, ''Analysis of two-dimensional thin structures (from micro- to nano-scales) using the boundary element method,'' Comput. Mech. **22**, 404–412 (1998).
[32] I. C. Mathews, ''Numerical techniques for three-dimensional steady-state fluid-structure interaction,'' J. Acoust. Soc. Am. **79**, 1317–1325 (1986).
[33] G. C. Everstine and F. M. Henderson, ''Coupled finite element/boundary element approach for fluid structure interaction,'' J. Acoust. Soc. Am. **87**, 1938–1947 (1990).
[34] L. Demkowicz and J. T. Oden, ''Application of hp-adaptive BE/FE methods to elastic scattering,'' Comput. Methods Appl. Mech. Eng. **133**, 287–317 (1996).
[35] P. P. Goswami, T. J. Rudolphi, F. J. Rizzo, and D. J. Shippy, ''A boundary element model for acoustic-elastic interaction with applications in ultrasonic NDE,'' J. Nondestruct. Eval. **9**, 101–112 (1990).
[36] A. F. Seybert, T. W. Wu, and X. F. Wu, ''Radiation and scattering of acoustic waves from elastic solids and shells using the boundary element method,'' J. Acoust. Soc. Am. **84**, 1906–1912 (1988).
[37] Q. Huang and T. A. Cruse, ''On the non-singular traction-BIE in elasticity,'' Int. J. Numer. Methods Eng. **37**, 2041–2072 (1994).
[38] T. A. Cruse and J. D. Richardson, ''Non-singular Somigliana stress identities in elasticity,'' Int. J. Numer. Methods Eng. **39**, 3273–3304 (1996).
[39] J. D. Richardson, T. A. Cruse, and Q. Huang, ''On the validity of conforming BEM algorithms for hypersingular boundary integral equations,'' Comput. Mech. **20**, 213–220 (1997).
[40] P. A. Martin, F. J. Rizzo, and T. A. Cruse, ''Smoothness-relaxation strategies for singular and hypersingular integral equations,'' Int. J. Numer. Methods Eng. **42**, 885–906 (1998).
[41] R. W. Freund and N. M. Nachtigal, ''QMRPACK: A package of QMR algorithms,'' ACM Trans. Math. Softw. **22**, 46–77 (1996).

The beam forming properties of a concave spherical reflector with an on-axis receiver

Grant B. Deane
Marine Physical Laboratory, Scripps Institution of Oceanography, UCSD, La Jolla, California 92093-0238

(Received 16 April 1998; revised 12 May 1999; accepted 10 June 1999)

A wave-theoretical formulation for the on-axis scattering of a concave, spherical reflector with a pressure release surface is presented. Expressions for the reflector gain and half-power beam width are derived using a stationary phase analysis of the radiation integral arising from the Kirchhoff–Helmholtz scalar diffraction theory. The gain is shown to scale with frequency and the beam width scales inversely with frequency. The results of the analysis compare well with data collected from a 3-m spherical dish in both air and underwater calibrations. © *1999 Acoustical Society of America.* [S0001-4966(99)04809-2]

PACS numbers: 43.20.Fn, 43.58.Ls [AN]

INTRODUCTION

The subject of this paper is a derivation of the on-axis gain and half-power beam width of a concave spherical reflector with a pressure release surface, insonified with plane waves. The motivation for the work was provided by a need to model the acoustical properties of a 3-m spherical-section dish used to focus ambient noise in the Acoustic Daylight Ocean Noise Imaging System (ADONIS).[1,2] The ADONIS system was built to explore the idea of forming acoustical images of objects in the ocean using the background ambient noise field, a concept which appeared in the literature a decade or so ago, and has been developed and studied by Buckingham and his colleagues.[3–6] Although geometrical treatments of spherical reflectors have been published (see Spencer and Hyde[7] and the references therein for a treatment of the focal region of a spherical reflector using geometrical optics), a survey of the literature failed to reveal analytical expressions for the gain and half-power beam width of a spherical reflector when the aperture of the dish may be only a few wavelengths across, and these are developed here.

There is an extensive electromagnetics literature dealing with the design and analysis of reflecting antennas. Current methods employ various computational techniques for evaluating the vector radiation integral for electromagnetic fields (see Scott[8] for a review of modern computational methods and Tsai *et al.*[9] for a comparison of the geometrical theory of diffraction and integral equation formulation for the analysis of reflector antennas). Because the present application is acoustical, and as we require closed-form, analytical expressions for the properties of the spherical reflector, our analysis is based on the scalar diffraction theory of Kirchhoff and Helmholtz (see Rusch and Potter,[10] p. 33, for example). This method of analysis is similar to that used to study the acoustic radiation patterns of transducers, as discussed by de Hoop *et al.*[11]

The material in the paper is arranged as follows. The theory and assumptions leading to a one-dimensional integral representation for the on-axis acoustic field scattered from a concave, pressure release, spherical-section reflector are presented in Sec. I. The asymptotic analysis of this integral in terms of large kR_0, where k is the wave number of the incident radiation and R_0 is the radius of the reflector, is developed in the Appendix. A physical interpretation of the asymptotic analysis and scaling laws for the on-axis reflector gain and half-power beam width are derived in Sec. II. A brief discussion of how reducing the solid angle subtended by the dish affects the focusing properties is given in Sec. III and the scattering theory is compared with experimental data in Sec. IV. Section V briefly discusses the focusing properties of the dish at a receiver point which lies off the dish axis, and a paper summary is given in Sec. VI.

I. THEORETICAL FORMULATION

The geometry for the scattering calculation is shown in Fig. 1. We will use whichever of spherical polar and Cartesian coordinates is appropriate for a given part of the analysis. The Cartesian coordinate variables are (x,y,z) with the x-y plane parallel to the dish rim. The spherical polar coordinates are (R,θ,ϕ). In spherical polar coordinates, the dish surface is defined by $R=R_0$ and $0\leqslant\theta\leqslant\alpha$, where α is a positive constant satisfying $0<\alpha\leqslant\pi/2$. An element of area on the surface of the dish is given by $d\underline{\Gamma}=\hat{\underline{n}}_\Gamma R_0^2\sin(\theta)d\theta\,d\phi$, where $\hat{\underline{n}}_\Gamma$ is a unit vector normal to the dish surface, pointing outward, away from the origin of the spherical section.

The Kirchhoff–Helmholtz integral theorem yields an expression for the field scattered out of an incident wave field by the dish. The pressure at the point z_r on the z-axis is

$$p(z_r)=-\frac{1}{4\pi}\int\int_\Gamma(G\nabla p-p\nabla G)\cdot d\underline{\Gamma}, \tag{1}$$

where the double integral is evaluated over the shell surface, a dot denotes the inner product between two vectors, $d\underline{\Gamma}$ is an elemental area vector on the shell (vectors will always be denoted by an underbar) and G is the free space Green's function given by

$$G=\frac{\exp(ik\overline{R})}{\overline{R}}, \tag{2}$$

where k is the incident field wave number and

$$\overline{R}^2=R_0^2\left(1+\frac{z_r^2}{R_0^2}-2\frac{z_r}{R_0}\cos(\theta)\right) \tag{3}$$

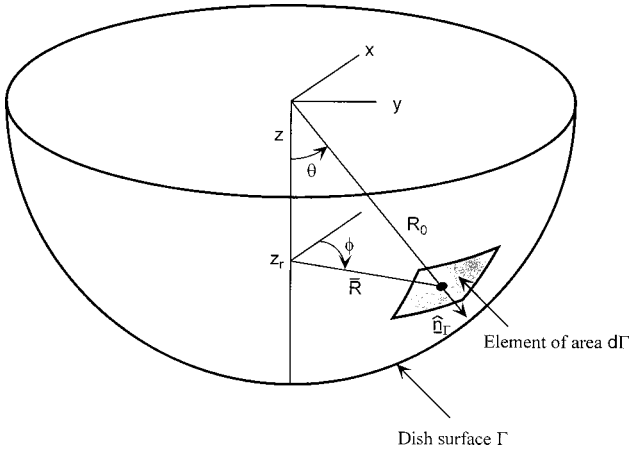


FIG. 1. The geometry of the spherical dish section and coordinate variables used in the scattering calculations.

is the square of the distance from the elemental area to the field evaluation point (see Fig. 1). We will assume the shell is a pressure release boundary, in which case $p=0$ on the shell and the second term in the integrand of Eq. (1) vanishes. As the nondimensional ratio z_r/R_0 plays an important role in the subsequent analysis, we introduce the normalized axial distance variable

$$\beta=\frac{z_r}{R_0}. \tag{4}$$

Then $\bar{R}=R_0\sqrt{s}$ where

$$s=1+\beta^2-2\beta\cos(\theta). \tag{5}$$

The beam forming properties of the dish can be calculated by considering an harmonic, plane wave with a propagation vector $\underline{k}$ incident on the shell. The pressure field that accompanies the wave is

$$p_i(\underline{R})=\exp(i\underline{k}\cdot\underline{R}), \tag{6}$$

where $\underline{R}=x\hat{\underline{x}}+y\hat{\underline{y}}+z\hat{\underline{z}}$ and $\hat{\underline{x}}$, $\hat{\underline{y}}$, and $\hat{\underline{z}}$, respectively, are unit vectors in the x, y and z directions.

An expression for the pressure on the shell boundary due to the incident wave field can be formulated as follows. Consider an element of area on the shell $d\underline{\Gamma}$. Locally, the surface acts as a planar, specular reflector and the pressure field can be expressed as the sum of incident and reflected plane wave fields. Since the shell is part of a spherical surface, the vector $\underline{R}$ is normal to the element of shell area to which it points, $d\underline{\Gamma}$. Let $\underline{R}_0$ be a vector with magnitude R_0 and pointing in the same direction as $\underline{R}$. Then the vector $\underline{R}-\underline{R}_0$ is also normal to $d\underline{\Gamma}$ and the sum of incident and reflected fields can be expressed as

$$p(\underline{R})=\exp(-i\underline{k}\cdot\underline{R}_0)\times[\exp(i\underline{k}\cdot(\underline{R}-\underline{R}_0))-\exp(-i\underline{k}\cdot(\underline{R}-\underline{R}_0))]. \tag{7}$$

The leading term on the right hand side of Eq. (7) accounts for the phase of the field on the surface relative to the coordinate system origin. Substituting Eqs. (7) and (2) into Eq. (1), we obtain an expression for the scattered pressure at z_r in terms of a double integral over the shell surface:

$$p(z_r)=\frac{ikR_0^2}{2\pi}\int_0^{\alpha}\int_0^{2\pi}\frac{\exp(ikR_0\sqrt{s})}{kR_0\sqrt{s}}\exp(i\underline{k}\cdot\underline{R}_0)\underline{k}\cdot\hat{\underline{n}}_\Gamma\sin(\theta)\,d\phi\,d\theta, \tag{8}$$

where α defines the angular extent of the shell. The remainder of the problem consists of evaluating this double integral.

Without loss of generality, we can assume that the incident plane wave shows no phase variation along the y-axis. Using Eq. (8) for the scattered field and evaluation of the resulting integral over ϕ yields, after some algebra, an expression for the scattered field pressure at z_r on the dish axis

$$p(z_r)=ik^2R_0^2\int_0^{\alpha}\frac{\exp(ikR_0\sqrt{s})}{kR_0\sqrt{s}}\times\exp(ikR_0\cos(\theta_0)\cos(\theta))\sin(\theta)\times\{\cos(\theta_0)\cos(\theta)J_0(kR_0|\sin(\theta_0)|\sin(\theta))-i|\sin(\theta_0)|\sin(\theta)J_1(kR_0|\sin(\theta_0)|\sin(\theta))\}d\theta, \tag{9}$$

where J_0 and J_1 are Bessel functions of the first kind, respectively of order 0 and 1, and $|\ |$ denotes the absolute value of the argument. Note that, from Eq. (5), s is a function of θ. Equation (9) is our final expression for the pressure field scattered by the dish. The two-dimensional integral has been reduced to a form which is suitable for numerical evaluation or asymptotic analysis, which are the topics of the next section.

II. APPLICATION OF THE ASYMPTOTIC ANALYSIS

An asymptotic analysis of the integral for the scattered field in terms of kR_0 is presented in the Appendix. The asymptotic analysis is a useful tool in interpreting the physics of the dish response and deriving scaling laws, and we will draw on the results of the analysis throughout the remaining sections.

A. Physical interpretation of the stationary phase points

We begin by considering the response of the dish to a normally incident plane wave, so that $\theta_0=0$. As discussed in the Appendix, there are two points of stationary phase which contribute to the integral for the scattered field. These are $\theta_1=0$ and

$$\cos(\theta_2)=\frac{1}{2\beta}. \tag{10}$$

Equation (10) has a simple physical interpretation; it describes the condition for the specular reflection of energy normally incident on the dish through the point z_r on the z-axis.

The contribution from the stationary point $\theta_1=0$ corresponds to energy reflected from the base of the dish, back along the z-axis. Energy following this path forms an interference pattern with the specularly reflected energy. At z_r

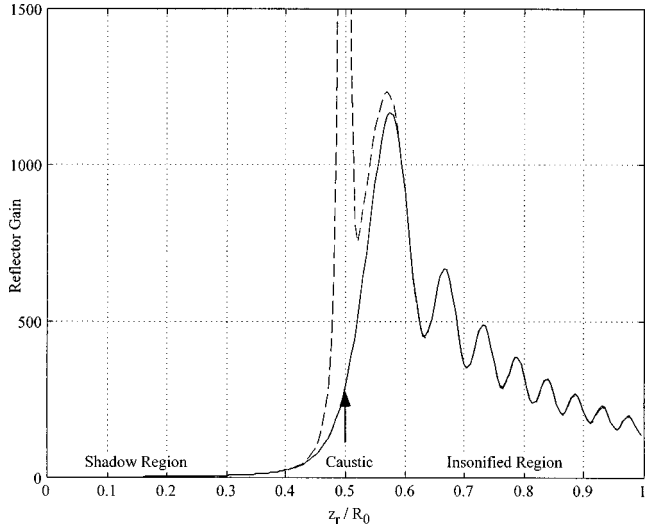


FIG. 2. The dish gain versus normalized axial distance for $kR_0=100$. The solid line was calculated by numerical integration of the scattering integral. The broken line shows the asymptotic expansion of the integral. The first-order asymptotics fail at $z_r/R_0=0.5$ where a caustic is formed in the scattered field. The formation of a shadow for $z_r/R_0<0.5$ can be seen.

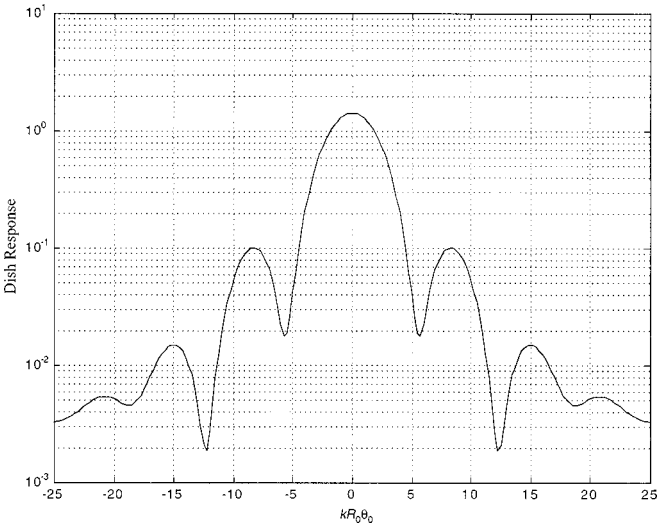


FIG. 3. The beam forming response of the dish as a function of $kR_0\theta_0$. The dish gain was calculated for $kR_0=100$ at $z_r/R_0=0.576$ where the gain is a maximum.

$=R_0/2$, corresponding to $\beta=1/2$, the two stationary points merge and the first-order asymptotic analysis fails.

B. The dish gain

For reasons discussed in the Appendix, the contributions to the scattered field from the two stationary phase points are of different order in terms of the expansion variable kR_0, with the specularly reflected energy from θ_2 dominating. To leading order in kR_0 the dish gain $A=p_r p_r^*$ is given by the squared magnitude of the specularly reflected energy. From Eq. (A12)

$$A=\frac{\pi kR_0}{2\beta^3}, \tag{11}$$

showing that the dish gain scales with frequency. Equation (11) is valid provided θ_2 lies in the path of integration, which limits values of β to $1/2<\beta<1/(2\cos\alpha)$.

Figure 2 shows the dish gain as a function of β for $kR_0=100$ and $\alpha=\pi/2$. The solid curve is a numerical calculation based on a Simpson's rule evaluation of the integral in Eq. (A1), and is essentially exact. The broken line shows the stationary phase point contributions to the integral, which amounts to p_1 for $0<\beta<1/2$ and p_1+p_2 for $1/2<\beta<1$, where p_1 and p_2, respectively, are given by Eqs. (A10) and (A12).

The agreement between the stationary phse analysis and the numerical evaluation of Eq. (A1) is very good except around $\beta=1/2$. As $\beta\to1/2$, the stationary phase points θ_1 and θ_2 merge and a caustic is formed in the field. Equation (A10) for p_1 becomes singular at the caustic, and the first-order asymptotic analysis fails. The caustic separates an insonified region for which $1/2<\beta<1$ from a shadow region defined by $0\leqslant\beta<1/2$. A maximum occurs in the dish gain in the insonified region just before the caustic at a value of β which depends on kR_0. The larger the value of kR_0, the closer the maximum lies to $\beta=1/2$. The oscillations in gain in the insonified region are caused by the interference between the specular reflections and the energy reflected from the base of the dish.

C. The dish beam forming response

As noted above, the on-axis response of the dish to normally incident energy has a maximum in the insonified region just before the caustic formed at $\beta=1/2$. As θ_0 increases, the scattered field intensity drops forming a beam about $\theta_0=0$. Figure 3 shows a direct numerical integration of Eq. (9) as a function of θ_0 for $kR_0=100$ and $\beta=0.576$ where the gain is a maximum. Anticipating the result that beam width scales inversely with kR_0, the abscissa has units of $kR_0\theta_0$. The formation of a beam about $\theta_0=0$ is quite evident.

We can derive an analytical expression for the dish beam width from Eq. (A24) which describes the beam forming response of the dish under the conditions $kR_0\gg1$ and $kR_0|\theta_0|\ll1$. The leading order term multiplying θ_0^2 in Eq. (A24) is the factor containing $k^2R_0^2$. Retaining only this term, we obtain the approximation

$$A\approx\overline{p_2}\overline{p_2^*}\approx\frac{\pi kR_0}{2\beta^3}\left(1-\frac{k^2R_0^2}{4}\left(1-\frac{1}{4\beta^2}\right)\theta_0^2\right)^2. \tag{12}$$

Let Θ be the half-power beam width of the dish. Then $A(\Theta)=A(0)/2$, and, from Eq. (12),

$$\Theta=\frac{4\beta\left(1-\frac{1}{\sqrt{2}}\right)}{kR_0(4\beta^2-1)^{1/2}}, \tag{13}$$

where $1/2<\beta<1$. From Eq. (13) we can deduce that the dish beam width is inversely proportional to frequency. Note that the beam width increases as β approaches 0.5 from above and there is a trade-off between increasing dish gain and increasing beam width as the caustic at $\beta=1/2$ is approached.

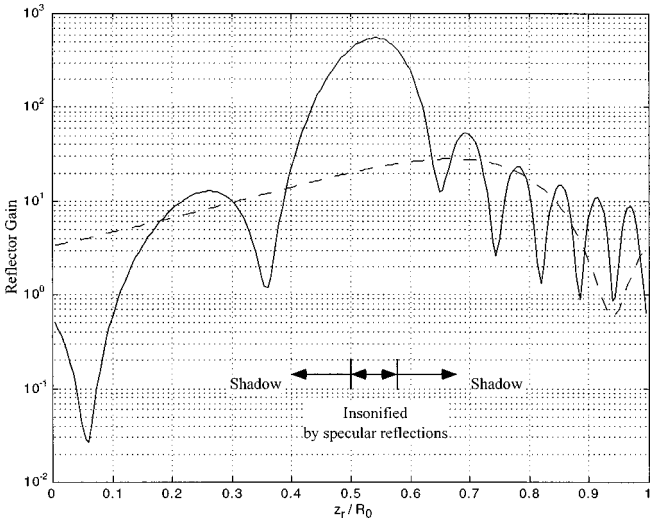


FIG. 4. The dish gain versus normalized axial distance for $\alpha=\pi/6$. The small dish has two shadow regions for $z_r/R_0<0.5$ and $z_r/R_0>1/(2\cos\alpha)=1/\sqrt{3}$. The solid and broken lines respectively correspond to kR_0 equal to 100 and 20.

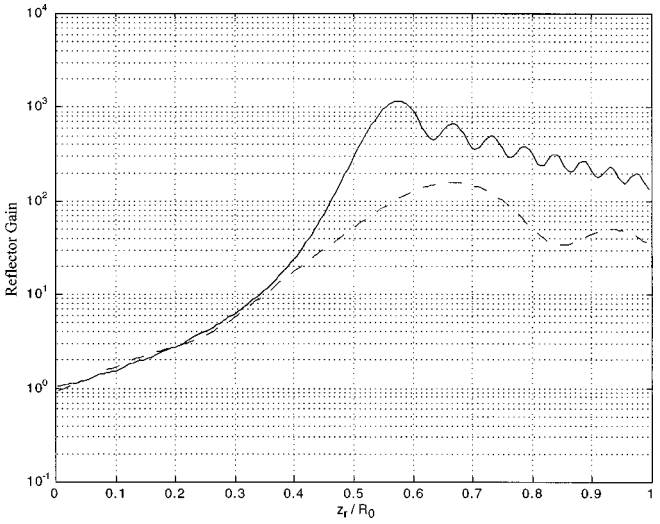


FIG. 5. The dish gain versus normalized axial distance for $\alpha=\pi/2$ for comparison with Fig. 4. Only one shadow region for $z_r/R_0<0.5$ is formed by the hemispherical dish. The solid and broken lines respectively correspond to kR_0 equal to 100 and 20.

An asymptotic expression for the dish side-lobe response could be obtained by replacing the Bessel functions in Eq. (A17) with their large-argument asymptotic forms and performing a stationary phase analysis. Such a derivation would have little practical use as the stationary points involve transcendental equations which need to be solved numerically.

III. THE EFFECT OF REDUCING THE DISH SIZE

The previous section considered the scattering response of a dish with $\alpha=\pi/2$, which corresponds to a hemisphere. A dish constructed for practical applications, however, is likely to subtend a smaller angle, which leads to a consideration of the scattering properties of a dish with $\alpha<\pi/2$. The effects of decreasing α can be discussed in terms of the asymptotic analysis presented in the Appendix. The main effect is to reduce the range of β for which there are specular reflections. When $z_r=R_0/2$, $\beta=1/2$ and for normally incident energy, the stationary phase angle $\theta_2=0$. As β increases, θ_2 increases until, at $\beta=1/(2\cos\alpha)$, the stationary point reaches the rim of the dish. Any further increase in β results in the stationary point moving beyond the dish rim and an acoustic shadow is formed. Thus the net effect is the introduction of a second shadow region close to the dish surface, in addition to the shadow formed when $\beta<1/2$ (by the term ''shadow'' we mean there are no specular reflections contributing to the scattered field).

Reducing the dish size also introduces a nonzero end point contribution from $\theta=\alpha$, denoted by p_3 and corresponding to energy reflected from the dish rim [see Eq. (A13) in the Appendix].

Figure 4 shows the dish gain versus β for $kR_0=100$ and 20, and $\alpha=\pi/6$. The curves were calculated by a Simpson's rule evaluation of the integral in Eq. (A1). Figure 5 shows the same curves calculated for a dish with $\alpha=\pi/2$ for comparison. The region of the dish axis that is insonified by specular reflections from the dish surface is labeled ''insonified by specular reflections.'' The oscillations within the two shadow regions are caused by interference between the energy reflected from the base of the dish and the energy scattered from the dish rim. The contribution from the dish rim is identically zero when $\alpha=\pi/2$, which is why the oscillations are absent in the shadow region of Figs. 2 and 5.

IV. COMPARISON WITH EXPERIMENT

Figure 6 shows a comparison of a numerical evaluation of the scattering integral and the measured response of the ADONIS dish.[1,2] The ADONIS dish is a concave, spherical reflector with $R_0=3$ m and $\alpha\approx\pi/6$, designed for passive acoustic imaging using underwater ambient noise. The re-

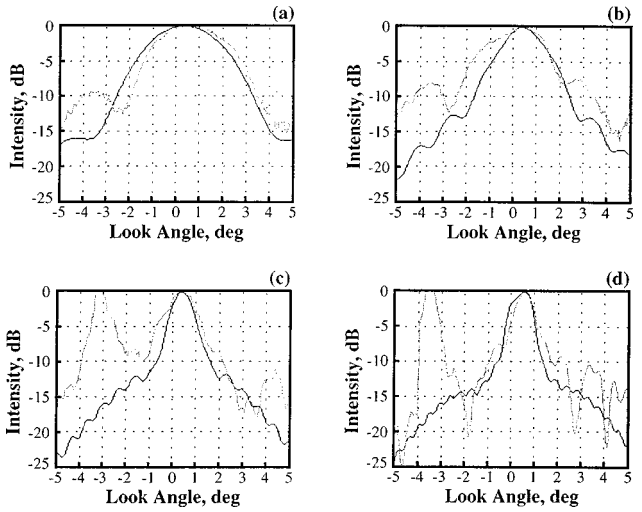


FIG. 6. Comparison between the scattering calculations and a measurement of the ADONIS dish response taken underwater during the ORB 2 experiments. The solid line is a numerical evaluation of the scattering integral and the broken line is the measured response. The dish parameters are $R_0=3$ m and $\alpha\approx\pi/6$. The four source frequencies of 10 kHz, 25 kHz, 50 kHz and 75 kHz correspond to (a) $kR_0=126$, (b) $kR_0=314$, (c) $kR_0=630$ and (d) $kR_0=945$. The large spikes at -3 deg in (c) and (d) were caused by boat noise contamination. (Reproduced with permission from Epifanio, Ref. 1).

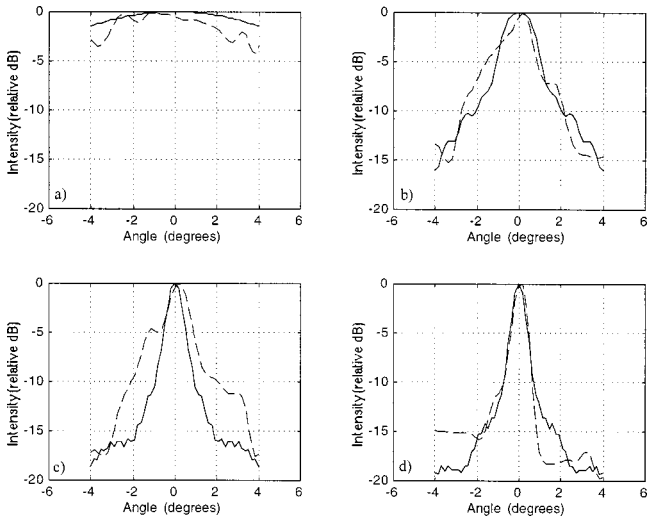


FIG. 7. Comparison between the scattering calculations and a measurement of the ADONIS dish response taken in air. The solid line is a numerical evaluation of the scattering integral and the broken line is the measured response. The four source frequencies of 0.6 kHz, 8.1 kHz, 16.8 kHz and 24.3 kHz correspond to (a) $kR_0=34$, (b) $kR_0=464$, (c) $kR_0=967$ and (d) $kR_0=1400$.

flecting face of the dish is covered in neoprene sections. The air enclosed by the neoprene gives the dish surface a pressure release surface, which is the boundary condition considered for the scattering analysis. The dish calibration data was collected as part of the ORB 2 experiments described by Epifanio.[1] The curves are plotted in dB relative to the sound intensity at the beam peak, and show the response of the dish to a 40-m distant spherical sound source. The scattering theory has accurately reproduced the systematic reduction in beam width through the four frequencies shown. Detailed agreement cannot be argued for the side-lobe structure where boat noise contamination has interfered with the calibrations.

The ADONIS dish was also calibrated with an air experiment. The dish was mounted on Scripps Pier facing a broadband acoustic source. The acoustic response at the dish for various angles of incidence was divided by a free-space reference spectrum to obtain the dish beam forming response. As for the underwater measurements, absolute gain levels are not available and the curves are plotted in dB relative to the sound intensity at the beam peak. The results of the air calibration and the scattering theory are shown in Fig. 7. All the experimental curves were shifted by 0.8 deg to compensate for a light breeze which was blowing across the propagation path. The agreement between theory and experiment is not compromised by spurious sources, and in some respects shows better agreement than the underwater case.

V. OFF-AXIS-RESPONSE

The integral representation for the acoustic field scattered by the dish developed in Sec. I is based on the assumption of symmetry about the dish axis. This is a computationally convenient restriction that allows an analytical evaluation of one of the double integrals for the scattered field and a stationary phase analysis of the remaining integral. The applicability of the analysis can, however, be extended in the following way.

Consider the off-axis receiver point $\overline{z_r}$ shown in Fig. 8. A double integral for the scattered field can be developed for this point in a spherical polar coordinate system where the z-axis is aligned with the radial passing through $\overline{z_r}$. Because of the symmetry of the shell, the expression will be the same as Eq. (8), with the exception of the limits of integration. Figure 8 shows one cross-section through the dish, illustrating how the angular measure between the z-axis and one rim increases while the other decreases.

From the asymptotic analysis of the symmetrical case, we know that for $1/2<\beta<1$ and $\theta_0 \ll 1$ the main contributions to the scattered field come from energy reflected from the base of the dish and a specular reflection. Near the caustic where the dish gain is a maximum, the energy scattered from the rim of the dish does not significantly affect the beam forming properties of the shell. It follows that the perturbed limits of integration, which account for energy scattered from the rim, will have little effect on focused field, and the focusing properties of the dish at $\overline{z_r}$ in a rotated coordinate system will be similar to those at z_r.

Thus the focusing properties of the spherical dish at an off-axis receiver point are very similar to the on-axis response. Instead of the single focus point achieved with a parabolic reflector, the spherical reflector has a focal surface, which is the surface of a sphere centered on the reflector origin. Each point on the focal surface has a maximum beam forming response to plane waves incident along an axis passing through the point of interest and the reflector origin. The existence of a focal surface, as opposed to the single focal point of a parabolic reflector, is an attractive property if an application calls for the formation of multiple focal points corresponding to different look directions without moving the reflector. The 1000-f diameter spherical reflector radio telescope at Arecibo, Puerto Rico relies on this property to focus on radio sources in different parts of the sky by moving the receiver element rather than the dish. In the ADONIS system, multiple receivers are placed on the focal surface to form the acoustical analogue of an optical image. The disadvantages of spherical reflectors include reduced gain over parabolic reflectors, inherent spherical aberration in the focal region and chromatic aberration.[1,7,12]

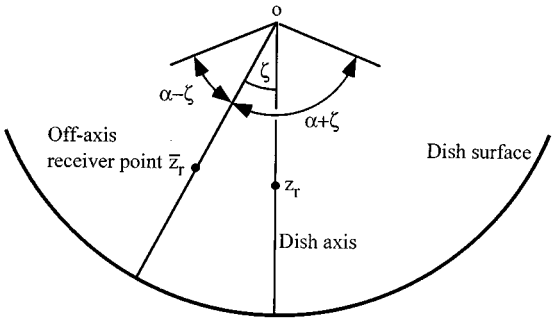


FIG. 8. The geometry of an off-axis receiver point $\overline{z_r}$. The figure illustrates how the angular measure between the dish rim and radial through $\overline{z_r}$ is a function of azimuthal angle for off-axis points. The angles illustrated are $\alpha+\zeta$ and $\alpha-\zeta$, respectively, for $\phi=0$ and $\phi=\pi$.

VI. SUMMARY

The main result of this work is the derivation of expressions for the gain and half-power beam with of a concave, spherical reflector with a pressure-release boundary. The reflector gain scales with frequency and the beam width scales inversely with frequency. The beam patterns predicted by the scattering theory compare well with air and underwater data collected from the 3-m spherical reflector used in the ADONIS experiments. Although the analytical expressions developed here for the gain and beam width apply to a receiver on the reflector axis, the scattering integral can be extended to off-axis points (see Sec. V). A numerical study of off-axis receiver points based on this approach has been presented by Epifanio, who also considers the chromatic aberration and depth-of-field associated with the ADONIS reflector.[1]

ACKNOWLEDGMENTS

I am pleased to thank Dr. Chad Epifanio and Dr. John Potter for generously making the experimental data available and to Professor Michael Buckingham for helpful discussions. This work was supported under the Office of Naval Research Contract No. N00014-96-1-0120.

APPENDIX A. ASYMPTOTIC ANALYSIS OF THE FIELD INTEGRAL

An approximate form for the field integral in Eq. (9) can be obtained by the method of stationary phase. We will begin by considering the simplified case of normally incident plane waves for which $\theta_0=0$.

1. Normally incident plane waves

For normally incident plane waves, the field integral in Eq. (9) can be expressed in the canonical form

$$p=\int_0^{\alpha}\exp(ikR_0\xi(\theta))f(\theta)d\theta, \tag{A1}$$

where

$$\xi(\theta)=s^{1/2}+\cos(\theta), \tag{A2}$$

$$f(\theta)=\frac{ikR_0}{s^{1/2}}\sin(\theta)\cos(\theta), \tag{A3}$$

and

$$s=1+\beta^2-2\beta\cos(\theta). \tag{A4}$$

The critical points of Eq. (A1) are the end points of the contour of integration, $\theta_0=0$ and $\theta=\alpha$, and the points of stationary phase at which

$$\xi'=\sin(\theta)\left(\frac{\beta}{s^{1/2}}-1\right)=0, \tag{A5}$$

where $'$ indicates taking the derivative with respect to θ. Later we will need the second derivative of the phase function with respect to θ, which is:

$$\xi''=\frac{\beta\cos(\theta)}{s^{1/2}}-\frac{\beta^2\sin^2(\theta)}{s^{3/2}}-\cos(\theta). \tag{A6}$$

The points of stationary phase are solutions to Eq. (A5). There are two of them at

$$\theta_1=0, \tag{A7}$$

and

$$\theta_2=\arccos\left(\frac{1}{2\beta}\right). \tag{A8}$$

The first stationary point lies on an end point of integration, and also corresponds to a point at which $f=0$ and $f'\neq 0$. In this case the asymptotic contribution to the integral is[13]

$$p_1=\frac{f_1'}{kR_0|\xi_1''|}\exp\left(ikR_0\xi_1+i\,\mathrm{sgn}(\xi_1'')\frac{\pi}{2}\right), \tag{A9}$$

where $\xi_1=\xi(0)$, $\xi_1''=d^2\xi/d\theta^2|_{\theta=0}$, $f_1'=df/d\theta|_{\theta=0}$ and $\mathrm{sgn}(z)$ is the Signum function which is -1 if $z<0$ and 1 if $z\geqslant 0$. Evaluating the various terms in Eq. (A9), we obtain

$$p_1=\frac{1}{1-2\beta}\exp(ikR_0(2-\beta)) \tag{A10}$$

provided $\beta\neq 1/2$. The behavior of the integral at $\beta=1/2$ is discussed later.

The stationary point lies in the interior of the path of integration if $1/2<\beta<1/(2\cos\alpha)$. Provided β lies in this range, the contribution to the integral is for an isolated, first-order stationary point, which is

$$p_2=f_2\left(\frac{2\pi}{kR_0|\xi_2''|}\right)^{1/2}\exp\left(ikR_0\xi_2+i\,\mathrm{sgn}(\xi_2'')\frac{\pi}{4}\right), \tag{A11}$$

where $\xi_2=\xi(\theta_2)$, $\xi_2''=d^2\xi/d\theta^2|_{\theta=\theta_2}$ and $f_2=f(\theta_2)$. Making the appropriate substitutions, we find that

$$p_2=i\left(\frac{\pi kR}{2\beta^3}\right)^{1/2}\exp\left(ikR_0\left(\beta+\frac{1}{2\beta}\right)-i\frac{\pi}{4}\right). \tag{A12}$$

The asymptotic expansion for p depends on β. For $\beta<1/2$ there is only one stationary point contribution from $\theta=0$, which is given by Eq. (A10). For $1/2<\beta<1/(2\cos\alpha)$ there are two stationary points and the asymptotic form is given by the sum of Eqs. (A10) and (A12). At $\beta=1/2$, the two stationary points merge, and the first-order asymptotic analysis fails. A uniform asymptotic analysis would provide an analytical form for the integral valid for all values of β. We have not undertaken such an analysis as it would significantly complicate the results without adding any further physical insight.

Comparing Eqs. (A10) and (A12), it is apparent that the stationary point contributions are of different order in the expansion variable kR_0. The contribution from $\theta=0$ is independent of kR_0 whereas that from θ_2 scales as $(kR_0)^{1/2}$. The difference in power of kR_0 arises from the fact that $f(0)=0$, which reduces the order of the stationary point contribution at $\theta=0$.

There is a third critical point at the end point of integration $\theta=\alpha$, at which $f\neq 0$. The contribution to the integral from this point is

$$p_3=\frac{\cos\alpha}{\beta-(1+\beta^2-2\beta\cos\alpha)^{1/2}}$$
$$\times\exp(ikR_0(1-\beta+\cos\alpha)). \quad \text{(A13)}$$

For $\alpha=\pi/2$ (a hemispherical dish), p_3 is zero and the end point does not contribute to the integral. If $\alpha<\pi/2$, then p_3 must be included in the asymptotic expansion of the integral. Note that at $\beta=1/(2\cos\alpha)$, the stationary point θ_2 is equal to α and p_3 is singular.

2. Inclined plane waves

To study the beam forming properties of the dish, we must consider incident plane waves for which $\theta_0\neq 0$. As the nature of the critical points is unchanged from the case for normally incident energy, the analysis proceeds largely as before. The canonical form for the integral is Eq. (A1) with

$$\xi(\theta)=s^{1/2}+\cos(\theta_0)\cos(\theta), \quad \text{(A14)}$$

$$\xi'=\frac{\beta\sin(\theta)}{s^{1/2}}-\cos(\theta_0)\sin(\theta), \quad \text{(A15)}$$

$$\xi''=\frac{\beta\cos(\theta)}{s^{1/2}}-\frac{\beta^2\sin^2(\theta)}{s^{3/2}}-\cos(\theta_0)\cos(\theta), \quad \text{(A16)}$$

and

$$f(\theta)=\frac{ikR_0}{s^{1/2}}\sin(\theta)$$
$$\times\{\cos(\theta_0)\cos(\theta)J_0(kR|\sin(\theta_0)|\sin(\theta))$$
$$-i|\sin(\theta_0)|\sin(\theta)J_1(kR|\sin(\theta_0)|\sin(\theta))\}. \quad \text{(A17)}$$

The Bessel functions in Eq. (A17) have not been included in the phase function as we are interested in the behavior of the integral for small θ_0 where the main beam forming response occurs. Consequently, we will be concerned with the Bessel function properties for small argument which do not affect the phase of the integrand.

The stationary phase points are solutions of Eq. (A15), which are

$$\overline{\theta_1}=0 \quad \text{(A18)}$$

and

$$\overline{\theta_2}=\arccos\left(\frac{1}{2\beta}-\frac{\beta}{2}\tan^2(\theta_0)\right), \quad \text{(A19)}$$

where we have used an overbar to distinguish the stationary points from the case when $\theta_0=0$. The stationary point contributions are evaluated using Eqs. (A9) and (A11). Once found, the expressions can be expanded in a Taylor series about $\theta_0=0$. The calculations are straightforward and yield the approximate forms, correct to order θ_0^2:

$$\overline{p_1}\approx\frac{1}{1-2\beta}\left(1+\frac{1}{2}\left(\frac{1-\beta}{1-2\beta}-1\right)\theta_0^2\right)$$
$$\times\exp(ikR_0(2-\beta-\theta_0^2/2)) \quad \text{(A20)}$$

and

$$\overline{p_2}\approx i\left(\frac{\pi kR_0}{2\beta^3}\right)^{1/2}$$
$$\times\{(1-(\beta^2+1/4)\theta_0^2)J_0(kR|\sin(\theta)|\sin(\theta_2))$$
$$-i(4\beta^2-1)^{1/2}|\theta_0|J_1(kR|\sin(\theta)|\sin(\theta_2))\}$$
$$\times\exp(ikR(\beta+1/(2\beta)-1/(4\beta)\theta_0^2-i\pi/4). \quad \text{(A21)}$$

For $\theta_0=0$, Eqs. (A20) and (A21) respectively reduce to Eqs. (A10) and (A12).

3. Expansion of the Bessel functions for large *kR* and small θ_0

Equation (A21) is asymptotically valid for $kR_0\gg 1$ and $\theta_0\ll 1$, to order θ_0^2. It can be simplified further by imposing the stronger restriction $kR_0|\theta_0|\ll 1$. Assuming this inequality holds, we can use the small argument expansions

$$J_0(z)\approx 1-z^2/4 \quad \text{(A22)}$$

and

$$J_1(z)\approx z/2 \quad \text{(A23)}$$

to obtain

$$\overline{p_2}\approx i\left(\frac{\pi kR_0}{2\beta^3}\right)\left\{1-\left(1+4\beta^2+k^2R_0^2\left(1-\frac{1}{4\beta}\right)\right)\right.$$
$$\times\theta_0^2/4-2ikR_0\left(\frac{4\beta^3-\beta}{4\beta-1}\right)^{1/2}\theta_0^2\Big\}$$
$$\times\exp\left(ikR\left(\beta+\frac{1}{2\beta}+\frac{1}{4\beta}\theta_0^2\right)-i\pi/4\right). \quad \text{(A24)}$$

[1] C. L. Epifanio, ''Acoustic Daylight: Passive Acoustic Imaging using Ambient Noise,'' Ph.D. Thesis, University of California, San Diego, 1997, pp. 311.

[2] J. R. Potter, ''Acoustic imaging using ambient noise: Some theory and simulation results,'' J. Acoust. Soc. Am. **95**, 21–33 (1994).

[3] S. Flatté and W. Munk, ''Submarine detection: Acoustic contrast versus acoustic glow,'' MITRE Corporation, McLean, Virginia, JASON JSR-85-108, 1985.

[4] M. J. Buckingham, B. V. Berkhout, and S. A. L. Glegg, ''Imaging the ocean with ambient noise,'' Nature (London) **356**, 327–329 (1992).

[5] S. A. L. Glegg and M. J. Buckingham, ''Acoustic Daylight: A new way to explore the ocean?,'' Florida Atlantic University Report, 1990.

[6] M. J. Buckingham, ''Theory of acoustic imaging in the ocean with ambient noise,'' J. Comput. Acoust. **1**, 117–140 (1993).

[7] R. C. Spencer and G. Hyde, ''Studies of the focal region of a spherical reflector: Geometric optics,'' IEEE Trans. Antennas Propag. **AP-16**, 317–324 (1968).

[8] C. Scott, *Modern Methods of Reflector and Antenna Analysis and Design* (Artech House, Norwood, 1990), pp. 130.

[9] L. L. Tsai, D. R. Wilton, M. G. Harrison, and E. H. Wright, ''A comparison of geometrical theory of diffraction and integral equation formulation for analysis of reflector antennas,'' IEEE Trans. Antennas Propag. **AP-20**, 705–712 (1972).

[10] W. V. T. Rusch and P. D. Potter, *Analysis of Reflector Antennas* (Academic, New York, 1970), pp. 178.

[11] A. T. de Hoop, S. Zeroug, and S. Kostek, ''Transient analysis of the transmitting properties of a focused acoustic transducer with an arbitrary rim,'' J. Acoust. Soc. Am. **98**, 1767–1777 (1995).

[12] J. R. Potter, ''Acoustic daylight: Theory and numerical simulations for an imaging system using ambient noise in the ocean,'' J. Acoust. Soc. Am. **92**, 2351 (1992).

[13] N. Bleistein and R. A. Handelsman, *Asymptotic Expansions of Integrals* (Holt, Rinehart and Winston, New York, 1975), pp. 425.

Variation in the dispersion of axisymmetric waves in infinite circular rods with crystallographic wire texture

T. A. Mason[a)]
Materials Science and Technology Division, MS G755, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

(Received 8 September 1998; revised 6 May 1999; accepted 18 May 1999)

This paper presents the solution to the frequency equation for a number of polycrystalline, textured circular rods having transverse isotropy. The effective, second-order elastic stiffness tensors were estimated using the recursive general Hill arithmetic mean (GHAM). The velocity dispersion curves for a number of combinations of materials and crystallographic fiber or wire textures were calculated and the variation due to texture displayed. At large wavelengths, the velocity dispersion of fiber textured materials exhibits a lowest-order axisymmetric mode which varies only with the directional Poisson's ratios in a manner similar to that of isotropic aggregates. In this wavelength regime, the waves propagate nondispersively at the wave speed, C_0, as dictated by the directional Young's modulus. At wavelengths smaller than the rod radius, the dispersion curves were more influenced by the full anisotropy of the wire textures. At these wavelengths, the dispersion curves for the anisotropic materials deviated significantly from those of the isotropic materials and one another with the higher axisymmetric vibration modes exhibiting extreme differences. This deviation is a function of the single crystal anisotropy and nature of the wire textures.
[S0001-4966(99)00309-4]

PACS numbers: 43.20.Jr, 43.20.Mv [AN]

INTRODUCTION

An extensive bibliography exists regarding the velocity dispersion relations for axisymmetric compression waves in infinite circular rods.[1–5] The study of vibration and dispersion in infinite, isotropic circular rods is often referred to as the Pochhammer–Chree theory in reference to the early, independent work of Pochhammer[1] and Chree.[2] Chree[6] subsequently extended the isotropic analysis to the study of the vibration of rods with transverse-isotropic or hexagonal symmetry. At the level of approximation employed to make the problem tractable, the form of the solution was identical to that for an isotropic bar but evaluated with the values of Young's moduli and Poisson's ratio which correspond to the properties of a transverse-isotropic material. More recently, the exact solution for the dispersion of axisymmetric waves in transverse-isotropic, infinite bars has been presented by a number of authors.[7–10]

The purpose of the present study is to demonstrate the solution of the Pochhammer–Chree equations for a number of polycrystalline, transverse-isotropic circular rods exhibiting elastic stiffness tensors representative of real materials with crystallographic texture. A textured polycrystalline material is one which has a non-random distribution of crystal lattices in its constituent grains. This texture or preferred lattice orientation results from the thermomechanical processing used to reduce a large cast billet to a smaller geometrical shape, such as a cylindrical rod in this case.[11] Details of the thermo-mechanical processing such as work piece geometry, stress state and temperature, are directly related to the nature of the crystallographic texture that evolves during processing. It is the combination of the texture and the symmetry of the single crystal which dictates the nature of the macroscopic properties of the material. In this study, the variation of the velocity dispersion of compression waves in infinite cylindrical rods will be studied as a function of the anisotropic elastic stiffness tensor. The key point of departure of this study from other dispersion studies on transverse-isotropic materials is that appropriate combinations of materials and crystallographic textures are used to more accurately simulate the macroscopic response of real materials.

First, the frequency equation for infinite, transverse-isotropic circular rods will be briefly presented and the material parameters required for its solution will be discussed. The methods used to obtain the effective elastic properties for textured polycrystals will then be outlined and applied to a number of appropriate combinations of materials and crystallographic textures. The velocity dispersion curves for these fictive textured bars will then be calculated and compared to the results for isotropic materials.

I. FREQUENCY EQUATION FOR TRANSVERSE-ISOTROPIC, INFINITE CIRCULAR RODS

Morse[7] was the first to present the exact solution of the frequency equation for a general transverse-isotropic, solid circular rod. The transcendental dispersion equation, which relates the phase velocity and the angular frequency, is obtained by solving the equations of motion with the application of several boundary conditions on the allowed displacements and stresses and the use of the transverse-isotropic symmetry in the elastic stiffness matrix. It is assumed that the displacement in the rod is independent of the cylindrical angle, θ, but still a function of the radius and axial location.

a)Electronic mail: tmason@lanl.gov

Additionally, it is assumed that the cylindrical surface of the rod ($r=a$) is stress free. The result of the derivation is reported in Eqs. (1)–(3). In the interest of space, the interested reader is referred to the detailed presentations of Morse,[7] Mirsky,[8] Eliot and Mott[9] or Mengi and McNiven.[10]

The general transverse-isotropic frequency equation is given in Eq. (1). Here J_0 and J_1 are the zeroth- and first-order Bessel functions of the first kind, a is the bar radius, γ is the wave number, and C_{ij} are components of the effective stiffness tensor in Voigt matrix notation. Note that a capital ''C'' without subscripts refers to the phase velocity which is related to the frequency, f, and wave length, λ, as shown in Eq. (2a). The remaining variables, α, k, p and q, are defined in Eqs. (2b), (2c) and (3):

$$ap(\alpha\gamma+k^2-p^2)\left(q^2+\frac{C_{13}}{C_{44}}k^2\right)\frac{J_0(qa)}{J_1(qa)}-aq(\alpha\gamma+k^2-q^2)\times\left(p^2+\frac{C_{13}}{C_{44}}k^2\right)\frac{J_0(pa)}{J_1(pa)}-qp(p^2-q^2)\left(\frac{C_{12}}{C_{11}}-1\right)\times\left(\frac{C_{13}}{C_{44}}+1\right)=0, \tag{1}$$

$$C=\frac{\lambda f}{2\pi}=\frac{f}{\gamma}, \tag{2a}$$

$$\alpha=\frac{\gamma}{C_{11}}(C_{13}+C_{44}), \tag{2b}$$

$$k^2=\frac{\gamma^2}{C_{11}}(\rho C^2-C_{44}), \tag{2c}$$

$$\begin{Bmatrix} q^2 \\ p^2 \end{Bmatrix}=\frac{\gamma^2}{2C_{11}C_{44}}(h\pm[h^2-4C_{11}C_{44}(\rho C^2-C_{33})\times(\rho C^2-C_{44})]^{1/2}), \tag{3}$$

where

$$h=C_{44}(\rho C^2-C_{44})+C_{11}(\rho C^2-C_{33})+(C_{13}+C_{44})^2. \tag{4}$$

The discussion of Mengi and McNiven[10] on the real and imaginary domains of the Bessel functions was particularly helpful in solving for the roots of the frequency equation. The roots were obtained using a bisection routine with an acceptable error of less than 10^{-5}.

The solution of Eq. (1) requires the knowledge of six material parameters: the five independent components from the transversely isotropic elastic stiffness, C_{ij}, and the mass density, ρ. While these parameters are well known for single crystals,[12,13] the effective elastic stiffness tensor must be measured or estimated for each polycrystal with an anisotropic microstructure or texture.

Some reports on the experimental determination of velocity dispersion in long slender rods have commented on the possible influence that texture or anisotropy could have on wave propagation and dispersion.[14–16] While anisotropy has been suggested as a possible reason for deviation from the theoretical dispersion relations, few of the researchers reported any characterization of the materials' static properties or microstructures. A number of these will be discussed in Sec. VI below.

TABLE I. Wire textures resulting from swagging, extrusion and wire drawing for three types of crystal lattice.

Lattice	Texture
fcc	[100]+[111]
bcc	[110]
hcp	$[10\bar{1}0]$

II. CRYSTALLOGRAPHIC FIBER TEXTURES

As mentioned above, the effective elastic properties of textured polycrystals have received little attention in the study of the vibration of circular rods. The purpose of this study was to report the range of properties obtainable for a number of metals processed via conventional metallurgical practices. This simulation will proceed by first noting the crystallographic textures which result from producing a circular bar through deformation processing and then examining the range of effective elastic properties obtained for several symmetry classes of materials exhibiting these crystallographic textures.

Excluding bar rolling, the dominant processing techniques to produce long cylindrical rods are extrusion, swaging and drawing. These processes generate preferred alignment of the crystal lattices in the deformed metals which correspond to the rod or symmetry axis. Table I lists the dominant textures which result from these deformation processes for the cubic and hexagonal symmetry classes. The $\langle uvw\rangle$ notation refers to a fiber texture where the $\langle uvw\rangle$ directions of the individual grains are aligned with the rod axis. A fiber texture is defined by the alignment of a given crystallographic direction in each grain with some macroscopic or sample direction. For a complete or theoretical fiber texture, the crystal directions perpendicular to this fiber axis are assumed to be uniformly distributed.

The texture resulting from production of the rods is here referred to as ''wire texture.'' For a face-centered cubic (fcc) material, an experimental wire texture would be a combination of the [100] and [111] fibers along with a collection of random orientations. The proportion of [100] and [111] fibers in a fcc wire texture is a function of the metal and level of alloying. Typical wire-textured copper may contain approximately 25%–35% [100] fiber while aluminum and lead will contain less than 10% and silver more than 90% [100] fiber.[11] The respective fiber textures were simulated by rotating appropriate reference orientations about the fiber axis in small increments. These theoretical distributions of orientations were randomized by the addition of a random walk within a 5° diameter sphere centered on each of the three Euler angles used to describe the orientation. To realistically simulate the properties of a fcc wire texture, the elastic properties of the theoretical [100] and [111] fibers must be combined in a ratio appropriate for the material. This was done by constructing sets of single orientations to reflect the appropriate wire texture. It should be noted that throughout this

study it is assumed that the symmetry axis of the transversely isotropic material is parallel to the X_3 axis of the sample.

III. EFFECTIVE PROPERTIES OF TEXTURED POLYCRYSTALS

The Voigt[17] and Reuss[18] estimates have long been applied to approximating the elastic response of polycrystalline materials.[19,20] After proving that the Voigt and Reuss estimates were extremal values of the stored elastic energy in strained polycrystals, Hill[21] suggested using the arithmetic or geometric means of the Voigt and Reuss shear modulus estimates to better approximate the effective shear modulus of the polycrystal. The arithmetic mean of the Voigt and Reuss estimates of the elastic stiffness, here called the ''Hill average,'' has been widely applied and found to correlate well with experiment.[22,23] It has subsequently been shown that other estimation schemes, such as the self-consistent approach,[24–27] yield more precise estimates of the effective material properties of polycrystals when the microstructure exhibits aligned, elongated grains. It has also been shown that the differences between the Hill average and the self-consistent estimate are often almost vanishing.[27]

For the purposes of this study, a recent suggestion by Morawiec[28] will be used to calculate the effective stiffness tensors for the textured polycrystals. Morawiec suggested taking the recursive averages of the Voigt and Reuss estimates until the calculated stiffness and compliance tensors converge to an inverse relationship. This inverse relation in the second-order elastic properties is exhibited in nature by all mono- and polycrystals. It should not be forgotten that this recursive estimation scheme is neither introduced nor supported in a formal physics-based manner and that the use of the Voigt and Reuss estimates as the basis for this analysis inherently assumes equiaxed grains and no correlation between neighboring grains. Matthies and Humbert[28] have shown that the GHAM method, or ''super-Hill average'' as they called it, compares very well with the self-consistent and geometric means. Mason[29,30] has also used the GHAM to study the influence of texture on the second- and third-order elastic stiffness of textured polycrystals. Similarly, the GHAM technique was used in this study to estimate the effective properties of polycrystals for a number of wire-textured materials of different crystal symmetries. A discussion of the averaging procedure used to calculate the Voigt and Reuss estimates and the GHAM recursion is given in the Appendix.

The Young's modulus, E, Poisson's ratio, ν, and three elastic wave speeds for a number of isotropic materials are presented in Table II. The mass density and anisotropy factor for the single crystals are also listed. The anisotropy factor is defined in Eq. (5):

$$A=\frac{2C_{44}}{C_{11}-C_{12}}. \tag{5}$$

Additionally, the ratio of the elastic stiffness in the a- and c-axis directions is reported in parentheses for the hexagonal metals. The single crystal elastic stiffness values reported in Ref. 12 were employed in the calculation of the effective stiffness tensors for textured polycrystals. Table II shows that isotropic beryllium exhibits a Poisson's ratio very near zero and lead is near 0.4 while the remainder of the metals cluster in a band around the value 0.3. This shows that the use of a value of 0.3 for Poisson's ratio for most materials, assuming isotropy, is not too extreme of an assumption. The average value of the Poisson's ratios listed in Table II is 0.298. Exclusion of beryllium and lead would result in an average of 0.313.

TABLE II. Young's moduli, Poisson's ratio and elastic wave speeds for isotropic polycrystals.

	Density (Mg/m^3)	Anisotropy factor	E (GPa)	ν	C_O (km/s)	C_L (km/s)	C_S (km/s)
Mg	1.781	0.98 (0.96)	44.62	0.290	5.01	5.73	3.12
Be	1.850	1.22 (0.83)	311.5	0.034	12.97	13.01	9.04
Al	2.699	1.23	70.15	0.347	5.10	6.42	3.11
Ti	4.510	1.33 (0.88)	113.6	0.319	5.02	6.01	3.11
Zr	6.490	0.95 (0.88)	97.18	0.330	3.87	4.71	2.37
α-Fe	7.850	2.42	205.1	0.279	5.11	5.78	3.20
Ni	8.874	2.59	215.9	0.305	4.93	5.76	3.06
Cu	8.930	3.21	125.8	0.347	3.75	4.73	2.29
Mo	10.21	0.91	326.9	0.292	5.66	6.50	3.52
Pb	11.35	4.00	23.52	0.411	1.44	2.20	0.86
Ta	16.65	1.56	184.9	0.339	3.33	4.12	2.04
W	19.22	1.00	402.6	0.281	4.58	5.18	2.86

The elastic wave speeds presented in Table II are defined in Eqs. (6a)–(6c) in terms of the Lamé constants (λ, μ) for an isotropic material and the components of the elastic stiffness or compliance matrices (C_{ij},S_{ij}) for a transverse-isotropic material. Here C_0, C_L and C_S are the longitudinal elastic wave speeds in materials of finite and infinite extent and the shear wave speed, respectively. The notation in Eqs. (6a)–(6c) assume the wave propagation direction is parallel to the X_3 sample axis:

$$\rho C_0^2=E=\frac{\mu(3\lambda+2\mu)^{\text{Iso}}}{(\lambda+\mu)}=\frac{1^{\text{Trans Iso}}}{S_{33}}, \tag{6a}$$

$$\rho C_L^2=(\lambda+2\mu)^{\text{Iso}}=C_{33}{}^{\text{Trans Iso}}, \tag{6b}$$

$$\rho C_S^2=\mu^{\text{Iso}}=C_{44}{}^{\text{Trans Iso}}. \tag{6c}$$

The variations in the estimated E, ν and elastic wave speeds according to the Voigt, Reuss, and Hill averages and GHAM method for isotropic polycrystals of copper, nickel and lead are presented in Table III. Comparison of the values for

TABLE III. Young's modulus and Poisson's ratio for isotropic aggregates of fcc crystals by four estimation methods.

	Cu		Ni		Pb	
Estimation method	E (GPa)	ν	E (GPa)	ν	E (GPa)	ν
Voigt	144.7	0.324	236.6	0.286	28.81	0.391
Hill	127.5	0.345	217.2	0.303	23.55	0.408
GHAM	125.9	0.347	215.9	0.305	23.52	0.411
Reuss	109.4	0.367	196.9	0.322	19.19	0.427

TABLE IV. Young's moduli, Poisson's ratio and elastic wave speeds for fictive [110] fiber textured bcc polycrystals. Fiber axis is assumed parallel to the three-axis.

	E_3 (GPa)	ν_{13}	C_O (km/s)	C_L (km/s)	C_S (km/s)
α-Fe	228.9	0.253	5.40	5.94	3.06
Mo	322.8	0.295	5.62	6.48	3.54
Ta	196.0	0.329	3.43	4.17	1.99
W	402.7	0.281	4.58	5.18	2.86

Young's modulus and Poisson's ratio reveals that while the use of the Voigt and Reuss estimates would result in noticeably different dispersion calculations, the difference between the Hill average and the converged GHAM would be vanishing. For all three metals reported, the Reuss estimate would underestimate the elastic moduli while over predicting Poisson's ratio. The opposite trends are seen for the Voigt estimate.

While the relative difference between the Voigt and Reuss estimates for each metal in Table III is clearly dependent on the anisotropy of the single crystal, the difference between the first Hill average and the GHAM is almost insignificant for the three metals. This rapid convergence of the recursive averages to a reciprocal relationship between the stiffness and compliance tensors was exhibited by all of the metals listed in Table II.

The GHAM method was next used to obtain the effective properties for fiber-textured bcc, fcc and hcp polycrystals. Tables IV–VI report the respective data. The inclusion of the [0001] fiber texture in Table VI was made in order to facilitate comparison to the hexagonal single crystal values. The second-order elastic response of a hexagonal single crystal and an ideal hcp [0001] fiber-textured polycrystal are identical. The data presented in Table VI for a [0001] fiber are slightly randomized to match the degree of noise present in the other simulations. Examination of the [110] fiber-textured bcc metals in Table IV shows that materials with an anisotropy factor greater than unity exhibit ''texture stiffening'' while molybdenum experiences a decrease. The magnitudes of the variation are directly proportional to the anisotropy of the single crystals. As expected, the values for isotropic tungsten remain constant.

TABLE V. Young's moduli, Poisson's ratio and elastic wave speeds for fictive [100] and [111] fiber textured fcc polycrystals. Fiber axis is assumed parallel to the three-axis.

	E_3 (GPa)	ν_{13}	C_O (km/s)	C_L (km/s)	C_S (km/s)
[100]					
Al	62.66	0.363	4.82	6.30	3.24
Ni	130.4	0.381	3.83	5.28	3.70
Cu	66.87	0.418	2.74	4.34	2.90
Pb	10.83	0.458	0.98	2.07	1.14
[111]					
Al	75.59	0.336	5.29	6.51	3.02
Ni	298.8	0.230	5.80	6.25	2.72
Cu	190.7	0.269	4.62	5.16	1.99
Pb	39.77	0.350	1.87	2.37	0.73

TABLE VI. Young's moduli, Poisson's ratio and elastic wave speeds for fictive $[10\bar{1}0]$ and [0001] fiber textured hcp polycrystals. Fiber axis is assumed parallel to the three-axis.

	E_3 (GPa)	ν_{13}	C_O (km/s)	C_L (km/s)	C_S (km/s)
$[10\bar{1}0]$					
Mg	45.60	0.289	5.06	5.77	3.05
Be	289.9	0.047	12.5	12.6	8.93
Ti	105.8	0.333	4.84	5.92	2.99
Zr	99.11	0.317	3.91	4.70	2.30
[0001]					
Mg	50.74	0.253	5.34	5.88	3.03
Be	348.4	0.018	13.7	13.7	9.38
Ti	145.9	0.264	5.69	6.33	3.21
Zr	124.8	0.303	4.39	5.04	2.27

As all the fcc materials under consideration have anisotropy factors in excess of unity, they all exhibit the same trends in the variation of their material property parameters with texture. Comparison of the Young's moduli for the two fibers reiterates Mason's earlier report[29] regarding the interaction of the longitudinal and shear stiffness. The [100] fibers exhibit Young's moduli which range from 46% to 90% of the isotropic values while the [111] fibers exhibit moduli which range from 7% to 70% stiffer than the isotropic values.

In comparison to isotropic aggregates, those hexagonal materials with an anisotropy factor greater than unity experienced a decrease in their axial Young's moduli for the $[10\bar{1}0]$ wire texture while the other hexagonal materials exhibited an increase. The axial Young's moduli for the [0001] fibered materials were all significantly larger than the isotropic values. The magnitude of the axial moduli for a $[10\bar{1}0]$ fibered material is dominated by the elastic stiffness in the *a* axis of the single crystal while the transverse stiffness is a function of the variation in the elastic stiffness from the *c* axis to the *a* axis.

As mentioned above, the ideal wire textures for bcc and hcp metals are simply the [110] and $[10\bar{1}0]$ fiber textures. The effective properties of three fcc materials, copper, nickel and lead, were calculated using appropriate weights applied to each of the two fiber textures. For ease of reference, the

TABLE VII. Young's moduli, Poisson's ratio and elastic wave speeds for fictive wire-textured polycrystals. Fiber axis is assumed parallel to the three-axis. The balance of the wire textures for the fcc materials are [111] fiber textures.

	Wire texture	E_3 (GPa)	ν_3	C_O (km/s)	C_L (km/s)	C_S (km/s)
Cu	30%[100]	156.4	0.310	4.18	4.93	2.30
Ni	30%[100]	252.3	0.272	5.33	5.97	3.05
Pb	10%[100]	13.94	0.447	1.11	2.11	1.11
α-Fe	[110]	228.9	0.253	5.40	5.94	3.06
Ta		196.0	0.329	3.43	4.17	1.99
Mg	$[10\bar{1}0]$	45.60	0.289	5.06	5.77	3.05
Be		289.9	0.047	12.5	12.6	8.93
Ti		105.8	0.333	4.84	5.92	2.99

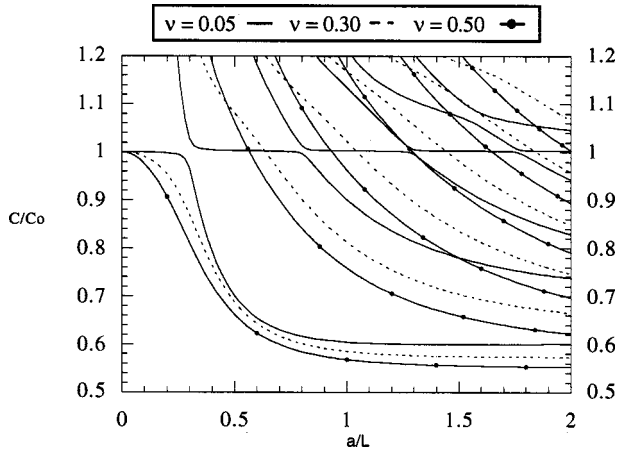


FIG. 1. Velocity dispersion curves for the first six axisymmetric vibration modes in isotropic materials with varying Poisson's ratios.

results are reported in Table VII along with a number of bcc and hcp wire-textured materials presented above in Tables IV and VI.

It should be noted that the concept of an effective medium holds in this case when the grain size is smaller than the wavelength of the propagating disturbance. Otherwise, each grain can be viewed as a local wave guide and the wave would experience extreme dispersion as it passes from one grain to another. Follansbee and Frantz[31] considered the radial variation of the longitudinal displacement as a function of frequency and examined the amplitudes of the Fourier components which fell on either side of Davies' criteria for acceptable radial variation of less than 5%. The cutoff value for acceptable variation was an a/λ ratio of less than 0.1. Their analysis shows that when the incident wave components (i.e., $n=22$) exceed this critical value, the wave amplitudes have fallen to about 2% of the lowest-order component amplitude. This magnitude of wave amplitude is in the experimental noise regime. This shows that the potential for complications due to localized wave guides is small unless the average grain size is on the order of one-tenth of the bar radius.

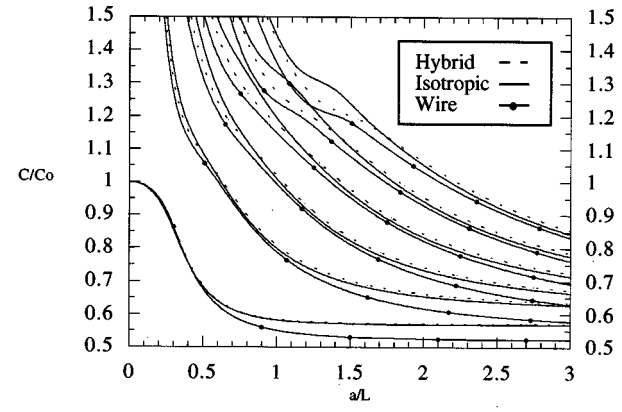


FIG. 2. Velocity dispersion curves for the first six axisymmetric vibration modes in copper polycrystals calculated using the isotropic, hybrid-isotropic, and transverse-isotropic methods.

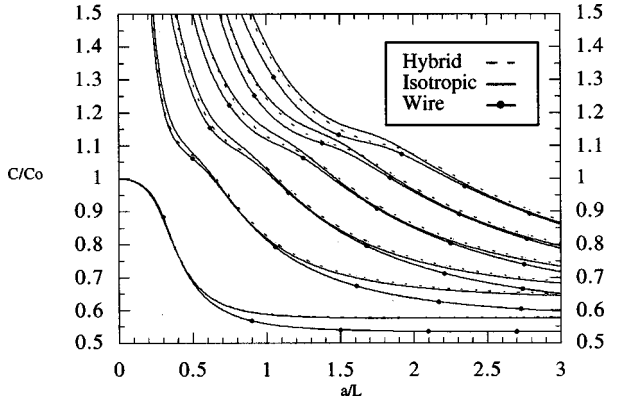


FIG. 3. Velocity dispersion curves for the first six axisymmetric vibration modes in alpha-iron polycrystals calculated using the isotropic, hybrid-isotropic and transverse-isotropic methods.

IV. DISPERSION CORRECTION CURVES FOR TEXTURED RODS

Bancroft[4] and Flinn,[5] among others, have shown the variation in the velocity dispersion curves for isotropic materials as a function of Poisson's ratio. The dispersion curves for several of the lowest-order axisymmetric modes for various Poisson's ratios are shown in Fig. 1. The relative velocity (C/C_0) is plotted versus the radius normalized by the wavelength (a/λ). The vibration modes correspond to those denoted by $LT(m,0)$ by Eliot and Mott[9] or ''axial-radial'' and ''axially symmetric with $p=0$'' by Thurston.[6]

The calculated velocity dispersion curves for wire-textured copper, alpha-iron, magnesium and titanium are presented in Figs. 2–5, respectively. In each case, the dispersion curves for the first six axisymmetric vibration modes are displayed. The velocity dispersion curves for copper and alpha-iron were calculated using three different approaches: the isotropic, transverse-isotropic and hybrid isotropic calculations. The isotropic and transverse-isotropic calculations proceeded using the formulations presented above with stiffness tensors of the appropriate symmetry. The hybrid method, originally suggested by Chree,[6] uses the isotropic formulation of the problem but applies the transverse-isotropic values for the axial Young's modulus and Poisson's ratio in its

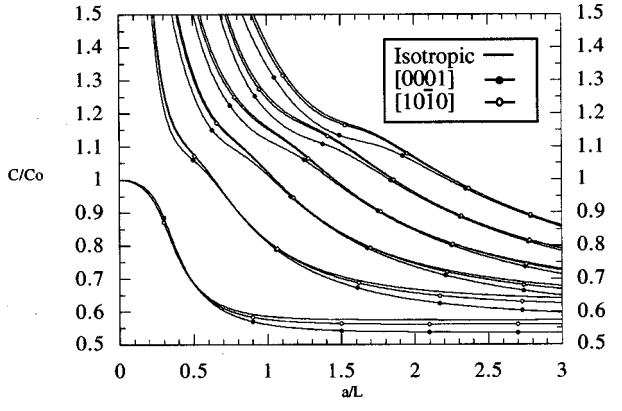


FIG. 4. Velocity dispersion curves for the first six axisymmetric vibration modes in magnesium polycrystals. Comparison of isotropic and [0001] and $[10\bar{1}0]$ wire textures.

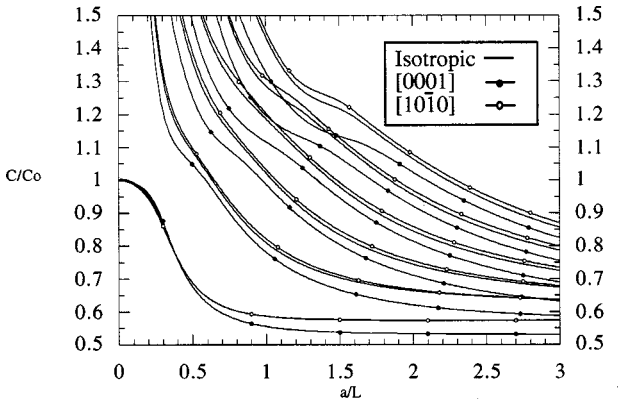


FIG. 5. Velocity dispersion curves for the first six axisymmetric vibration modes in titanium polycrystals. Comparison of isotropic and [0001] and $[10\bar{1}0]$ wire textures.

evaluation. Examination of Figs. 2 and 3 reveals that the agreement between the lowest-order, axisymmetric vibration modes for all three methods is very good in the wavelength regime that was of practical interest in Chree's time [i.e., long wavelength or $(a/\lambda)<0.5$]. At smaller wavelengths, the lowest-order hybrid isotropic and isotropic curves are nearly identical. The dispersion curves for the higher-order modes show a similar behavior, the hybrid isotropic and isotropic formulations are most similar.

The velocity dispersion curves for isotropic, [0001] and $[10\bar{1}0]$ wire fibers for magnesium and titanium are given in Figs. 4 and 5, respectively. Surprisingly, the lowest-order dispersion curve for $[10\bar{1}0]$ fiber-textured titanium ($A=1.33$) more closely matches its isotropic counterpart over the range examined than does magnesium's ($A=0.98$). The three curves for the higher-order vibration modes in magnesium lie closely together over most of the wavelength range studied while the [0001] curves for titanium appear to be offset from the isotropic and $[10\bar{1}0]$ curves by a near-constant value over the same range. As in the cases of the wire textured copper and iron, there appears to be little difference between the velocity dispersion for isotropic and textured bars at wavelengths greater than the bar diameter. At smaller wavelengths (i.e., $a/\lambda>0.5$), the influence of the anisotropy is exhibited as increasing divergence of the isotropic and transverse-isotropic dispersion curves. The lowest-order curves for both hexagonal materials cross at values near a/λ equal one-half.

V. DISCUSSION AND CONCLUSIONS

This paper presents the first report of the use of the effective properties of textured polycrystals to calculate the velocity dispersion corrections for infinite cylindrical bars. While the wire textures employed in the study are strictly theoretical idealizations, the application of the constructed polycrystals' effective properties clearly demonstrates the effect that crystallographic texture can have on the axisymmetric vibration modes of cylindrical bars.

Where the large wavelengths dominate, the difference between the lowest-order isotropic and transverse-isotropic velocity dispersions were small for all of the materials studied. These high-frequency waves propagate in a nondispersive manner at wave speeds, C_0, which are functions of the directional Young's modulus of the textured polycrystals. The difference in this nondispersive wave speed for isotropic and textured polycrystals can be seen by referring to Tables II and VII. For example, isotropic copper has an elastic wave speed of 3.75 km/s while the wire-textured polycrystal exhibits a wave speed of 4.18 km/s. This trend continues for all of the cubic metals except for lead and molybdenum. The slight decrease in the wave speed in molybdenum is explained by the anisotropy factor of the single crystal being less than unity ($A=0.91$). In the case of lead, the single crystal anisotropy is high ($A=4.0$), but the wire texture depresses the longitudinal wave speed approximately 23% below that of its isotropic speed. The anisotropies of the lead single crystal and crystallographic wire texture combine to not only depress the value of the longitudinal wave speed, C_0, but to increase the shear wave speed, C_s, to the same value, 1.1 km/s. As mentioned above, the $[10\bar{1}0]$ textures in the hexagonal materials depress the elastic wave speed proportional to the degree of anisotropy of the single crystal stiffness tensor. While the absolute value of the bulk wave speed, C_0, varies with material and texture, the velocity dispersion curves for $a/\lambda<0.4$ are very similar for all of the materials.

At smaller wavelengths, the differences between the dispersion curves for the isotropic and textured materials became larger. This can be rationalized by considering the interaction of higher-order vibration modes as well as splitting of the initially pure longitudinal waves into waves of mixed longitudinal and radial natures. While the differences were small for magnesium and titanium, the deviations between the curves for the cubic polycrystals were more pronounced. The exhibition of more uniformity between the results for the isotropic and anisotropic hexagonal materials can be justified by their relatively small single crystal elastic anisotropies. The cubic materials, on the other hand, exhibit relatively large anisotropies which are evidenced in the shift of the small wavelength portion of the first, but also in the higher-mode vibrations as well. Any physical problem where the higher-order vibration modes play a major role would require the use of the formal transverse-isotropic treatment due to this significant deviation of the velocity dispersion curves from those of an isotropic material.

Booker and Sagar[14] have reported similar deviations in the experimental velocity dispersion of the lowest-order axisymmetric mode for two alpha-brass rods. One rod was reported to exhibit crystallographic texture and a microstructure with grains elongated in the extrusion direction. A second rod exhibited macroscopic isotropy and uniform grain shape. While no specific information was presented regarding the texture of the first rod, it was reported that after the anisotropic rod was annealed, its velocity dispersion curve corresponded to that of the isotropic brass rod. The authors took this as proof that the disparity between the dispersion curves was caused by the crystallographic texture of the first sample.

Hudson[15] also reports an attempt to match the experimental velocity dispersion curves for a number of metals[16] to

theoretical predictions by varying the value of Poisson's ratio in the calculation of isotropic dispersion curves. He concluded that the experimental dispersion curves must correspond to materials with anisotropic properties due to the failure to consistently match the experimental curves to any theoretical, isotropic calculations. In the case of cold-drawn silver wires, the theoretical dispersion curve for the first flexural vibration mode with $\nu=0.49$ matched experimental data very well while the first ''elongational'' mode seemed to fit the experiment best using $\nu=0.39$. It is quite possible that the coupling of the two different vibration modes with the texture of the material would yield such a result.

As was discussed in the examples above, the differences between the isotropic and transverse-isotropic cases were greater for the higher-order axisymmetric vibration modes. It is clear that the shape of the dispersion curves for the higher-order modes is dependent not only on Poisson's ratio but the full anisotropic, elastic stiffness tensor. The change in the value of the shear wave speed is also worthy of note. This is the speed that surface waves approach at very small wavelengths. Transitioning from an isotropic to a transverse-isotropic material depresses the bulk shear wave speed by 2% on average. This average excludes lead, which exhibits a 29% increase in the shear speed. Wire-textured copper and nickel, on the other hand, exhibit shear speed changes of less then half of a percent.

The calculation of the effective properties for the fictive textured rods can be used to provide baseline values for further research on vibration and wave propagation in textured materials. Additionally, materials which exhibit a texture gradient or localized texture differences can be treated as composite or functionally graded materials. The application of these growing fields of analysis to the solution of vibration problems for more complex materials greatly broadens the class of solvable problems. A number of vibration studies have been made of two- and multi-layer composites in solid and hollow rods[6,32,33] where the bulk elastic properties of the layers are selected by the choice of metals employed. This analysis can be extended to include the variations induced by inhomogeneous deformation and recrystallization within a given material to produce single or polyphase ''composite materials'' which are optimized to yield a desired wave speed, vibration mode or energy dispersion characteristic.

ACKNOWLEDGMENTS

The author would like to thank G. C. Kaschner, J. F. Bingert, and R. D. Dixon (MST-LANL) and M. J. Beran (Tel Aviv University) for their thoughtful reviews of the manuscript. This work was performed under the auspices of the U.S. Department of Energy.

APPENDIX: VOIGT AND REUSS ESTIMATES AND THE GHAM RECURSION

1. The Voigt and Reuss estimates of the tensoral properties of textured polycrystals

The Voigt[17] and Reuss[18] assumptions are often used to calculate estimates of material properties based on the volume fraction of phases or constituents present in the material. A simple example of this analysis is presented in the introductory materials text by Ashby and Jones[34] where they introduce the topic to estimate the elastic stiffness of a two-phase composite material. The pair of basic underlying assumptions employed in the analysis physically contradict one another. The Voigt assumption is that each constituent or part of the material experiences the same strain while the Reuss assumption claims that all parts of the material see the same stress. The physical reality is that the local stress and strain fields vary through the interior of the body and that each of the above assumptions is too extreme. An expression for the strain energy of an elastically deformed polycrystal is given in Eq. (A1). Here the strain energy is defined as the volume integral of the local strain energy at each point of the body. The angle brackets ($\langle\rangle$) denote an ensemble average which is assumed to be equivalent to the respective volume average:

$$\begin{aligned} U &= \frac{1}{2}\int \epsilon_{ij}(\mathbf{x})\sigma_{ij}(\mathbf{x})d\mathbf{x} \\ &= \frac{1}{2}\int \epsilon_{ij}(\mathbf{x})C_{ijkl}(\mathbf{x})\epsilon_{kl}(\mathbf{x})d\mathbf{x} \\ &= \langle \epsilon_{ij}(\mathbf{x})C_{ijkl}(\mathbf{x})\epsilon_{kl}(\mathbf{x})\rangle \\ &= \frac{1}{2}\int \sigma_{ij}(\mathbf{x})S_{ijkl}(\mathbf{x})\sigma_{kl}(\mathbf{x})d\mathbf{x} \\ &= \langle \sigma_{ij}(\mathbf{x})S_{ijkl}(\mathbf{x})\sigma_{kl}(\mathbf{x})\rangle. \end{aligned} \tag{A1}$$

The true natures of the locally varying stress and strain fields in Eq. (A1) are not known and the information required to describe them in any detail within a representative control volume would likely be untractable. One of the simplest trial fields that can be used to study influence that the character of the strain and stress fields have on the strain energy is a constant field with a spatially varying component. The substitution of this trial field into the strain energy equation is shown in Eq. (A2):

$$U=\frac{1}{2}\int (\langle \epsilon_{ij}\rangle+\alpha\epsilon'_{ij}(\mathbf{x}))C_{ijkl}(\mathbf{x})(\langle \epsilon_{kl}\rangle+\alpha\epsilon'_{kl}(\mathbf{x}))d\mathbf{x}. \tag{A2}$$

The principles of minimum potential and complementary energy state that the calculation of the strain energy using a trial field which deviates from the actual, physical strain and stress fields would result in a value for the strain energy which exceeds the true value. This allows bounds on the strain energy to be constructed. When the parameter α is set to zero, the local strain is taken to be equal to the average strain for the entire body, $\langle \epsilon_{ij}\rangle$. The resulting expression, Eq. (A3a), is now an overestimate of the strain energy. A similar estimate exists using the average stress field and the elastic compliance tensor, Eq. (A3b):

$$U\leqslant \tfrac{1}{2}\langle\langle \epsilon_{ij}\rangle C_{ijkl}\langle \epsilon_{kl}\rangle\rangle=\tfrac{1}{2}\langle \epsilon_{ij}\rangle\langle C_{ijkl}\rangle\langle \epsilon_{kl}\rangle, \tag{A3a}$$

$$U\leqslant \tfrac{1}{2}\langle\langle \sigma_{ij}\rangle S_{ijkl}\langle \sigma_{kl}\rangle\rangle=\tfrac{1}{2}\langle \sigma_{ij}\rangle\langle S_{ijkl}\rangle\langle \sigma_{kl}\rangle. \tag{A3b}$$

Other choices of the parameter α lead to the derivation of perturbation estimates for the strain energy which is beyond the scope of the present study. Hill[21] showed that the over-

estimates obtained through the application of the Voigt and Reuss assumptions are actually rigorous bounds on the strain energy stored in the body, the Voigt estimate being the upper bound and the Reuss estimate the lower. While these bounds are often presented schematically in the form given by Eq. (A4), it should be noted that the bounds are scalar in nature and require the specification of a strain field for evaluation. The quantities of interest are usually the components of the effective bulk elastic stiffness tensor, C^*,

$$\langle S_{ijkl}\rangle^{-1} \leqslant C^*_{ijkl} \leqslant \langle C_{ijkl}\rangle. \tag{A4}$$

The average elastic stiffness and compliance tensors can be obtained through an averaging scheme using lattice orientation information as weights in the summing or integration. The local elastic stiffness or compliance is a fourth-rank tensor which is related to a reference tensor through the standard tensor transformation shown in Eq. (A5). The matrices, g_{ij}, describe the inclination of the reference frame with respect to the local crystal lattice. This formulation is passive and is the most popular description in the crystallographic texture community:

$$T_{ijkl}(g) = g_{ai} g_{bj} g_{ck} g_{dl} T^0_{abcd}. \tag{A5}$$

The average of the local tensors can be obtained in a number of ways. The most popular method has been to calculate the integral of the local elasticity tensor using the orientation distribution function (ODF) as a weighting function. The ODF, denoted by $f(g)$, is a probability density function which describes the probability of any point within a polycrystal containing any given lattice orientation. This integral is given in Eq. (A6). The ODF is usually obtained from the inversion of crystallographic pole figures measured in x-ray or neutron diffraction experiments. See the work of Bunge *et al.*,[21] Morris[24] or Matthies and Humbert[27] for details on the application of this approach:

$$\langle T_{ijkl}\rangle = \int T_{ijkl}(g) f(g) dg, \quad \int f(g) dg = 1. \tag{A6}$$

The average can also be calculated by simply taking the weighted sum of the individual stiffness tensors with appropriate weights applied in the sum as shown in Eq. (A7). This approach can be used to directly apply fictive or experimental single orientation measurements to the averaging process,

$$\langle T_{ijkl}\rangle = \frac{1}{W} \sum_{n=1}^{N} T_{ijkl}(g_n) w(g_n), \quad W = \sum_{n=1}^{N} w(g_n). \tag{A7}$$

2. The generalized Hill arithmetic mean

One of the weaknesses of the Voigt and Reuss bounds is that they are the extreme bounds of the envelope which encloses all possible strain energy states for a polycrystal with a given orientation population. No further information can be distilled from the bounds. Hill[21] noted that the arithmetic or geometric average of the parameters which describe the bounds, the bulk and shear moduli in Hill's isotropic case, might yield a more useful estimate of the macroscopic elastic response.

The arithmetic average of the bounds has been shown to compare favorably with experimental measurements and has come to be called the ''Hill average.'' This average of the bounds still suffers from its lack of reciprocity with the corresponding average defined in terms of the compliance tensors. Let C_0^+ and S_0^+ denote the average stiffness and compliance tensors, $\langle C\rangle$ and $\langle S\rangle$, respectively. Voigt and Reuss bounds would then be C_0^+ and C_0^- as defined in Eq. (A8). The Hill average is shown in Eq. (A9). A similar Hill average defined using the compliance tensors is also presented in Eq. (A9). These two averages are not related in a reciprocal manner as the case for the true effective stiffness and compliance tensors.

$$\begin{aligned} &C_0^+ = \langle C\rangle, \quad S_0^+ = \langle S\rangle, \\ &C_0^- = (S_0^-)^{-1} = \langle S\rangle^{-1}, \quad S_0^- = (C_0^-)^{-1} = \langle C\rangle^{-1}, \end{aligned} \tag{A8}$$

$$C_1^+ = \tfrac{1}{2}[C_0^+ + C_0^-], \quad S_1^+ = \tfrac{1}{2}[S_0^+ + S_0^-]. \tag{A9}$$

Morawiec[28] and others have suggested that continuing the averaging process would yield a pair of elasticity tensors which are inversely related. This recursive averaging is not grounded in any physical or mathematical law other than the fact that the effective stiffness and compliance should be reciprocals of one another. The generalization of the recursion is given in Eq. (A10):

$$\begin{aligned} &C_i^+ = \tfrac{1}{2}[C_{i-1}^+ + C_{i-1}^-], \quad C_1^- = (S_1^+)^{-1}, \\ &S_i^+ = \tfrac{1}{2}[S_{i-1}^+ + S_{i-1}^-]. \end{aligned} \tag{A10}$$

Matthies and Humbert[27] have shown that this recursion yields estimates of the elastic properties of polycrystals which are almost exactly the same as those obtained using more complex estimation schemes.

[1] L. Pochhammer, ''Über Fortpflanzungsgeschwindigkeiten kleiner Schwingungen in einem unbegrenzten isotropen Kreiszylinder,'' J. Reine Angew. Math. **81**, 324–336 (1876).
[2] C. Chree, ''Longitudinal vibrations of a circular bar,'' Q. J. Pure Appl. Math. **21**, 287–298 (1886).
[3] D. Bancroft, ''The velocity of longitudinal waves in cylindrical bars,'' Phys. Rev. **59**, 588–593 (1941).
[4] E. A. Flinn, ''Dispersion Curves for Longitudinal and Flexural Waves in Solid Circular Cylinders,'' J. Appl. Phys. **29**, 1261–1262 (1958).
[5] R. N. Thurston, ''Elastic waves in rods and clad rods,'' J. Acoust. Soc. Am. **64**, 1–37 (1978).
[6] C. Chree, ''On the longitudinal vibrations of aelotropic bars with one axis of material symmetry,'' Q. J. Pure Appl. Math. **24**, 340–359 (1890).
[7] R. W. Morse, ''Compressional waves along an anisotropic circular cylinder having hexagonal symmetry,'' J. Acoust. Soc. Am. **26**, 1018–1021 (1954).
[8] I. Mirsky, ''Wave propagation in transversely isotropic circular cylinders, Part I: Theory,'' J. Acoust. Soc. Am. **37**, 1016–1021 (1965).
[9] F. C. Eliot and G. Mott, ''Elastic waves propagation in circular cylinders having hexagonal crystal symmetry,'' J. Acoust. Soc. Am. **44**, 423–430 (1968).
[10] H. D. McNiven and Y. Mengi, ''Dispersion of waves in transversely isotropic rods,'' J. Acoust. Soc. Am. **49**, 229–236 (1971).
[11] I. L. Dillamore and W. T. Roberts, ''Preferred Orientation in Wrought and Annealed Metals,'' Metall. Rev. **10**, 271–380 (1965).
[12] Landolt-Börnstein, *Numerical Data and Functional Relationships in Science and Technology, New Series, Group III*, edited by A. G. Avery and A. K. McCurdy (Springer-Verlag, New York, 1992), Vol. 29. See also volumes 11 and 18 for earlier citations.
[13] G. Simmons and H. Wang, *Single Crystal Elastic Constants and Calculated Aggregate Properties: A Handbook* (MIT, Cambridge, MA, 1971).

[14] R. E. Booker and F. H. Sagar, ''Velocity dispersion of the lowest-order longitudinal mode in finite rods of circular cross section,'' J. Acoust. Soc. Am. **49**, 1491–1498 (1971).
[15] G. E. Hudson, Dispersion of Elastic Waves in Solid Circular Cylinders, Phys. Rev. **63**, 46–51 (1943).
[16] S. K. Shear and A. B. Focke, ''The Dispersion of Supersonic Waves in Cylindrical Rods of Polycrystalline Silver, Nickel and Magnesium,'' Phys. Rev. **57**, 532–537 (1940).
[17] W. Voigt, in *Lehrbuch der Kristallphysik* (Taubner, Leipzig, 1928), p. 962.
[18] A. Reuss, ''Berechnung der Fliedßgrenze von Mischkristallen auf Grund der Plastizitäsbedingung für Einkristalle,'' Z. Angew. Math. Mech. **9**, 103–108 (1929).
[19] L. Zuo, M. Humbert, and C. Esling, ''Elastic properties of polycrystals in the Voigt-Reuss-Hill approximation,'' J. Appl. Crystallogr. **25**, 751–755 (1992).
[20] S. Hirsekorn, ''Elastic Properties of Polycrystals, a Review,'' Textures Microstruct. **12**, 1–14 (1990).
[21] R. Hill, ''The Elastic Behaviour of a Crystalline Aggregate,'' Proc. Phys. Soc. London, Sec. A **A65**, 349–354 (1952).
[22] H.-J. Bunge and W. T. Roberts, ''Orientation distribution, elastic and plastic anisotropy in stabilized steel sheet,'' J. Appl. Crystallogr. **2**, 116–128 (1969).
[23] G. J. Davies, D. J. Goodwill, and J. S. Kallend, ''Elastic and plastic anisotropy in sheets of cubic metals,'' Metall. Trans. A **3**, 1627–1631 (1972).
[24] P. R. Morris, ''Elastic Constants of Polycrystals,'' Int. J. Eng. Sci. **8**, 49–61 (1970).
[25] M. Humbert and J. Diz, ''Some practical features for calculating the polycrystalline elastic properties from texture,'' J. Appl. Crystallogr. **24**, 978–981 (1991).
[26] B. C. Hendrix and L. G. Yu, ''Self-consistent elastic properties for transversely isotropic polycrystals,'' Acta Mater. **46**, 127–135 (1998).
[27] S. Matthies and M. Humbert, ''On the principle of a geometric mean of even-rank symmetric tensors for textured polycrystals,'' J. Appl. Crystallogr. **28**, 254–266 (1995).
[28] A. Morawiec, ''Review of deterministic methods of calculation of polycrystal elastic constants,'' Textures Microstruct. **22**, 139–167 (1994).
[29] T. A. Mason, ''Simulation of the Variation of Material Tensor Properties of Polycrystals Achieved Through Modification of the Crystallographic Texture,'' Scr. Mater. **39**, 1537–1543 (1998).
[30] T. A. Mason, ''Effective Third-Order Elastic Constants of Textured Polycrystals,'' Los Alamos Unclassified Report #98-2035 (1998).
[31] P. S. Follansbee and C. Frantz, ''Wave propagation in the split Hopkinson pressure bar,'' J. Eng. Mater. Technol. **105**, 61–66 (1983).
[32] H. D. McNiven, J. L. Sackman, and A. H. Shah, ''Dispersion of axially symmetric waves in composite elastic rods,'' J. Acoust. Soc. Am. **35**, 1602–1609 (1963).
[33] R. Parnes, ''Torsional dispersion relations if waves in an infinitely long clad cylindrical rod,'' J. Acoust. Soc. Am. **71**, 1347–1351 (1982).
[34] M. F. Ashby and D. R. H. Jones, *Engineering Materials* (Pergamon, New York, 1985), p. 58.

Scattering from fluid-loaded cylindrical shell with periodic circumferential constraints using analytical/numerical matching

R. C. Loftman and D. B. Bliss
Department of Mechanical Engineering, Duke University, Durham, North Carolina 27708-0300

(Received 6 October 1998; revised 15 April 1999; accepted 3 June 1999)

Acoustic scattering from an infinitely long, thin elastic cylindrical shell with axially periodic circumferential constraints leads to the excitation of flexural, longitudinal, and shear waves on the shell. Of interest is the interaction of these structural waves with the discontinuous constraints, and the effect of this interaction on the resultant scattered sound field. This three-dimensional, fully coupled, structural acoustic problem is solved using a method called analytical/numerical matching (ANM). The method combines high-resolution local analytical solutions and low-resolution global numerical solutions to more efficiently model structural discontinuities. The ANM local solutions have been developed to efficiently capture the rapid variation in system response across the discontinuous constraint. The local solutions are combined with a smooth global solution, modeled by modal methods, to form an accurate, uniformly valid composite solution. The ANM composite solution is more accurate and converges much more rapidly than the traditional modal approach. An important observation is that this scattering problem exhibits considerable sensitivity to modeling accuracy and convergence of structural response in the region of the discontinuities. Errors in these regions have an overall effect on the structure and the associated scattered field. © *1999 Acoustical Society of America.* [S0001-4966(99)02209-2]

PACS numbers: 43.20.Tb, 43.40.Rj [CBB]

INTRODUCTION

The vibratory motion of fluid-coupled shells with discrete structural constraints such as stiffeners and attachments is of significant practical interest. There is a need to predict the dynamic behavior of large and complicated systems of this type. Discrete structural constraints, however, are responsible for high-resolution effects that lead to poor convergence. The present work investigates a method developed to alleviate this problem using an analytic reformulation of the discrete constraint that removes high-resolution content from the problem before the overall computational evaluation is performed. The method is investigated in the context of scattering from an infinite periodically ring constrained cylindrical shell. This problem contains many of the elements that challenge the implementation of ANM in general three-dimensional problems. However, it is a special case that allows for a particularly simple classical solution and a straightforward ANM treatment. The present work is considered an important step toward the development of a practical tool for structural-acoustic modeling.

The vibration of a fluid-coupled cylindrical shell is a fundamental problem in structural acoustics which has received significant treatment by many investigators. Only a few references with direct relevance to the present work will be cited here. The infinite cylindrical shell is one of few geometries for which an analytic solution of vibratory motion exists in the presence of a coupled acoustic medium. Junger discovered the modal solution for radiation and scattering from the cylindrical shell in 1952.[1,2] The forced response, as well as scattering, from a periodically ring constrained shell was treated soon afterward.[3,4] Analytic modal solutions are possible because the separable natural modes of the shell and the fluid coincide at their interface. In the presence of constraints, analytic solution is possible so long as the nonuniformities lie with regularity and along coordinate lines.[5] The analysis of the cylindrical shell geometry was treated with significant sophistication in Junger's initial work. Junger's original modal solution for beam incident scattering from the periodically ring constrained cylindrical shell[3,4] is essentially the same solution that will be called the classical modal solution and used as a benchmark throughout this work.

Although significant and rapid progress was made initially, the subject of scattering from cylindrical shells with internal complexity is still very much an active area of research. Many permutations of problem geometry and solution method are being investigated by many authors. For instance, the physics of the fluid-loaded cylindrical shell is being investigated by authors such as Scott,[6] Photiadis,[7] and Rumerman.[8] The nature of the structurally constrained shell and the constraints themselves is the subject of work by authors such as Guo,[9] Skelton,[10] and Mead and Bardell.[11] Of particular interest with regard to the present work, Cuschieri and Feit have formulated hybrid numerical and analytical Green's functions for discrete line forces on a plate[12] and cylindrical shell.[13] The Green's function for the cylindrical shell was then used to model discrete constraints in scattering problems with single and multiple constraints.[14,15] Their method handles mathematical singularities arising in transform space due to fluid loading coincidence effects analytically, while the overall evaluation of the inverse transform is performed numerically.

The present work is most similar to the mode acceleration method[16,17] with regard to how computational advantage is gained. Both methods improve computational efficiency by capturing most of the high-wave-number content of a

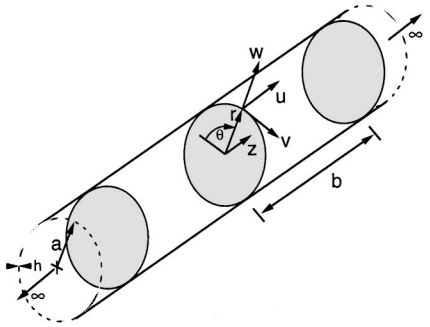


FIG. 1. Shell geometry and coordinate system.

total solution in a separate constituent solution. In the mode acceleration method, the high-resolution content is contained in the solution of the equivalent pseudo-static problem. In the present work, the high-resolution content is contained in a separate analytic solution of local extent, which is associated directly with a discontinuity. This local solution has the advantages of being derived independently of overall problem geometry and being very efficiently expressed.

Analytical/numerical matching (ANM) is a general approach designed to achieve high accuracy using low-resolution numerics by treating the most rapidly varying parts of a problem analytically, and separately, from the overall computational evaluation of the problem. The method was first applied to problems in vortex dynamics and rotorcraft free wake analysis.[18–21] More recently, ANM has been applied to rotorcraft aerodynamics,[22] wing aerodynamics for compressible unsteady flow[23] and to the development of an acoustic boundary element method.[24] The present work is part of an effort to develop ANM as a broadly applicable method for structural acoustic problems with structural discontinuities.[25–27] A simple abstraction of the method was presented in the context of a flat membrane.[25] Scattering from a two-dimensional shell with a variety of constraint configurations was then treated.[26] The present work treats scattering from a three-dimensional contained shell. Also, an alternative ANM formulation based on the Green's function formalism has been developed to treat general constraint configurations.[27] Publication of the results of this more general three-dimensional ANM formulation in comparison to the present approach to the same problem geometry is forthcoming. The ANM approach leads to an accurate alternative formulation of the basic problem, with a large reduction in required computational resolution. In addition, the alternative formulation can lead to an improved understanding of the effect of discontinuities.

I. MATHEMATICAL MODEL

The shell geometry and coordinate system are illustrated in Fig. 1. The infinitely long circular cylindrical shell with radius a is immersed in acoustic medium of non-negligible density. Discrete rigid bulkheads are affixed periodically along the entire length of the shell, a distance b apart. A bulkhead spacing of $b=2\pi a$ is chosen for this example and is used throughout. Standard z, θ, r, circular cylindrical coordinates are used. The z axis is aligned along the axis of the cylinder with zero at a bulkhead. Displacements from mean position are called u, v, and w in the axial (z), tangential (θ), and radial (r) directions, respectively. Small-disturbance linear theory is used to model the dynamics. The plane that bisects the shell along the axis and through $\theta=0,\pi$ is called the mid-plane of the problem.

The equations governing shell dynamics are the same as those in the text by Junger and Feit,[28] described there as a ''dynamic counterpart of Donnell's formulation, and are taken from Kraus.''[29] Using shell radius, a, and reciprocal of frequency, $1/\omega$, as characteristic length and time scales, the dimensionless form of the equations for shell motion are as follows.

$$-\nu\frac{\partial\bar{w}}{\partial\bar{z}}-\frac{\partial^2\bar{u}}{\partial\bar{z}^2}-\frac{1-\nu}{2}\frac{\partial^2\bar{u}}{\partial\theta^2}-\frac{1+\nu}{2}\frac{\partial^2\bar{v}}{\partial\bar{z}\partial\theta}+\bar{K}_p^2\frac{\partial^2\bar{u}}{\partial\bar{t}^2}=\bar{p}_u, \tag{1}$$

$$-\frac{\partial\bar{w}}{\partial\theta}-\frac{1+\nu}{2}\frac{\partial^2\bar{u}}{\partial\bar{z}\partial\theta}-\frac{1-\nu}{2}\frac{\partial^2\bar{v}}{\partial\bar{z}^2}-\frac{\partial^2\bar{v}}{\partial\theta^2}+\bar{K}_p^2\frac{\partial^2\bar{v}}{\partial\bar{t}^2}=\bar{p}_v, \tag{2}$$

$$\nu\frac{\partial\bar{u}}{\partial\bar{z}}+\frac{\partial\bar{v}}{\partial\theta}+\bar{w}+\bar{\beta}^2\left(\frac{\partial^2}{\partial\bar{z}^2}+\frac{\partial^2}{\partial\theta^2}\right)^2\bar{w}+\bar{K}_p^2\frac{\partial^2\bar{w}}{\partial\bar{t}^2}=\bar{p}_w-\bar{p}_f, \tag{3}$$

where $(\bar{u},\bar{v},\bar{w})=(u,v,w)/a$, $\bar{z}=z/a$, and $\bar{t}=\omega t$. Dimensionless frequency is $\bar{K}_p=k_pa=(\omega/c_p)a$ related to shell radius and the longitudinal wave speed in a plate, c_p. The dimensionless measure of thickness is $\bar{\beta}^2=h^2/12a^2$. The dimensionless pressure, defined by $\bar{p}=ap/h\rho_sc_p^2=ap(1-\nu^2)/E$, gives applied pressure related to the material stiffness and dimensions of the shell. Fluid pressure due to shell motion, denoted by p_f, is written explicitly, separate from other pressures acting on the shell. External pressures (force/area) applied to the shell and acting in the positive coordinate directions are p_u, p_v, and p_w. These external driving pressures will be used to describe the influence of structural constraints.

The rigid bulkheads are modeled as discrete constraints on the shell. The bulkheads influence the shell by applying force and moment along their discrete lines (no thickness) of intersection with the shell as necessary to yield displacements and rotation there that are zero. The motions associated with the constraints can be solved separately from the motions of the unconstrained shell, which simplifies the ANM treatment of the constraint. This is explained subsequently in Sec. II C.

In the surrounding fluid, the acoustic wave equation governs the pressure field, p_f, and the acoustic momentum equation relates the pressure solution to the velocity field, $\mathbf{V}_f$.

$$\nabla^2p_f=\frac{1}{c_f^2}\frac{\partial^2p_f}{\partial t^2}, \tag{4}$$

$$\frac{\partial}{\partial t}\rho_f \mathbf{V}_f = -\nabla p_f, \tag{5}$$

where c_f is the acoustic wave speed and ρ_f is the fluid density.

Harmonic time dependence of the form $e^{i\omega t}$ is assumed throughout this work.

II. MODAL SOLUTION

Beam incident scattering from the periodically ring constrained shell is constructed from a simple set of constituent solutions. Each constituent is a modally represented solution of the unconstrained shell forced by a symmetric and periodic pressure distribution. The solution for general periodic forcing distribution is described in the following section, Sec. II A. The constituent problems required to construct the constrained scattering problem are described in the next two sections. In Sec. II B we describe the solution of unconstrained scattering and in Sec. II C we describe the influence of the constraints.

A. Periodically forced unconstrained fluid-loaded shell

The solution of the unconstrained shell driven by an externally applied forcing, $\{\overline{p_u},\overline{p_v},\overline{p_w}\}$, that is periodic in z is the fundamental result that will be used to construct the full constrained scattering problem. Given this periodicity, the problem can be solved by assuming trigonometric series expansions in the two normalized surface coordinate directions, $\overline{z}$ and θ, for the three components of displacement, $\{\overline{u},\overline{v},\overline{w}\}$. In the present case of symmetry with respect to $z=0$ and the mid-plane, the necessary expansions for displacement reduce to the following set. Attention is restricted to this doubly symmetric case,

$$\begin{Bmatrix} \overline{u} \\ \overline{v} \\ \overline{w} \end{Bmatrix} = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \begin{Bmatrix} AB_{nm}^{\overline{u}}\cos(n\theta)\sin(K_m\overline{z}) \\ BA_{nm}^{\overline{v}}\sin(n\theta)\cos(K_m\overline{z}) \\ AA_{nm}^{\overline{w}}\cos(n\theta)\cos(K_m\overline{z}) \end{Bmatrix} e^{i\omega t}, \tag{6}$$

where $K_m = m/\overline{b}$.

At this point, the acoustic field that results from shell displacement can be determined. Determining the fluid loading on the shell surface due to the as yet unknown coefficients of displacement allows the entire coupled system to be determined in terms of shell displacement. Performing separation of variables on the wave equation in cylindrical coordinates reveals that the fluid solution is naturally expressed as an expansion of Bessel functions in r and trigonometric functions in θ and z. The trigonometric functions are chosen so that periodicity and symmetry are met in θ and z. The radial dependence is chosen so that the Sommerfeld radiation conditions for outgoing waves or decaying solutions are satisfied for large r. The following solution for the acoustic pressure results,

$$p_f = \sum_{n=0}^{\infty} AA_{nm}^{\overline{p_f}}\cos(n\theta)\cos(K_m\overline{z}) \times \begin{cases} H_n^{(2)}(K_r\overline{r})e^{i\omega t}, & K_f > K_m, \\ \mathcal{K}_n(K_r\overline{r})e^{i\omega t}, & K_f < K_m, \end{cases} \tag{7}$$

where $K_f = k_f a = (\omega/c_f)a$ is the fluid wave number normalized by shell radius a, and fluid sound speed c_f. The "radial wave number" is given by $K_r = |\sqrt{K_f^2 - K_m^2}|$. This quantity scales r in the argument of the special functions $H_n^{(2)}$ and $\mathcal{K}$. The function $H_n^{(2)} = J_n - iY_n$ is the Hankel function of the second kind of order n, and gives outwardly propagating solutions for our time convention when $K_f > K_m$. The function $\mathcal{K}_n$ is the Kelvin function, or modified Bessel function of the second kind of order n, and gives solutions that decay in r when $K_f < K_m$. Note that the functional dependence at the shell surface is identical to that of the natural modes of shell vibration.

The momentum equation allows continuity of radial velocity at the shell surface to be enforced, determining the coefficients of the acoustic solution. Acoustic loading is given by the following relationship between each mode of radial shell response and the resulting acoustic pressure mode,

$$AA_{nm}^{\overline{p_f}} = \mathcal{S}_{nm}^f AA_{nm}^{\overline{w}}, \tag{8}$$

$$\mathcal{S}_{nm}^f = \alpha\frac{K_f^2}{K_r}\begin{cases} \dfrac{H_n^{(2)}(K_r)}{H_n'^{(2)}(K_r)}, & K_f > K_m, \\ \dfrac{\mathcal{K}_n(K_r)}{\mathcal{K}_n'(K_r)}, & K_f < K_m, \end{cases} \tag{9}$$

where $\alpha = a\rho_f c_f^2/h\rho_s c_p^2$ is a dimensionless measure of the degree of fluid loading. The prime in the above denotes differentiation with respect to argument.

Substitution of the expansions for displacement shown in Eq. (6) into the governing equations of shell motion yields an independent algebraic system of linear equations for each combination of n,m. This system is solved in closed form to give the coefficients of the series exactly, provided the coefficients of the external forcing are known. The equations of motions for given n,m are expressed by the following matrix equation,

$$[D_{nm}]\{d_{nm}\} = \{p_{nm}\}, \tag{10}$$

where the matrix representing the physics of the unconstrained fluid-loaded structure, $[D_{nm}]$, for a given mode set n,m, is as follows:

$$[D_{nm}]=\begin{bmatrix} K_m^2+\frac{1-\nu}{2}n^2-K_p^2 & \frac{1+\nu}{2}nK_m & \nu K_m \\ \frac{1+\nu}{2}nK_m & \frac{1-\nu}{2}K_m^2+n^2-K_p^2 & n \\ \nu K_m & n & \beta^2(n^2+K_m^2)^2+1-K_p^2+\mathcal{S}_{nm}^f \end{bmatrix}. \tag{11}$$

The vector representing the coefficients of displacement, $\{d_{nm}\}$, for this set of modes is defined by the following:

$$\{d_{nm}\}=\begin{Bmatrix} AB_{nm}^{\overline{u}} \\ BA_{nm}^{\overline{v}} \\ AA_{nm}^{\overline{w}} \end{Bmatrix}. \tag{12}$$

The forcing coefficients are given by a double Fourier decomposition of the forcing distribution, $\{\overline{p_u},\overline{p_v},\overline{p_w}\}$, in θ and $\overline{z}$. The vector representing the forcing coefficients for this set of modes, $\{p_{nm}\}$, is given by the following:

$$\{p_{nm}\}=\begin{Bmatrix} AB_{nm}^{\overline{p_u}} \\ BA_{nm}^{\overline{p_v}} \\ AA_{nm}^{\overline{p_w}} \end{Bmatrix} = \frac{\epsilon_n\epsilon_m}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi\overline{b}}^{\pi\overline{b}}\begin{Bmatrix} \overline{p_u}\cos(n\theta)\sin(K_m\overline{z}) \\ \overline{p_v}\sin(n\theta)\cos(K_m\overline{z}) \\ \overline{p_w}\cos(n\theta)\cos(K_m\overline{z}) \end{Bmatrix} d\theta d\overline{z}, \tag{13}$$

where ϵ_n is the Neumann factor, defined to be 1 when $n=0$ and 2 otherwise.

B. Scattering formulation

Acoustic scattering from the unconstrained elastic shell is formulated by a two-step process. First, scattering from a geometrically equivalent rigid cylinder is calculated. Second, the solution of the fluid-loaded elastic cylinder driven by the surface pressure found in the rigid scattering problem is added. Separating the scattering problem in this way allows it to be modeled using the forced shell solution already presented (Sec. II A).

With the help of a mathematical identity,[30] the incoming plane wave train can be written as follows. The incident wave has magnitude P_i, originates in the $\theta=0$ direction, and is beam incident (parallel to the cylinder):

$$\overline{p}_{\text{incident}}=\overline{P}e^{i(\omega t+K_f\overline{r}\cos(\theta))} = P\sum_{n=0}^{\infty}\epsilon_n i^n J_n(K_f\overline{r})\cos(n\theta)e^{i\omega t}. \tag{14}$$

Balancing radial velocity at the surface of the cylinder with outgoing waves yields the solution for scattering from the rigid cylinder:

$$p=P\sum_{n=0}^{\infty}\epsilon_n i^n\left\{J_n(K_f\overline{r}) - \frac{J_n{}'(K_f)}{H_n'^{(2)}(K_f)}H_n^{(2)}(K_f\overline{r})\right\}\cos(n\theta)e^{i\omega t}. \tag{15}$$

C. Modeling periodic rigid bulkhead constraints

The constraint requirement of the rigid bulkheads is satisfied by adding constituent solutions of ring forcing to the unconstrained result. In general, a discrete bulkhead can apply radial, tangential, and axial forces, as well as moment along its line of intersection with the shell to ensure no displacements or rotation along that line. In the present case, symmetry implies that only radial force, $F^w(\theta)$, that is even, and tangential force, $F^v(\theta)$, that is odd are required to satisfy the constraint. It is necessary to represent the forcing influence of the constraint as an applied pressure field that is consistent with the driven problem formulation of Sec. II A. The azimuthal dependence of the ring of force imparted by the constraint can be represented by Fourier decomposition in θ. The axial dependence of the constraint forcing, which is represented by periodic Dirac delta functions in z, $\delta(z-qb)$ where q is any integer and b is the constraint spacing, can also be represented by Fourier decomposition in z. This leads to the following description of constraint influence:

$$\overline{p_w^{\text{ring}}}=F^w(\theta)\,\delta(z-qb) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\frac{\epsilon_m}{2\pi}\overline{A_n^{Fw}}\cos(n\theta)\cos(K_m\overline{z}), \tag{16}$$

$$\overline{p_v^{\text{ring}}}=F^v(\theta)\,\delta(z-qb) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\frac{\epsilon_m}{2\pi}\overline{B_n^{Fv}}\sin(n\theta)\cos(K_m\overline{z}). \tag{17}$$

The axial dependence, and therefore the relative magnitude of the coefficients as m varies for a given n, is fixed by the discrete periodic nature of the constraints. The azimuthal dependence (the coefficients $\{\overline{A_n^{Fw}},\overline{B_n^{Fv}}\}$) is determined by satisfaction of the constraint definition. The present constraint configuration, does not lead to coupling between separate azimuthal modes. The coefficients $\{\overline{A_n^{Fw}},\overline{B_n^{Fv}}\}$ are independent for each n.

To solve for the ring forcing coefficients, the system response is calculated for radial and tangential ring forcing of unit strength for each azimuthal order, n. The displacements at the ring due to general ring forcing, called $\{d_n^{\text{ring}}\}$, can then be expressed by the following equation,

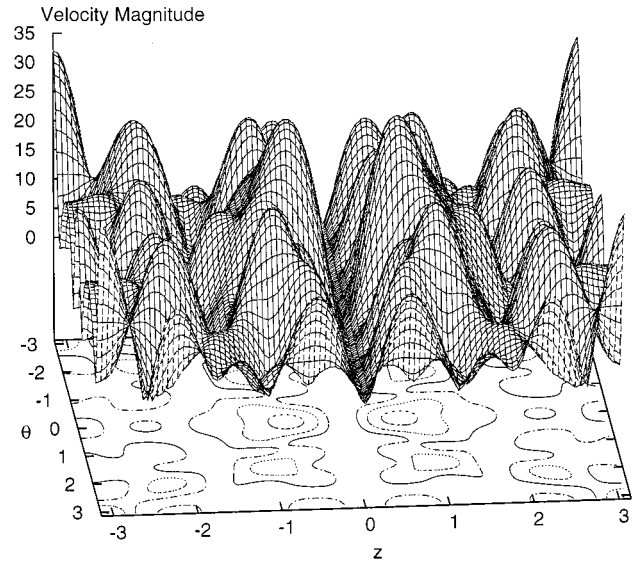


FIG. 2. Radial surface velocity: scattering from ring constrained shell.

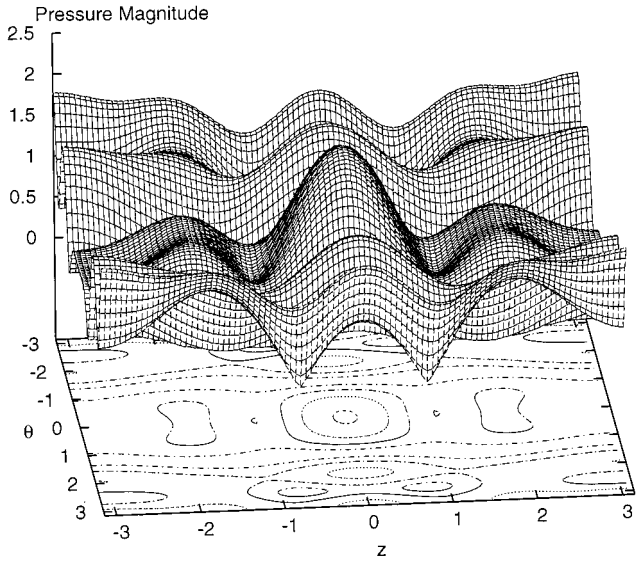


FIG. 3. Scattered pressure at $r=100a$: scattering from ring constrained shell.

$$\{d_n^{\text{ring}}\}=[dU_n^{\text{ring}}]\{F_n^{\text{ring}}\}, \tag{18}$$

where $\{F_n^{\text{ring}}\}=\{\overline{A}_n^{Fw},\overline{B}_n^{Fv}\}$ is the vector of forcing coefficients, and the matrix $[dU_n^{\text{ring}}]$ contains the radial and tangential displacements at the ring resulting from each unit ring forcing. Now, the ring forcing strength required to satisfy the constraint condition of no displacement at the ring can be determined for each azimuthal mode by the following relationship,

$$\{d_n^{\text{total}}\}=\{d_n^{\text{ring}}\}+\{d_n^{\text{unconstrained}}\}=\{0\}, \tag{19}$$

$$\{F_n^{\text{ring}}\}=-[dU_n^{\text{ring}}]^{-1}\{d_n^{\text{unconstrained}}\}, \tag{20}$$

where $\{d_n^{\text{unconstrained}}\}$ is the unconstrained displacement at the ring constraint location.

D. Sample case

In this section, a sample result for beam incident acoustic scattering from the cylindrical shell with periodic rigid bulkheads is shown. Figure 2 shows the magnitude of radial surface velocity in one periodic bay. Figure 3 shows the magnitude of far field scattered pressure. Parameters in the range of typical interest in the mid-frequency range with subsonic flexural waves are used. Acoustic wave number normalized by radius K_f is 2.6. Shell thickness, given by $\beta^2=0.000\,05$, is about 2% of radius. The shell material is characterized by longitudinal wave speed c_p that is five times the acoustic wave speed. This yields $K_p=0.52$. The mass ratio $\rho_f a/\rho_s h$ is 1, which yields the fluid loading parameter $\alpha=0.04$. This set of parameters is used as a representative sample throughout this work.

Note, in particular, that scattered pressure is characterized by significantly larger wavelengths than surface velocity. Only the lowest harmonics of radial displacement are supersonic, and therefore efficient acoustic radiators. This is important with regard to the resolution required to accurately compute scattered pressure. It may be assumed, at first, that because only lower harmonics radiate well, it is not necessary to resolve the fine scale features of shell response to calculate scattered pressure. However, a constraint on shell response at a discrete location causes even the lower harmonics to be sensitive to the accuracy with which fine scale local features are modeled.

III. ANM FORMULATION

The ANM formulation deals specifically with the discretely forced portion of the problem. The method uses superposition to separate local high-resolution content associated with a discrete forcing from the evaluation of overall system dynamics. To convey the principal of ANM before discussing the problem at hand the following mathematical abstraction of the method has been included. Consider a generic structure with displacement solution, w, governed by a linear operator, L, and driven by a discrete forcing, F_d.

$$L(w)=F_d. \tag{21}$$

Allow that this problem must be solved computationally, due to complexity of domain, for instance, and that a smoother driving force would result in more rapid convergence.

The method proceeds by construction of an analytical local solution, w_l, that incorporates the discrete forcing. This local solution is brought smoothly to rest at the edge of a prescribed smoothing region by an opposing smooth applied force, F_{gf}:

$$L(w_l)=F_d-F_{gf.} \tag{22}$$

The local solution is the most important part of the method. It is not obtained by the solution of a differential equation. Rather, it is a prescribed solution field. The nonunique local solution is carefully constructed so that it is analytically expressed, only locally nonzero, contains the original discrete forcing, and assures the smoothness of F_{gf}.

The overall system response, what is left after the local solution has been subtracted out, is the complementary glo-

bal solution, w_g. This solution is obtained by replacing the discrete force in the original formulation by F_{gf}.

$$L(w_g)=F_{gf}. \tag{23}$$

The smooth force, F_{gf}, called the global forcing pressure, is simply obtained from Eq. (22) where L, w_l, and F_d are all known.

Linearity requires that the total solution is unchanged, $w_l+w_g=w$ [by adding Eqs. (22) and (23)]. The advantage of this process is due to the efficient expression of w_l which contains the high-resolution features that would limit convergence. Eq. (23) is solved instead of Eq. (21) and requires far less computational resolution.

Previously, ANM has been demonstrated in the case of mathematically two-dimensional structural acoustic problems.[25,26] Although somewhat different, a more detailed and perhaps more direct explanation of the method can be found in this previous work. In two dimensions, the ANM local displacement is easily generated as a function of one variable. In the present three-dimensional case, however, variance along the line of constraint makes direct application of the previous method impossible. The simplest solution for the present case is to reduce the problem to a set of two-dimensional ones. This is possible because of the particular uncoupled nature of the shell, fluid, and constraint influence with respect to azimuthal modal decomposition. Azimuthal decomposition of the problem can be performed straight away, giving rise to a set of independent two-dimensional problems. Each two-dimensional problem can then be treated as a smoothing in one coordinate.

A. Reducing to set of two-dimensional problems

This special three-dimensional problem is reduced to a set of independent two-dimensional problems by modal decomposition in azimuth. The azimuthal expansions for shell displacement are as follows. Similar expansions represent all other quantities:

$$\begin{Bmatrix} \bar{u}(\bar{z},\theta) \\ \bar{v}(\bar{z},\theta) \\ \bar{w}(\bar{z},\theta) \end{Bmatrix} = \sum_{n=0}^{\infty} \begin{Bmatrix} \bar{u}_n(\bar{z})\cos(n\theta) \\ \bar{v}_n(\bar{z})\sin(n\theta) \\ \bar{w}_n(\bar{z})\cos(n\theta) \end{Bmatrix}. \tag{24}$$

These azimuthal expansions result in the following independent set of θ-reduced governing equations for each azimuthal order of shell displacement from Eqs. (1)–(3):

$$-\nu\frac{d\bar{w}_n}{d\bar{z}}-\frac{d^2\bar{u}_n}{d\bar{z}^2}+\frac{1-\nu}{2}n^2\bar{u}_n-\frac{1+\nu}{2}n\frac{d\bar{v}_n}{d\bar{z}}-\bar{K}_p^2\bar{u}_n=\bar{p}_{un}, \tag{25}$$

$$n\bar{w}_n+\frac{1+\nu}{2}n\frac{d\bar{u}_n}{d\bar{z}}-\frac{1-\nu}{2}\frac{d^2\bar{v}_n}{d\bar{z}^2}+n^2\bar{v}_n-\bar{K}_p^2\bar{v}_n=\bar{p}_{vn}, \tag{26}$$

$$\nu\frac{d\bar{u}_n}{d\bar{z}}+n\bar{v}_n+\bar{w}_n+\bar{\beta}^2\left(\frac{d^2}{d\bar{z}^2}-n^2\right)^2\bar{w}_n-\bar{K}_p^2\bar{w}_n = \bar{p}_{wn}-\bar{p}_{fn}. \tag{27}$$

Recall that the influence of the discrete ring constraint can be represented by a similar expansion in θ, with a Dirac delta function modeling the z dependence [Eqs. (16) and (17)]. This yields the following applied pressure field for a single azimuthal order:

$$\bar{p}_{wn}^{\text{ring}}=\bar{A}_n^{\text{ring}}\delta(z-q2\pi a), \tag{28}$$

$$\bar{p}_{vn}^{\text{ring}}=\bar{B}_n^{\text{ring}}\delta(z-q2\pi a). \tag{29}$$

The influence of this discontinuity in z will be replaced by the smoothed ANM composite solution for each azimuthal order n using the preceding θ-reduced shell equations.

The following development of the ANM problem formulation applies to the θ-reduced system. In other words, the ANM formulation is carried out on one general azimuthal component of the total problem.

B. Local solution

Each of the two pertinent components (radial and tangential) of the discrete ring load causes unique reactions in the shell at their point of application. In each case, ANM moves this high-resolution effect into an analytic local solution to enhance the efficiency of the overall problem calculation.

The reaction of the structure at the point of application of each discrete load component is found by integrating the inhomogeneous governing equation across this point and taking the limit as the range of integration is reduced to include only the point of application. This procedure is demonstrated in the case of a radially applied load at $z=0$. Integrating the θ-reduced radial governing equation [Eq. (27)] across the discrete radial load for azimuthal order n from $\tilde{z}=-\epsilon$ to $\tilde{z}=\epsilon$ yields the following:

$$\int_{-\epsilon}^{\epsilon}(n\bar{v}_n+\bar{w}_n-\bar{K}_p^2\bar{w}_n+n^4\bar{w}_n)\,d\bar{z} + \left[\nu\bar{u}_n+\bar{\beta}^2\left(\frac{d^3\bar{w}_n}{d\bar{z}^3}-\frac{d\bar{w}_n}{d\bar{z}}\right)\right]_{-\epsilon}^{\epsilon} = \int_{-\epsilon}^{\epsilon}(\bar{A}_n^{\text{ring}}\delta(\bar{z})-\bar{p}_f)\,d\bar{z}. \tag{30}$$

Taking the limit as $\epsilon\to 0$ assuming $\bar{u}$ and $d\bar{w}/d\bar{z}$ are continuous, and $\bar{v},\bar{w}$, and $\bar{p}_f$ are finite, yields the following jump in derivative of radial displacement:

$$\bar{\beta}^2\left[\frac{d^3\bar{w}_n}{d\bar{z}^3}\right]_{0^-}^{0^+}=\bar{A}_n^{\text{ring}}. \tag{31}$$

By similar procedure applied to Eq. (26), the relationship for the tangentially applied force is given as follows:

$$-\frac{1-\nu}{2}\left[\frac{d\bar{v}_n}{d\bar{z}}\right]_{0^-}^{0^+}=\bar{B}_n^{\text{ring}}. \tag{32}$$

These relationships allow the discrete loads to be included in the ANM local solution by prescription. Note that the coefficients $\bar{A}_n^{\text{ring}}, \bar{B}_n^{\text{ring}}$ will merely scale the entire ANM subproblem. Calculation of these coefficients is performed ex-

actly as in Sec. II C. We could, therefore, equally well think of treating a discrete force of unit strength.

The generation of each ANM composite solution begins with the construction of the local solution. A polynomial form is assumed in the positive half of the smoothing width for the local shell displacements in each coordinate direction, $\overline{u}_{ln}$, $\overline{v}_{ln}$, and $\overline{w}_{ln}$. These local displacements are completed within the smoothing region by the appropriate even or odd reflection across the discontinuity. Outside the smoothing region, local displacement is prescribed to be zero. For convenience in generating the local solution, the axial coordinate is chosen with the discontinuity of interest at zero and is normalized by the smoothing half-width,

$$\tilde{z}=z/\sigma=\overline{z}/\overline{\sigma}, \tag{33}$$

where σ is the smoothing half-width. The local radial displacement, which is even, is given as an example of the assumed polynomial form as follows. A similar expression is assumed for the local tangential displacement, and also for the local axial displacement except that axial displacement is odd:

$$\overline{w}_{ln}=\begin{cases}\displaystyle\sum_{q=0}^{Q} c_q|\tilde{z}|^q, & |\tilde{z}|<1,\\ 0, & |\tilde{z}|>1.\end{cases} \tag{34}$$

Once the form of the local solution is determined, the appropriate requirements of its properties must be formulated. These property requirements are written in a manner consistent with the chosen form, often as continuity of derivative conditions at the boundaries of the description regions. Once the property requirements are explicitly written, the assumed form can be substituted in order to arrive at equations in terms of the unknown coefficients of the assumed local solution. These coefficients are then found by solving the resulting system of linear equations. The formulation of the property requirements proceeds as follows.

The discrete load applied to the system, and therefore its high-resolution effect, is included in the local solution by requiring the appropriate jump in derivative value across the forcing point. For example, the discrete radial load is included by requiring the previously derived jump in third derivative of radial displacement, Eq. (31). Also, there must be no extraneous discrete loads in the local solution. These loads are disallowed by ensuring continuity of the appropriate derivatives of displacement.

Continuity conditions must be applied in order to ensure a physically compatible local solution. Displacement in each coordinate direction and slope of radial displacement must be continuous.

The monopole strength of the local solution is eliminated by requiring no net area under the curve of radial local displacement. This condition significantly reduces the extent of influence of the local solution which will be chosen to be acoustically compact.

Finally, using the governing equations, conditions are imposed on the local solution to guarantee that the global forcing pressure, which is a result of the forcing necessary to create the local solution, is smooth in all three coordinate directions. Namely, the value and slope of the applied forcing pressure necessary to create the local solution are chosen to be continuous. This condition enhances the convergence rate of the global solution.

Some of the conditions required of the local solution are satisfied automatically at $\tilde{z}=0$ by the symmetry or antisymmetry. In general, though, each requirement leads to a linear equation in terms of the coefficients of the assumed polynomials. The resulting system of linear equations determines the local solution.

In the case of a radially acting ring force, a local radial displacement is found as a result of ten explicitly written independent constraint equations, $Q=9$. No displacement is required in the axial or tangential directions. The ten explicit constraint equations which determine radial local displacement in the upper half of the smoothing region are as follows:

$$\overline{\beta}^2\frac{1}{\overline{\sigma}^3}\left.\frac{d^3\overline{w}_{ln}}{d\tilde{z}^3}\right|_0=\frac{1}{2}\overline{A}_n^{\text{ring}}, \tag{35}$$

$$\left.\frac{d^3\overline{w}_{ln}}{d\tilde{z}^3}\right|_1=0, \tag{36}$$

$$\left.\frac{d^2\overline{w}_{ln}}{d\tilde{z}^2}\right|_1=0, \tag{37}$$

$$\left.\frac{d\overline{w}_{ln}}{d\tilde{z}}\right|_0=0, \tag{38}$$

$$\left.\frac{d\overline{w}_{ln}}{d\tilde{z}}\right|_1=0, \tag{39}$$

$$\overline{w}_{ln}|_1=0, \tag{40}$$

$$\int_0^1\overline{w}_{ln}d\tilde{z}=0, \tag{41}$$

$$\overline{p}_{wn}|_1=0, \tag{42}$$

$$\left.\frac{d\overline{p}_{wn}}{d\tilde{z}}\right|_0=0, \tag{43}$$

$$\left.\frac{d\overline{p}_{wn}}{d\tilde{z}}\right|_1=0. \tag{44}$$

Note that $\overline{p}_{wn}$ is the result of substitution of the local displacement into Eq. (27). Some of the preceding equations are shown in terms of $\overline{p}_{wn}$ for clarity, but when solving for the local solution, Eq. (27) is used to represent these relationships in terms of the local displacement. For instance, Eq. (42) becomes the following (leaving out fluid pressure which is known to be continuous):

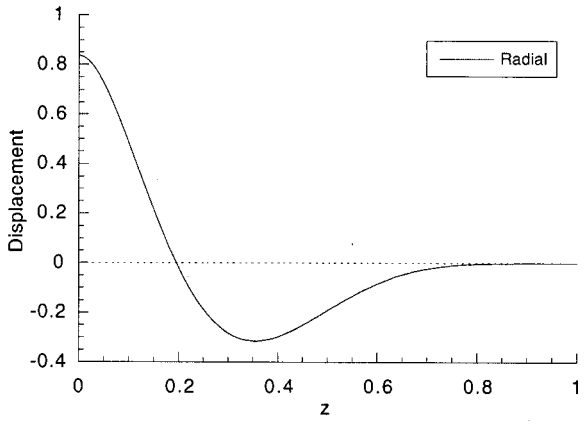


FIG. 4. Local displacement: radially acting ring force.

$$\overline{p}_{wn}|_1=\left[-\nu\frac{d\overline{u}_{ln}}{d\overline{z}}-n\overline{v}_{ln}-\overline{w}_{ln}-\overline{\beta}^2\left(\frac{d^2}{d\overline{z}^2}-n^2\right)^2\overline{w}_{ln}+\overline{K}_p^2\overline{w}_{ln}\right]_1=0. \tag{45}$$

Using the information that $\overline{u}_{ln}$ and $\overline{v}_{ln}$ are zero, and Eqs. (37) and (40), this reduces to a direct constraint on $\overline{w}_{ln}$ as follows:

$$\left.\frac{d^2\overline{w}_{ln}}{d\tilde{z}^2}\right|_1=0. \tag{46}$$

In the present case, the pressures $\overline{p}_{un}$ and $\overline{p}_{vn}$ are continuous and continuous in slope without the need for additional constraint equations.

The assumed form for local displacement is now substituted into the previous equations and the indicated operations performed. The result is a set of linear equations in terms of the unknown polynomial coefficients, c_q. The polynomial order is set by the number of independent constraints. The equations are arranged in matrix form and solved for the polynomial coefficients. The polynomial for local radial displacement in the case of a radial ring force of azimuthal order n is given below. This displacement is illustrated in Fig. 4 with $n=4$, $\overline{\sigma}=\pi/6$ and the previously used physical parameters:

$$\overline{w}_{ln}=\overline{A}_n^{\text{ring}}\frac{\overline{\sigma}^3}{12\overline{\beta}^2}\left[\frac{5}{1344}-\frac{27}{112}\tilde{z}^2+\tilde{z}^3-\frac{45}{32}\tilde{z}^4+\frac{35}{16}\tilde{z}^6-\frac{18}{7}\tilde{z}^7+\frac{81}{64}\tilde{z}^8-\frac{5}{21}\tilde{z}^9+\frac{n^2\overline{\sigma}^2}{32}\left(-\frac{1}{630}+\frac{2}{35}\tilde{z}^2-\tilde{z}^4+\frac{16}{5}\tilde{z}^5-\frac{14}{3}\tilde{z}^6+\frac{128}{35}\tilde{z}^7-\frac{3}{2}\tilde{z}^8+\frac{16}{63}\tilde{z}^9\right)\right]. \tag{47}$$

The development for the case of a tangential ring force follows similarly. The primary difference is that nonzero local displacement is required in all three coordinate directions to satisfy the prescribed requirement conditions. This local solution is given in the Appendix.

C. Global solution

For each azimuthal order, the global solution is essentially a smoothed version of the original problem. The global solution is calculated by the same conventional methods as the original discretely forced problem, except that the structure is driven by a new smooth pressure distribution, called the global forcing pressure. Because the system is driven by a smooth forcing rather than a discrete one it converges more rapidly.

The global forcing pressure is derived from the local solution. The local solution is, by the definition of the problem decomposition, the difference between the solution of the original discretely driven problem and the solution of the smooth global problem. Therefore, by linearity, the forcing that is required to create the local solution is the difference between the original discrete forcing and the global driving pressure. Substituting the known local displacement into the governing equations [Eqs. (25)–(27)] yields the external forcing that is required to create it. Solving for the global forcing pressure is a simple matter of rearranging terms in these equations. The global forcing pressure, $\overline{p}_{un}^{gf}$, $\overline{p}_{vn}^{gf}$, and $\overline{p}_{wn}^{gf}$, is given in general by the following equations. Often, many of the terms in these equations are zero such as in the case of the radial ring force, where only radial local displacement, $\overline{w}_{ln}$, is nonzero:

$$\overline{p}_{un}^{gf}=\nu\frac{d\overline{w}_{ln}}{d\overline{z}}+\frac{d^2\overline{u}_{ln}}{d\overline{z}^2}-\frac{1-\nu}{2}n^2\overline{u}_{ln}+\frac{1+\nu}{2}n\frac{d\overline{v}_{ln}}{d\overline{z}}+\overline{K}_p^2\overline{u}_{ln}, \tag{48}$$

$$\overline{p}_{vn}^{gf}=-n\overline{w}_{ln}-\frac{1+\nu}{2}n\frac{d\overline{u}_{ln}}{d\overline{z}}+\frac{1-\nu}{2}\frac{d^2\overline{v}_{ln}}{d\overline{z}^2}-n^2\overline{v}_{ln}+\overline{K}_p^2\overline{v}_{ln}+\overline{p}_{vn}^{\text{ring}}, \tag{49}$$

$$\overline{p}_{wn}^{gf}=-\nu\frac{d\overline{u}_{ln}}{d\overline{z}}-n\overline{v}_{ln}-\overline{w}_{ln}-\overline{\beta}^2\left(\frac{d^2}{d\overline{z}^2}-n^2\right)^2\overline{w}_{ln}+\overline{K}_p^2\overline{w}_{ln}+\overline{p}_{fn}+\overline{p}_{wn}^{\text{ring}}. \tag{50}$$

The original applied discrete load, $\overline{p}_{wn}^{\text{ring}}$ or $\overline{p}_{vn}^{\text{ring}}$, is canceled by the correctly prescribed discontinuous derivative of local displacement (which causes an infinite pressure of equal integrated strength and opposite sense with respect to the original discrete load when one more derivative is taken in the equation). A smooth global forcing pressure results.

The formulation results in a piecewise polynomial representation of the dominant part of the global forcing pressure, that due to structural terms. Except for fluid loading, each of the terms in Eqs. (48)–(50) is just differentiation of the polynomial local displacement. As an example, the global forcing pressure due to structural terms for the case of radial ring force [complementing the local displacement in Eq. (47)] is plotted in Fig. 5. The component of global forcing pressure due to fluid pressure is just the fluid pressure due to local displacement. This pressure is always smooth, and is generally small compared to the structural component. This pressure can be calculated by the same method, and in

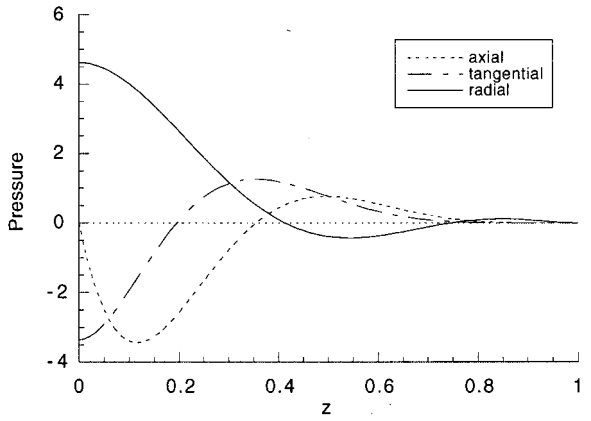


FIG. 5. Global forcing pressure: radially acting ring force.

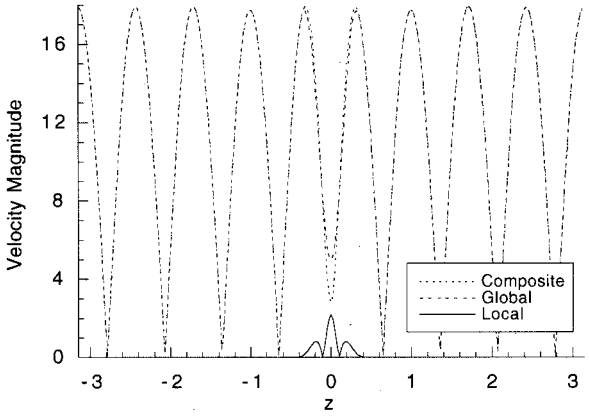


FIG. 6. Radial displacement: exact, global, and local.

fact at the same time, as the fluid pressure coupling to shell motion that must be calculated in the general solution of the forced shell. This fluid pressure has little to do with the smoothing of the discrete force, it is only important that it not be overlooked when calculating the coupled shell vibration.

The modal method described in Sec. II A is used to compute the global problem. The global forcing pressure is known, and its Fourier decomposition is given by Eq. (13). These integrals can be performed analytically. In fact, using a formula derived from integration by parts, Fourier decomposition of the piecewise polynomials can be determined without integration (see Ref. 31, Sec. 10.6). Only the values of derivative jumps in the function (of continual concern to this method) are necessary. Because the Fourier coefficients of global forcing pressure are expressed analytically, they require very little computational overhead to compute. There is, however, significant cancellation of terms in these analytic expressions that require a series approximation for small argument (small mode number times smoothing width) to eliminate numerical precision issues.

D. Composite solution

For each azimuthal order, assembling the composite solution is a matter of superposition. The local and global solutions are simply added throughout the domain. The local solution is a piecewise polynomial description of shell motion and the associated fluid reaction. The global solution is the result of a conventional solution of the driven system, but for the case of the special global forcing pressure. The series representation of the global solution contains the overall, low-resolution shell motion, while the polynomial description of the local solution contains high-resolution content in the immediate vicinity of the discrete load.

As a representative example, the complete solution is shown together with the two ANM constituent solutions for radially acting ring force and the representative parameter case in Figs. 6 and 7. Figure 6 shows the composite, global, and local radial displacement magnitudes. Figure 7 shows the harmonic decomposition of each of the solutions. Comparison of the harmonic magnitudes of the two ANM constituent solutions and the original discretely driven solution shows efficient separation of the shell response into high- and low-resolution parts. The local solution captures the higher harmonic content, or the high-resolution part of the solution, while the global solution captures the lower harmonic modes, or low-resolution content. Removal of the dominant high harmonic content in the global solution allows the harmonics of the global solution to fall at a much faster rate, which, in turn, causes the solution to converge more rapidly.

Varying the smoothing width changes the point at which the local solution begins to dominate the harmonic content of the composite solution. This is illustrated by Fig. 8, which shows the harmonic magnitudes for several choices of smoothing width. If the width is decreased, the point of crossover is shifted towards higher harmonic number. Choosing a smoothing width that is similar to half the length of the small structural wave in the problem produces good convergence results. However, the ANM composite solution given in the present work is without approximation for any choice of smoothing. The choice of smoothing width affects the separation of features and the rate of convergence, but not the converged solution itself.

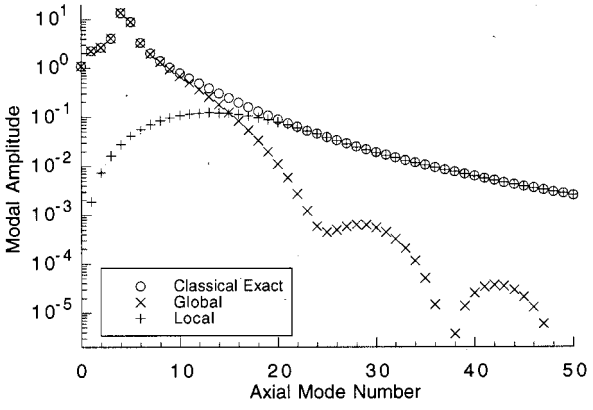


FIG. 7. Radial displacement coefficients: exact, global, and local.

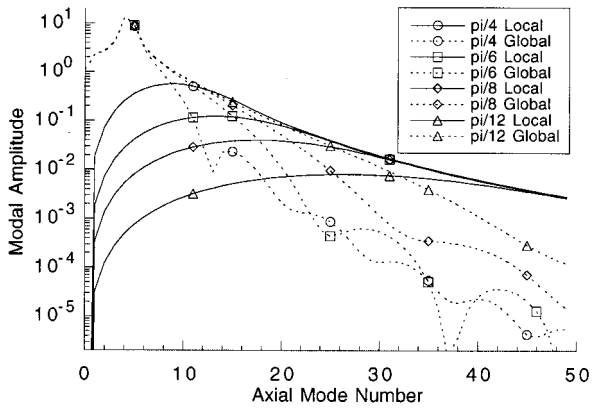


FIG. 8. Radial displacement coefficients: several smoothing widths.

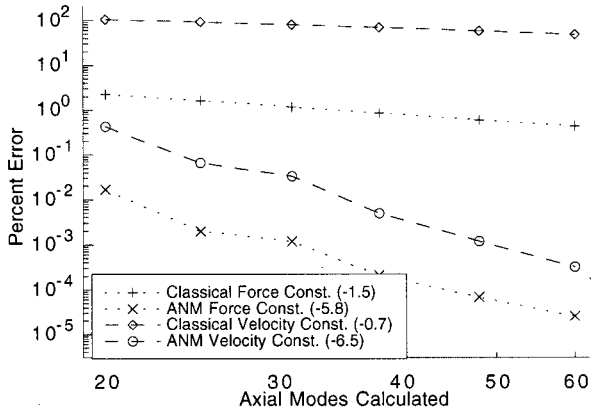


FIG. 10. Convergence of tangential velocity: tangential ring force.

E. Comparison of classical and ANM driven solutions

For each azimuthal mode the discrete influences of the discontinuity were smoothed using the ANM method. In this section, the effectiveness of the ANM treatment of the forcing components is examined.

The convergence, and therefore the computational efficiency of the method, is evaluated by plotting on log-log scales the percent error in surface velocity, averaged in azimuth using an L2 norm (or generalized rms value), versus the number of modes calculated. Error convergence for radial and tangential forcing cases are shown in Figs. 9 and 10 for the representative case. Error of the velocity component in the same direction as the driving force is calculated. In the case of a single driven mode, however, error magnitude does vary significantly dependent upon the particular parameters chosen. The order of convergence is estimated by linear regression of log of error versus log of modes, the slope of which is indicated on the legend. Error is measured relative to a solution that is converged well beyond the scale of error plotted. In each case, error convergence is plotted for the ANM composite driven solution and for the classical discretely driven solution for comparison. Two types of driven problems are shown as well. First, the strength of the discrete forcing is held constant. Second, the response (velocity) of the structure at the point of application of the discrete forcing is held constant. The constant forcing strength case is perhaps the more obvious case; however, the constant response case is pertinent with regard to a constrained system because the constraint condition is written in terms of shell response. The ANM solutions are uniformly more accurate and converge several times faster than the classical modal solutions.

In general, the case of constant drive point response is more difficult to converge than the case of constant forcing strength. In the case of constant forcing strength, the first several modes of the solution are evaluated without error regardless of the higher harmonic content. Therefore, the overall shape of the solution is correct with only a few modes calculated and only the local region about the forcing point is slow to converge. In the constant response case, though, the entire solution, even the first several modes, depends directly on the response of the system at the forcing point. Therefore, the overall solution is slower to converge. Constraining the drive point velocity leads to significant error throughout because a generally well-behaved solution is scaled by an isolated quantity that is difficult to converge.

The prevalence of the local solution and the effect of the ANM smoothing in the radially driven case is relatively small compared to the tangential case. This is due to the stiffness of the shell to bending, which already acts as a fairly efficient mechanical filter to the radially applied discrete load. The tangential local solution sees far less stiffness and contains more significant high-resolution effects. In situations where both discrete forces are needed to satisfy the constraint conditions, the tangential discrete loads generally hinder convergence more than the radial.

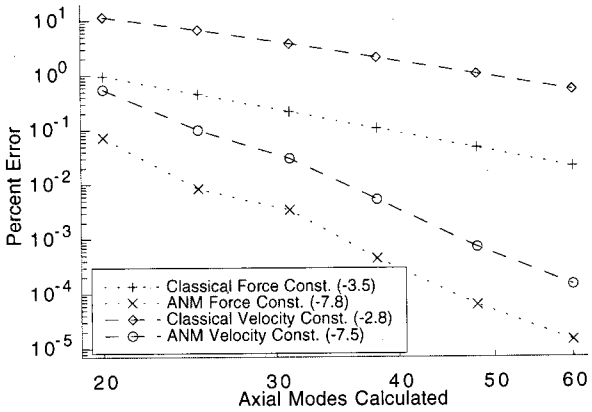


FIG. 9. Convergence of radial velocity: radial ring force.

IV. SCATTERING RESULTS AND DISCUSSION

The assembled three-dimensional ring constrained scattering problem consists of the unconstrained problem plus a host of properly scaled discretely forced problems. For each mode in azimuth, radial and tangential ring driven solutions are added with the proper scaling to eliminate motion at the bulkheads. Each of the several discretely driven problems is sensitive to convergence, especially because the scaling strength is dependent upon the response value right at the point of discrete force application. Assembling the full constrained problem yields the convergence behavior of radial

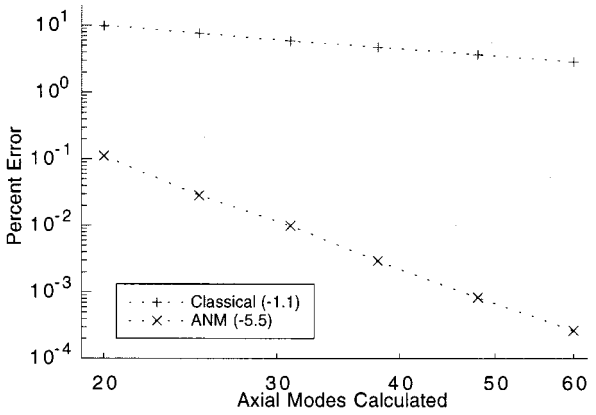


FIG. 11. Convergence of radial velocity: constrained scattering.

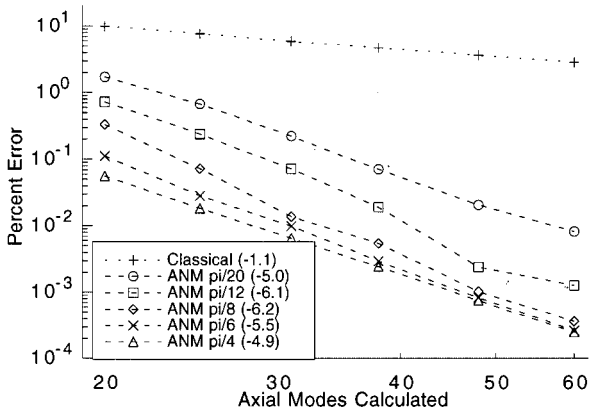


FIG. 13. Convergence of radial velocity: several smoothing widths.

velocity shown in Fig. 11 for the representative case. The rate limiting factor with regard to convergence is the constraint, which introduces sensitivity to the local details of response. Convergence of scattered pressure is shown in Fig. 12. The scattered pressure is a quantity of significant interest. It is also slow to converge due to the sensitivity of even the lowest harmonics to the fine scale details of response, because of the local nature of the discrete constraint.

Consider the issue of choosing the smoothing width in the ANM formulation. The convergence results for several smoothing widths are shown in Fig. 13. The primary effect of increasing the smoothing width is to shift the errors lower. This is because the analytic local solution captures more of the solution in Fourier space. The point at which the global solution begins to fall off (see Fig. 8 for a sample azimuthal mode) is shifted to lower mode number. Smoothing width should be chosen to be approximately as large as the smallest of the characteristic length scales in the problem other than that of the discontinuity. Figure 2, the plot of radial surface velocity, indicates that the characteristic wavelength of the rapid variations of shell motion is approximately $\frac{1}{4}$ the circumference. Figure 13 verifies that a smoothing width close to the small structural wavelength produces good convergence results.

It should be noted that the present ANM formulation can be called exact in the sense that no approximation is made to divide the discretely forced problem into local and global solutions. In the limit of infinite modes calculated, the ANM result is equivalent to the classical modal solution. For this reason, it is possible to use essentially any smoothing width. The guideline of choosing the smoothing width to be smaller than the characteristic length scales of the problem can even be violated. In the case of such an impractically large local solution, although the problem would not be separated as was intended, the composite solution would still converge to the correct result.

Consider the choice of the degree of smoothness required in the generation of the local solution. The ANM formulation until now has required continuity and continuity of slope of global forcing pressure. It is possible, however, to create a useful local solution with less stringent requirements. It is also possible to require higher derivatives to be continuous. Figure 14 shows the convergence of the constrained scattering problem with several ANM formulations and the classical method for comparison ($\bar{\sigma}=\pi/8$ here). The version requiring continuity and continuity of slope, a version requiring only continuity, and a version that does not require continuity of global forcing pressure are all shown. All versions lead to significant gains in accuracy and convergence rate. Additional degrees of smoothness tend to yield increased convergence rates. This is expected because func-

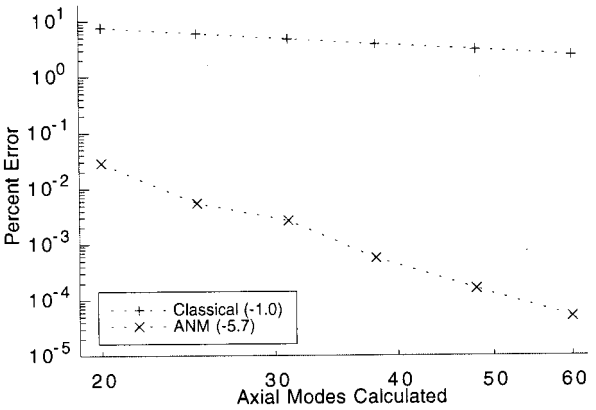


FIG. 12. Convergence of scattered pressure: constrained scattering.

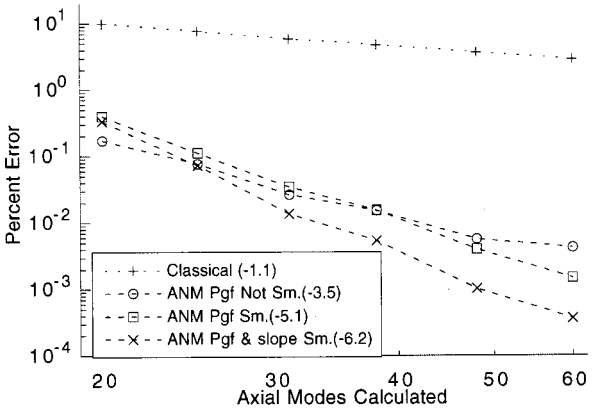


FIG. 14. Convergence of radial velocity: several ANM formulations p_{gf} discontinuous, continuous, and slope continuous.

tions that are continuous to higher derivatives generally yield Fourier coefficients that fall by greater power of mode number. The absolute error of the less smooth versions is perhaps surprisingly low. In fact, with only 20 modes calculated, the least smooth version is most accurate, even though it exhibits the slowest rate of convergence. The less smooth version can be more accurate under certain conditions because it displays an effectively larger smoothing region, shifting error lower. Additional smoothness constraints create a stiffer local solution (stiff is used here in the same sense as when discussing splines) which leads to a length at the edge of the smoothing region with nearly zero displacement, effectively narrowing the local displacement.

The primary disadvantage of increasing the level of smoothness is that the complication involved in the ANM formulation increases sharply. Conversely, less smooth solutions can be much simpler. For instance, if only continuity of global forcing pressure is required instead of both continuity and continuity of slope, the resulting local solutions are significantly simpler. Local displacement for radially acting ring force is given as follows for this less stringent condition [as compared to Eq. (47)].

$$\overline{w}_{ln}=\overline{A}_n^{\text{ring}}\frac{\overline{\sigma}^3}{12\overline{\beta}^2}\left[\frac{1}{280}-\frac{9}{40}\tilde{z}^2+\tilde{z}^3-\frac{15}{8}\tilde{z}^4+\frac{9}{5}\tilde{z}^5-\frac{7}{8}\tilde{z}^6+\frac{6}{35}\tilde{z}^7\right]. \tag{51}$$

The order of the polynomial is reduced from ninth to seventh, and the dependence of the local solution on mode number is eliminated. In the case of tangential ring force (local displacements given in the Appendix), the order of the local solution polynomials is reduced, the dependence on mode number is eliminated, and the need for nonzero local radial displacement is eliminated. This version requiring only continuity of global forcing pressure is perhaps the best trade off. It is significantly simpler, and as shown in Fig. 14, retains most of the convergence advantage of the more complicated case.

In the event of further relaxation of the smoothness requirement, where continuity of global forcing pressure is not even enforced, the local solutions become very simple. The local displacement for radially acting ring force for this version is given below:

$$\overline{w}_{ln}=\overline{A}_n^{\text{ring}}\frac{\overline{\sigma}^3}{12\overline{\beta}^2}\left[\frac{1}{160}-\frac{9}{32}\tilde{z}^2+\tilde{z}^3-\frac{45}{32}\tilde{z}^4+\frac{9}{10}\tilde{z}^5-\frac{7}{32}\tilde{z}^6\right]. \tag{52}$$

For this version, in the case of both radial and tangential ring forces, polynomial orders are significantly reduced, and no local displacement is necessary in any coordinate but the one the discrete force acts in. However, convergence rate is significantly compromised.

V. CONCLUSION

The solution of a representative problem of acoustic scattering from an elastic shell with structural constraints was obtained very efficiently using the method called analytical/numerical matching (ANM). The dominant part of the high-wave-number portion of the solution is captured in an artificial, efficiently expressed, analytic local solution. The remainder of the solution, called the global solution, consequently converges much more rapidly. The ANM composite solution exhibits orders of magnitude smaller error for a given number of modes summed and a rate of convergence that is as much as five times faster. The error level achieved by the ANM method with 20 modes calculated requires hundreds of modes to achieve using the classical formulation. While the current method has demonstrated success in the case of a three-dimensional problem geometry, it relies on a modal decomposition that is a special feature of the current geometry. An ANM composite Green's function approach, which is the subject of ongoing research,[27] will alleviate this reliance the special geometry.

ACKNOWLEDGMENTS

This work was sponsored by the Office of Naval Research. The authors would especially like to thank Dr. Luise Couchman, the technical monitor.

APPENDIX: TANGENTIAL FORCE LOCAL SOLUTION

The local solution for a single azimuthal mode of tangentially acting ring force is given as follows. Tangential and radial displacements are even and axial displacement is odd. The local solution begins with the proportionality between applied force and first derivative of $\overline{v}_l$, as in Eq. (32). The additional requirements of the local solution are set following the procedure outlined in Sec. III B:

$$\overline{u}_{ln}=\overline{B}_n^{\text{ring}}\frac{n\overline{z}_s^2}{4}\frac{1+\nu}{1-\nu}\left[-\frac{1}{4}\tilde{z}+\tilde{z}^2-\frac{3}{2}\tilde{z}^3+\tilde{z}^4-\frac{1}{4}\tilde{z}^5\right], \tag{A1}$$

$$\overline{v}_{ln}=\overline{B}_n^{\text{ring}}\frac{\overline{z}_s}{1-\nu}\left[\frac{3}{10}-\tilde{z}+\tilde{z}^2-\frac{1}{2}\tilde{z}^4+\frac{1}{5}\tilde{z}^5+\frac{z_s^2}{3(1-\nu)}\left(\overline{K}_p^2-n^2+\frac{n^2(1+\nu)^2}{4}\right)\times\left(\frac{1}{20}-\frac{1}{2}\tilde{z}^2+\tilde{z}^3-\frac{3}{4}\tilde{z}^4+\frac{1}{5}\tilde{z}^5\right)\right], \tag{A2}$$

$$\overline{w}_{ln}=\overline{B}_n^{\text{ring}}\frac{M}{120\overline{\beta}^2}\left[-\frac{20}{2016}+\frac{1}{56}\tilde{z}^2-\frac{5}{16}\tilde{z}^4+\tilde{z}^5-\frac{35}{24}\tilde{z}^6+\frac{8}{7}\tilde{z}^7-\frac{15}{32}\tilde{z}^8+\frac{5}{63}\tilde{z}^9\right]. \tag{A3}$$

With only the continuity of global forcing pressure required, the nonzero local displacements are as follows:

$$\overline{u}_{ln}=\overline{B}_n^{\text{ring}}\frac{n\overline{z}_s^2}{4}\frac{1+\nu}{1-\nu}\left[-\frac{1}{3}\tilde{z}+\tilde{z}^2-\tilde{z}^3+\frac{1}{3}\tilde{z}^4\right], \tag{A4}$$

$$\overline{v}_{ln} = \overline{B}_n^{\text{ring}} \frac{\overline{z}_s}{1-\nu}\left[\frac{1}{3} - \tilde{z} + \tilde{z}^2 - \frac{1}{3}\tilde{z}^3\right]. \tag{A5}$$

With no continuity required of global forcing pressure, the nonzero local displacement is as follows:

$$\overline{v}_{ln} = \overline{B}_n^{\text{ring}} \frac{\overline{z}_s}{1-\nu}\left[\frac{1}{2} - \tilde{z} + \frac{1}{2}\tilde{z}^2\right]. \tag{A6}$$

[1] M. C. Junger, ''Sound scattering by thin elastic shells,'' J. Acoust. Soc. Am. **24**, 366–373 (1952).
[2] M. C. Junger, ''Vibrations of elastic shells in a fluid medium and the associated radiation of sound,'' J. Appl. Mech. **19**, 439–445 (1952).
[3] M. C. Junger, ''The concept of radiation scattering and its application to reinforced cylindrical shells,'' J. Acoust. Soc. Am. **25**, 890–903 (1953).
[4] M. C. Junger, ''Dynamic behavior of reinforced cylindrical shells in a vacuum and in a fluid,'' J. Appl. Mech. **53**, 35–41 (1954).
[5] A. Klauson and J. Mesaveer, ''Sound scattering by cylindrical shell reinforced by lengthwise ribs and walls,'' J. Acoust. Soc. Am. **91**, 1834–1843 (1992).
[6] J. F. M. Scott, ''The free modes of propagation of an infinite fluid-loaded cylindrical shell,'' J. Sound Vib. **125**, 241–280 (1988).
[7] D. M. Photiadis, ''The propagation of axisymmetric waves on a fluid-loaded cylindrical shell,'' J. Acoust. Soc. Am. **88**, 239–250 (1990).
[8] M. L. Rumerman, ''Contribution of membrane wave reradiation to scattering from finite cylindrical steel shells in water,'' J. Acoust. Soc. Am. **93**, 56–65 (1993).
[9] Y. P. Guo, ''Sound scattering by bulkheads in cylindrical shells,'' J. Acoust. Soc. Am. **95**, 2550–2559 (1994).
[10] E. A. Skelton, ''Acoustic scattering by a disc constraining an infinite fluid-loaded cylindrical shell,'' J. Sound Vib. **148**, 243–264 (1991).
[11] D. J. Mead and N. S. Bardell, ''Free vibration of a thin cylindrical shell with periodic circumferential stiffners,'' J. Sound Vib. **115**, 499–520 (1987).
[12] J. M. Cushieri and D. Feit, ''A hybrid numerical and analytical solution for the Green's function of a fluid-loaded elastic plate,'' J. Acoust. Soc. Am. **95**, 1998–2005 (1994).
[13] J. M. Cushieri and D. Feit, ''A hybrid solution for the response Green's function of a fluid-loaded cylindrical shell,'' J. Acoust. Soc. Am. **96**, 2776–2784 (1994).
[14] J. M. Cushieri and D. Feit, ''Acoustic scattering from a fluid-loaded cylindrical shell with discontinuities: Single plate bulkhead,'' J. Acoust. Soc. Am. **98**, 320–338 (1995).
[15] J. M. Cushieri and D. Feit, ''Acoustic scattering from a fluid-loaded cylindrical shell with discontinuities: Double plate bulkhead,'' J. Acoust. Soc. Am. **98**, 339–352 (1995).
[16] J. L. Humar, *Dynamics of Structures* (Prentice Hall, Engelwood Cliffs, NJ, 1968), pp. 567–572.
[17] R. L. Blisplinghoff and H. Ashley, *Principles of Aeroelasticity* (Dover, New York, 1962), pp. 350, 471.
[18] D. B. Bliss and W. O. Miller, ''Efficient free wake calculations using analytical/numerical matching,'' J. Am. Helicopter Soc. **38**, 870–879 (1993).
[19] T. R. Quackenbush and D. B. Bliss, ''Free wake calculation of rotor flow fields for interactional aerodynamics,'' Vertica **14**, 313–327 (1990).
[20] T. R. Quackenbush, C-M. G. Lam, and D. B. Bliss, ''Vortex methods for computational analysis of rotor/body interaction,'' J. Am. Helicopter Soc. **39**, 14–24 (1994).
[21] D. B. Bliss and R. J. Epstein, ''Free vortex problems using analytical/numerical matching with solution pyramiding,'' AIAA J. **33**, 894–903 (1995).
[22] John A. Rule, ''Rotorcraft acoustic radiation prediction based on a refined blade-vortex interaction model,'' doctoral dissertation, Duke University, 1997.
[23] D. B. Bliss and R. J. Epstein, ''A novel approach to aerodynamic analysis using analytical/numerical matching,'' AIAA J. **34**, 2225–2232 (1996).
[24] R. J. Epstein and D. B. Bliss, ''An acoustic boundary element method using analytical/numerical matching,'' J. Acoust. Soc. Am. **101**, 92–106 (1997).
[25] R. C. Loftman and D. B. Bliss, ''The application of analytical/numerical matching to structural discontinuities in structural/acoustic problems,'' J. Acoust. Soc. Am. **101**, 925–935 (1997).
[26] R. C. Loftman and D. B. Bliss, ''Analytical/numerical matching for efficient calculation of scattering from cylindrical shells with lengthwise constraints,'' J. Acoust. Soc. Am. **103**, 1885–1896 (1998).
[27] R. C. Loftman, ''A general method for efficient treatment of structural discontinuities in structural acoustic vibration problems,'' doctoral dissertation, Duke University, 1998.
[28] M. C. Junger and D. Feit, *Sound, Structures, and Their Interaction* (MIT, Cambridge, MA, 1972).
[29] H. Kraus, *Thin Elastic Shells* (Wiley, New York, 1967).
[30] M. Abromowitz and I. A. Stegun, *Handbook of Mathematical Functions* (U.S. Bureau of Standards, Washington DC, 1964).
[31] E. Kreyszig, *Advanced Engineering Mathematics* (Wiley, New York, 1979).

Comparison of theoretical and experimental ultrasonic velocity in three-component mixtures of methyl ethyl ketone and toluene with *n*-alkanols and estimation of acoustic nonlinearity parameter (B/A)

Roshan Abraham
School of Pure and Applied Physics, Mahatma Gandhi University, Priyadarsini Hills P.O., Kottayam-686 560, Kerala, India

M. Abdul Khadar[a)]
Department of Physics, University of Kerala, Kariavattom, Thiruvananthapuram-695 581, Kerala, India

J. Jugan
School of Pure and Applied Physics, Mahatma Gandhi University, Priyadarsini Hills P.O., Kottayam-686 560, Kerala, India

(Received 25 April 1997; revised 20 February 1999; accepted 14 May 1999)

A theoretical estimation of ultrasonic velocity using Flory–Patterson theory (FPT), Jacobson's free length theory, Schaaffs' collision factor theory, Nomoto's relation, and Van Dael's ideal mixture relation (IMR) has been done in the three-component (ternary) mixtures of methyl ethyl ketone and toluene with *n*-alkanols. The theoretically predicted ultrasonic velocities are compared with the experimentally determined velocities in the ternary mixtures at 30 °C. It is found that the velocities evaluated using all the theories except FPT and IMR are in fairly good agreement with the experimental velocity values. Acoustic nonlinearity parameter (B/A) and a few thermodynamic parameters have been evaluated theoretically for the ternary mixtures using the Tong and Dong equation combined with Flory's statistical theory and its extended version to ternary mixtures. © *1999 Acoustical Society of America.* [S0001-4966(99)05508-3]

PACS numbers: 43.25.Ba [MAB]

INTRODUCTION

A variety of empirical, semiempirical, and statistical formulas are available in the literature for predicting ultrasonic velocity in liquids and liquid mixtures. These include Flory–Patterson theory (FPT),[1–3] Jacobson's free length theory (FLT),[4–6] Schaaffs' collision factor theory (CFT),[7,8] Nomoto's relation (NR),[9] and Van Dael's ideal mixture relation (IMR).[10] A literature survey shows that an evaluation of ultrasonic velocity using these theories has been done in different three-component (ternary) liquid mixtures.[11–14] In this report, a comparative study of ultrasonic velocities predicted using FPT, FLT, CFT, NR, and IMR with the experimentally determined ultrasonic velocities has been made for ternary mixtures of methyl ethyl ketone and toluene with *n*-alkanols at 30 °C. It is found that the ultrasonic velocities evaluated using different theories are satisfactory to different extents in different mixtures. The acoustic nonlinearity parameter (B/A) and a few thermodynamic parameters are also evaluated theoretically for the ternary mixtures. The nonlinearity parameter B/A plays a significant role in nonlinear acoustics and hence its determination is of increasing interest in a number of areas ranging from underwater acoustics to medicine. A number of experimental and theoretical studies have been performed on the nonlinearity parameter of liquids, by making use of finite-amplitude[15] or thermodynamic methods.[16] The finite-amplitude method is based on the measurement of the amplitude of the second harmonic generated due to distortion of the propagating sinusoidal wave. Analysis of the finite-amplitude method carried out by Law *et al.* showed the total systematic error to be of the order of ±8%.[17] The calculation of B/A by the thermodynamic method requires the experimental determination of velocity at different temperatures and pressures which involve practical difficulty. It has been shown that an accuracy for the determination of the nonlinearity parameter (B/A) in liquids by means of the conventional thermodynamic method can be about 5%.[17] Such accuracy of both the finite-amplitude and the thermodynamic methods limits the usefulness of direct application of these methods to investigations in liquid mixtures.[18] Therefore, the theoretical evaluation of nonlinearity parameter (B/A) in ternary mixtures by making use of the experimental velocity values at a constant temperature and at atmospheric pressure is important. Tong *et al.*[19] calculated the B/A values for pure liquids by making use of Schaaffs' equation.[7,8] Although detailed studies of the nonlinearity parameter have been made in many pure liquids,[20–22] such studies are rare in the case of liquid mixtures,[23] particularly in ternary liquid mixtures. The method of calculation of B/A used in the present study is based on the Tong and Dong equation[19] combined with Flory's statistical theory[1,2] and its extended version to ternary mixtures.[24]

a)Author to whom correspondence should be addressed.

TABLE I. Experimental and theoretical values of velocity of the ternary mixtures at 30 °C.

Mole fraction								
x_1	x_2	EXPT (m/s)	ρ (g/cc)	FLT (m/s)	CFT (m/s)	FPT (m/s)	NR (m/s)	IMR (m/s)
Methyl ethyl ketone+toluene+methanol								
0.1240	0.4187	1216.00	.8325	1218.39	1194.43	1306.40	1222.04	1005.03
0.1642	0.3814	1209.00	.8306	1219.53	1191.52	1306.64	1216.04	1007.38
0.2068	0.3581	1207.00	.8285	1216.10	1189.83	1304.54	1212.56	1012.02
0.2700	0.3159	1206.00	.8256	1214.25	1186.88	1302.98	1206.12	1017.99
0.3588	0.2674	1207.00	.8215	1208.22	1183.53	1299.02	1199.21	1028.35
0.3912	0.3117	1213.00	.8225	1204.32	1191.08	1296.56	1207.76	1046.38
0.4101	0.3348	1218.00	.8242	1210.18	1196.98	1299.06	1211.90	1059.43
0.4537	0.3833	1229.00	.8266	1216.42	1208.16	1301.33	1219.90	1097.21
Methyl ethyl ketone+toluene+ethanol								
0.1129	0.4006	1211.00	.8261	1217.14	1200.97	1317.59	1215.91	1105.89
0.1570	0.3600	1206.00	.8226	1208.33	1195.38	1312.53	1209.94	1106.31
0.2129	0.3218	1199.00	.8198	1203.29	1191.40	1308.80	1204.60	1108.28
0.2707	0.2954	1198.00	.8173	1196.21	1188.51	1303.74	1201.38	1111.68
0.3294	0.2687	1197.00	.8156	1194.05	1186.76	1301.11	1198.17	1115.23
0.3622	0.2962	1202.00	.8182	1200.57	1193.05	1302.70	1203.64	1122.95
0.3971	0.3253	1209.00	.8221	1215.31	1201.54	1308.16	1209.12	1133.10
0.4416	0.3837	1221.00	.8257	1220.31	1211.88	1307.99	1218.88	1154.68
Methyl ethyl ketone+toluene+propanol								
0.1241	0.3985	1235.00	.8279	1237.76	1231.37	1338.18	1236.12	1193.79
0.1484	0.3760	1233.00	.8268	1237.42	1229.19	1337.70	1233.09	1192.31
0.2210	0.3215	1231.00	.8251	1242.84	1225.14	1338.72	1225.28	1188.98
0.2692	0.2895	1227.00	.8204	1222.22	1217.18	1327.35	1220.51	1187.14
0.3185	0.2691	1226.00	.8195	1222.16	1214.84	1325.39	1216.97	1185.97
0.3745	0.2953	1227.00	.8201	1216.20	1214.54	1317.95	1218.38	1187.57
0.4208	0.3340	1233.00	.8241	1228.42	1219.61	1319.39	1221.30	1190.69
0.4407	0.3723	1237.00	.8273	1236.80	1224.14	1320.44	1224.75	1194.29

I. EXPERIMENT

The systems chosen for the present study are:

(1) Methyl ethyl ketone (MEK)+toluene (TL)+methanol (ML)
(2) Methyl ethyl ketone (MEK)+toluene (TL)+ethanol (EL)
(3) Methyl ethyl ketone (MEK)+toluene (TL)+propanol (PL)

The liquids for the present investigation [MEK is of SD fine chemicals, AR grade, ML, PL, and TL are of SRL, AR grade, and EL is of Merck, absolute GR grade] were used as supplied. The liquid mixtures of different composition were prepared by mixing calculated volumes of each component. The liquid cell was filled with the sample without air bubbles. A piezoelectric (PZ) transducer of resonant frequency 2 MHz was attached to one end of the cell using the gel for general purpose NDT as bonding material. This transducer was coupled to the Matec Ultrasonic Velocity Measuring System (Matec Instruments, Inc.) and the velocity was measured by the pulse echo overlap method. The velocity measurements were accurate to $\pm 0.2\%$. The experiment was done at 30 °C, the temperature being kept constant within ± 0.1 °C using a laboratory-made temperature controller.[25] The density of the samples was measured using a 10-ml pyknometer. The density measurements were accurate to $\pm 0.02\%$.

II. RESULTS AND DISCUSSION

A. Ultrasonic velocity

The experimental values of the ultrasonic velocity in, and the density of, the three ternary liquid mixtures at 30 °C, together with the ultrasonic velocities in the ternary liquid mixtures computed from free length theory (FLT),[4–6] collision factor theory (CFT),[7,8] Flory–Patterson theory (FPT),[1–3] Nomoto's relation (NR),[9] and ideal mixture relation (IMR)[10] at 30 °C are recorded in Table I. The necessary data of the pure components at 30 °C, required for the calculation of the parameters of the ternary mixtures, were taken from the literature[26,27] and are listed in Table II along with the calculated parameters of the pure components.

The experimental and theoretical velocity values versus mole fraction of MEK in the ternary mixtures at 30 °C are plotted in Figs. 1–3. A close look at the figures shows that the ultrasonic velocity computed from FLT and NR agrees very closely with the experimental values, while those computed using FPT, CFT, and IMR exhibit different degrees of deviation from the experimental values.

Several factors may be responsible for the discrepancy between the experimental values of ultrasonic velocity and the values predicted by FPT. Some of them may be associated with the assumptions made in the theory. The most important assumption is the one related to the size and shape of the molecules, which plays a vital role in predicting various physicochemical properties. The spherical nature of mol-

TABLE II. Parameters of pure components at 30 °C.

Parameter	Pure components				
	Methyl ethyl ketone	Toluene	Methanol	Ethanol	Propanol
V (cc)	90.19	107.13	40.85	58.97	75.95
ρ (g/cc)	0.7995	0.8601	0.7844	0.7812	0.7965
u (m/s)	1171	1292	1093	1123	1192
T_c (K)	535	593	513	516	536
$10^3\ \alpha(K^{-1})$	1.3072	1.0804	1.2168	1.1014	0.9903
C_P (J/mol)	165	156.84	81.54	114.71	150.73
V_0 (cc/mol)	70.18	86.46	31.25	45.22	58.78
$10^{-8}Y$ (cm^2/mol)	66.38	70.56	29.47	43.25	56.18
S	3.17	3.46	3.04	3.08	3.23
B (cc/mol)	20.80	25.00	9.17	13.42	17.38
V^* (cc/mol)	68.77	84.52	31.56	46.37	60.40
$\tilde{V}$	1.31	1.27	1.29	1.27	1.25
T^* (K)	4600	5056	4760	5006	5298
P^* (J/ml)	570	560	478	448	453

ecules has been assumed in defining the equations for segment fractions.[2] This is far from reality and leads to remarkable deviations in the predicted values of the ultrasonic velocity from the experimental values. The possibility of only two-body interactions has been assumed during the extension of the theory for a ternary liquid system[28] but there is every possibility of three-body interactions also, which has been ignored in the derivation of the theory for the sake of simplicity. Although three-body interactions contribute very little to the energy of the system, their possibility cannot be ignored.[28] The possibility of three-body collisions increases as the area of contact, i.e., chain length, increases.[28] Therefore, in order to get comparable results with experimental velocity, a correction term is needed which will take care of three-body effects in the evaluation of characteristic and interaction parameters in the Flory theory. This effect can be included by adding the interaction parameters χ_{123}, χ_{234}, χ_{341}, and χ_{412} for the three-body interactions in the expression for characteristic pressure of the mixture. The large deviation of ultrasonic velocity predicted by FPT from the experimental velocity might be due to the fact that the three-body interactions have not been taken care of in the FPT extended to ternary liquid mixtures.

The ultrasonic velocities predicted by IMR show a large deviation from the experimental velocities in the three ternary mixtures, but the deviation in velocity decreases as the chain of alcohols increases. This must be due to the assumptions made in the derivation of IMR for ternary mixtures. For the derivation of this relation, it has been assumed that the ratio of specific heats (γ_1, γ_2, and γ_3) of the components is equal to the ratio of specific heats of ideal mixture (γ_{im}), i.e., $\gamma_1=\gamma_2=\gamma_3=\gamma_{\text{im}}$.[10] This may not be true for the ternary mixtures in the present study, which might be partly responsible for the large deviation of velocity evaluated using IMR

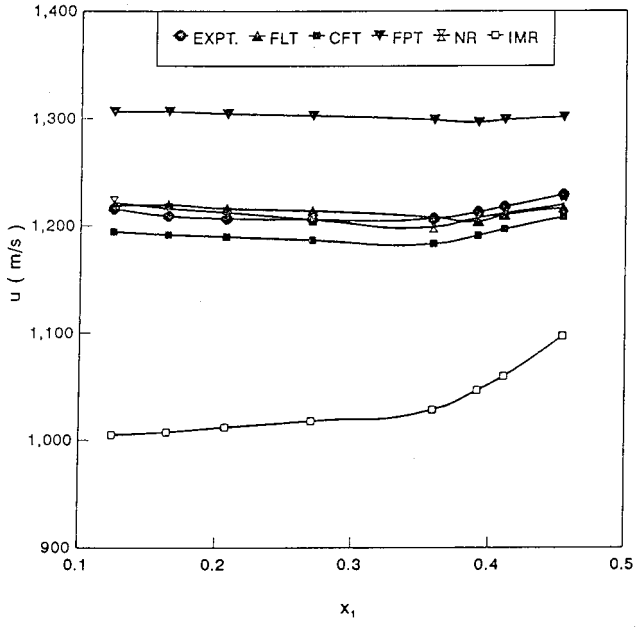


FIG. 1. Experimental and theoretical velocity versus mole fraction of MEK in the MEK+TL+ML ternary mixture at 30 °C.

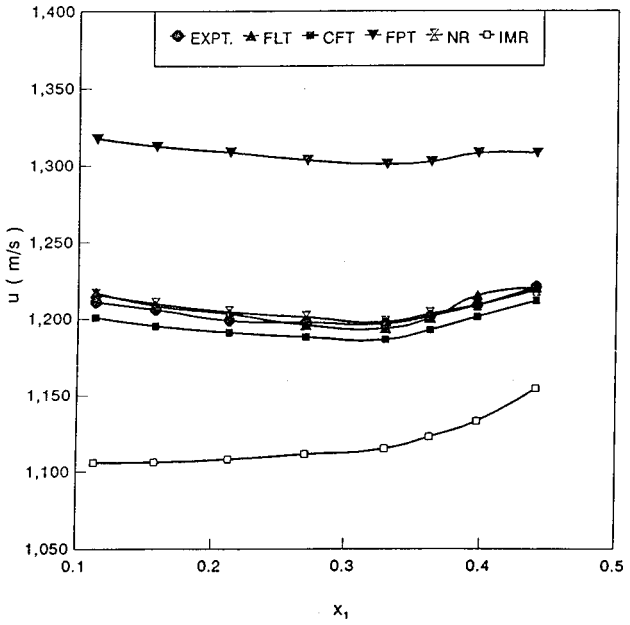


FIG. 2. Experimental and theoretical velocity versus mole fraction of MEK in the MEK+TL+EL ternary mixture at 30 °C.

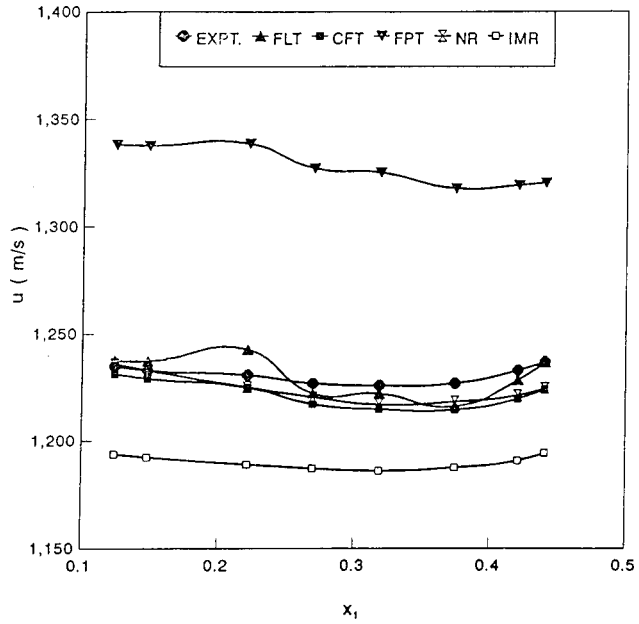


FIG. 3. Experimental and theoretical velocity versus mole fraction of MEK in the MEK+TL+PL ternary mixture at 30 °C.

from the experimental velocity. Another assumption used in the formulation of ideal mixture relation is that $V_1=V_2=V_3$, where V_1, V_2, and V_3 are the molar volumes of the components.[10] But in the ternary mixtures under present study, the molar volumes of alcohols are appreciably different from the molar volumes of the other two components, which are nearly equal (Table II). The large deviation of the theoretical velocities calculated using IMR form the experimental values in all the mixtures may be due to the large differences in molar volumes of alcohols compared to the molar volume of the other two components. The deviation in velocity calculated using IMR decreases with increase in chain length of alcohols. This may be attributed to the fact that the molar volume of alcohols increases and tends to approach the nearly equal molar volume of the other two components as the chain length of alcohols increases. This result strongly supports the reasoning given to the large deviation of theoretical velocity calculated using IMR from the experimental value.

The slight deviation in the value of ultrasonic velocity predicted by CFT from the experimental value can be ascribed to the fact that in collision factor theory, molecules are treated as real, nonelastic spheres.[7,8] The deviation in the ultrasonic velocity evaluated using IMR from the experimental value may be due to some sort of elasticity of the molecules.

B. Thermodynamic parameters and nonlinearity parameter (B/A)

The theoretically calculated values of the thermal expansivity (α), isobaric molar heat capacity (C_p), isochoric molar

TABLE III. Theoretically calculated values of thermal expansivity (α), isobaric heat capacity (C_P), isochoric heat capacity (C_V), ratio of specific heats (γ), isothermal compressibility (K_T) and acoustic nonlinearity parameter (B/A) of the ternary mixtures at 30 °C.

Mole fraction							
x_1	x_2	α (K^{-1})	C_p (J/mol·K)	C_v (J/mol·K)	γ	K_T (TPa^{-1})	B/A
Methyl ethyl ketone+toluene+methanol							
0.1240	0.4187	0.114E-02	123.29	95.20	1.295	1052.12	10.868
0.1642	0.3814	0.114E-02	123.82	96.08	1.289	1061.44	10.786
0.2068	0.3581	0.115E-02	125.67	97.45	1.290	1068.45	10.787
0.2700	0.3159	0.116E-02	127.80	99.15	1.289	1073.44	10.773
0.3588	0.2674	0.118E-02	131.65	101.82	1.293	1080.35	10.806
0.3912	0.3117	0.119E-02	137.83	105.47	1.307	1079.83	11.044
0.4101	0.3348	0.119E-02	141.14	107.75	1.310	1071.23	11.177
0.4537	0.3833	0.119E-02	148.47	112.48	1.320	1057.27	11.472
Methyl ethyl ketone+toluene+ethanol							
0.1129	0.4006	0.110E-02	137.26	108.35	1.267	1045.65	11.029
0.1570	0.3600	0.112E-02	137.84	108.75	1.268	1059.44	10.964
0.2129	0.3218	0.113E-02	139.09	109.77	1.267	1075.15	10.905
0.2707	0.2954	0.115E-02	140.98	110.75	1.273	1085.25	10.917
0.3294	0.2687	0.116E-02	142.85	111.95	1.276	1091.94	10.926
0.3622	0.2962	0.117E-02	145.66	113.52	1.283	1085.45	11.062
0.3971	0.3253	0.116E-02	148.54	115.46	1.286	1070.60	11.212
0.4416	0.3837	0.117E-02	153.29	117.57	1.304	1059.16	11.487
Methyl ethyl ketone+toluene+propanol							
0.1241	0.3985	0.106E-02	155.01	123.93	1.251	990.54	11.445
0.1484	0.3760	0.107E-02	155.21	124.25	1.249	993.80	11.408
0.2210	0.3215	0.107E-02	155.82	125.15	1.245	995.80	11.344
0.2692	0.2895	0.110E-02	156.51	124.72	1.255	1016.04	11.312
0.3185	0.2691	0.111E-02	157.08	124.92	1.257	1020.88	11.305
0.3745	0.2953	0.113E-02	158.15	124.09	1.274	1032.18	11.409
0.4208	0.3340	0.114E-02	158.94	123.76	1.284	1025.07	11.547
0.4407	0.3723	0.114E-02	159.39	123.40	1.292	1020.32	11.655

heat capacity (C_v), ratio of specific heats (γ), isothermal compressibility (K_T), and acoustic nonlinearity parameter (B/A) of the ternary mixtures at different temperatures are reported in Table III. The thermodynamical parameters of ternary mixtures are calculated using Flory's theory. Since the α, C_p, C_v, and K_T are very sensitive to temperature, pressure, and volume, the experimental determination of these parameters requires very accurate methods. In this context, the theoretical evaluation of these parameters is of considerable importance. The dependence of α, C_p, C_v, γ, and K_T on mole fraction can be analyzed from Table III. The values of these parameters of the mixtures vary regularly with change in mole fraction of alcohols. The parameters α, C_P, and K_T are used to calculate the nonlinearity parameter (B/A). The calculated values of B/A of ternary mixtures have the same order of the B/A values of the pure components.

III. CONCLUSION

Ultrasonic velocities predicted using FLT, CFT, FPT, NR, and IMR are compared with the experimentally measured velocity values at 30 °C in the three-component (ternary) mixtures of methyl ethyl ketone and toluene with *n*-alkanols. It is found that the velocities evaluated using all the theories except FPT and IMR are in fairly good agreement with the experimental velocity values. Acoustic nonlinearity parameter (B/A) and a few thermodynamic parameters have been evaluated theoretically for the tenrary mixtures.

[1] P. J. Flory, J. Am. Chem. Soc. **87**, 1833 (1965).
[2] A. Abe and P. J. Flory, J. Am. Chem. Soc. **87**, 1838 (1965).
[3] D. Patterson and A. K. Rastogi, J. Phys. Chem. **74**, 1069 (1970).
[4] B. Jacobson, Acta Chem. Scand. **6**, 1485 (1952).
[5] B. Jacobson, Acta Chem. Scand. **5**, 1214 (1951).
[6] B. Jacobson, J. Chem. Phys. **20**, 927 (1952).
[7] W. Schaaffs, Z. Phys. **114**, 110 (1939).
[8] W. Schaaffs, Z. Phys. **115**, 69 (1940).
[9] O. Nomoto, J. Phys. Soc. Jpn. **13**, 1528 (1958).
[10] W. Van Dael and E. Van Geel, Proceedings of the First International Conference on Calorimetry and Thermodynamics, Poland, 1969, p. 555.
[11] A. N. Kannappan and V. Rajendran, Acustica **75**, 192 (1991).
[12] J. D. Pandey, V. N. Srivastava, V. Vyas, and N. Panta, Indian J. Pure Appl. Phys. **25**, 1467 (1987).
[13] A. N. Kannappan, K. Ramalingam, and R. Palany, Indian J. Pure Appl. Phys. **29**, 43 (1991).
[14] J. D. Pandey, N. Pant, A. K. Shukla, Sarika, and V. Krishna, Indian J. Pure Appl. Phys. **27**, 246 (1989).
[15] W. N. Cobb, J. Acoust. Soc. Am. **73**, 1525 (1983).
[16] A. B. Coppens, R. T. Beyer, M. B. Reider, J. Donohue, F. Guepin, R. H. Hodson, and C. Townsend, J. Acoust. Soc. Am. **38**, 797 (1965).
[17] W. K. Law, L. A. Frizzell, and F. Dunn, Ultrasound Med. Biol. **11**, 307 (1985).
[18] A. P. Sarvazyan, T. V. Chalikian, and F. Dunn, J. Acoust. Soc. Am. **88**, 1555 (1990).
[19] Tong Jie and Dong Yanwu, Kexue Tongbao **33**, 1511 (1988).
[20] W. M. Madigosky, I. Rosenbaum, and R. Lucas, J. Acoust. Soc. Am. **69**, 1639 (1981).
[21] K. L. Narayana and K. M. Swamy, Acustica **47**, 51 (1980).
[22] B. Hartmann and E. Balizer, J. Acoust. Soc. Am. **82**, 614 (1987).
[23] C. M. Sehgal, B. R. Porter, and J. F. Greenleaf, J. Acoust. Soc. Am. **79**, 566 (1986).
[24] M. S. Khanwalkar, Acoust. Lett. **14**, 229 (1991).
[25] L. Godfrey and J. Philip, J. Phys. E **22**, 516 (1989).
[26] J. Timmermans, *Physicochemical Constants of Pure Organic Compounds* (Elsevier, London, 1950).
[27] *CRC Handbook of Chemistry and Physics*, 75th ed. (CRC, Boca Raton, 1994).
[28] J. D. Pandey, R. D. Rai, R. K. Shukla, A. K. Shukla, and N. Mishra, Indian J. Pure Appl. Phys. **31**, 84 (1993).

Modulation of torsional waves in a rod with a crack

A. E. Ekimov,[a] I. N. Didenkulov, and V. V. Kazakov
Institute of Applied Physics, Russian Academy of Sciences, Nizhny Novgorod 603600, Russia

(Received 12 September 1997; revised 1 February 1999; accepted 31 May 1999)

Results of experiments concerning nonlinear effects for elastic waves in metal rods with the crack-type defects are presented. The dependence of the nonlinear effect on the contact type is discussed. The crack has been modeled by cutting a rod and tightly filling the crack with metal plates. Two types of contact have been studied: a dry contact and a contact with lubricant. The modulation of high-frequency torsional waves (20 and 22.8 kHz) in the rod under the effect of low-frequency flexural vibrations has been investigated. Flexural vibrations have been excited by a shock and by a vibrator. The modulation of high-frequency torsional waves under the effect of low-frequency vibrations has been observed only in the rod with a crack. The level of modulation drastically decreases in the presence of liquid lubricant. © *1999 Acoustical Society of America.* [S0001-4966(99)03109-4]

PACS numbers: 43.25.Ed [MAB]

INTRODUCTION

Abnormally high acoustic nonlinearity exists in media with microstructures or defects (in the form of continuity disturbances).[1–7] In particular, strong nonlinear effects have been found in solids with cracks.[1,2,8–10] These effects include amplitude modulation of high-frequency longitudinal elastic waves by low-frequency vibrations,[3,4,8] generation of higher harmonics,[6,7,9,11] and combination scattering of two waves with different frequencies.[3,4] Nonlinear effects give a basis for developing different techniques of nondestructive testing (NDT).

It is usual to think that the physical nature of nonlinear effects in media with cracks is associated with the conditions of contact at points of continuity disturbances. Different models that describe abnormally high elastic nonlinearity of the cracks have been suggested.[5,12] One of the experiments, carried out at the Institute of Applied Physics (Nizhny Novgorod, Russia), has shown the effect of liquid filling of a crack in the lake ice on the decrease of the amplitude of ice nonlinear response to vibrational excitation.[10] The earlier experiments mainly dealt with longitudinal and flexural waves in solids.

The experiments, described in the present paper, have shown that a high degree of nonlinearity exists for torsional waves as well. The experiments have studied an influence of different contacts between touching surfaces of the crack on the strength of nonlinear response. The modulation of high-frequency torsional waves by low-frequency flexural vibrations in the metal rod with crack-type defects has been investigated. The crack has been especially modeled to study the influence of the contact type on modulation. Two types of contacts (a dry contact and a contact with a liquid lubricant) have been considered.

[a]At present on leave as visiting scientist at the Stevens Institute of Technology, Hoboken, NJ 07030.

I. EXPERIMENTS

Two cylindrical duralumin rods were used in the experiment. The length of the rods, *D*, was 91 cm, and their diameter, *d*, was 10 mm. The defects, modeling a crack, were made in one of the rods while the second one served as a reference rod without defects. To imitate the crack, a cut in the diameter plane (3.8-mm deep and 0.3-mm wide) was made at the middle of the rod. The cut was tightly filled in with three steel plates, each being 0.1 mm in thickness [see Fig. 1(a)].

Figure 1(b) and (c) give the schemes of experiments for

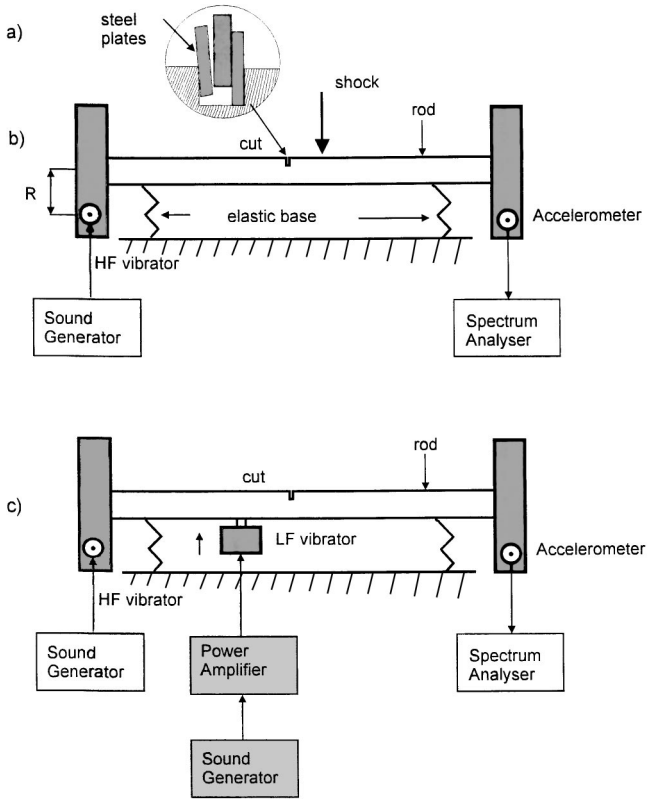


FIG. 1. A scheme of the experimental setup.

TABLE I. Eigenfrequencies of the rod flexural vibrations.

Type of rod	Flexural vibration mode			
	1	2	3	4
Rod with a cut	26 Hz	101 Hz	226 Hz	399 Hz
Rod without a cut	26 Hz	100 Hz	225 Hz	397 Hz

studying the interaction of high-frequency torsional waves with low-frequency flexural vibrations in the rods. The ends of the rods freely rested upon an elastic base. Foam rubber was used for this base. It is important to mention that the resonance frequency of the elastic base is much lower than the lowest eigenfrequency of the flexural vibrations of the rod. A piezoceramic source [high-frequency (HF) vibrator] of vibrations was used to excite high-frequency torsional waves. Transverse momentum, applied at the distance $R = 30$ mm from the rod axis, was responsible for the excitation. A piezoaccelerometer was fixed at the rod end, opposite to the excitation end, and was oriented to register torsional vibroaccelerations. In this experiment the level of the torsional accelerations was about 0.1 m/s^2.

Low-frequency flexural vibrations of the rods were excited in two different ways:

(a) by a shock (by a rubber coated hammer) normal to the rod surface near its middle [Fig. 1(b)];
(b) by an electrodynamical vibrator [190 g low-frequency (LF) vibrator] normal to the rod surface, hung near the rod center [Fig. 1(c)]. At this point, the level of acceleration produced by the LF vibrator was about 1 m/s^2.

Vibrational signals were registered by a light accelerometer (2 g in mass) fixed at the rod end. Vibrational signals were registered and analyzed in the frequency range 0–25.6 kHz by a spectrum analyzer (model B&K 2034).

Each way of excitation is characterized by its own advantages and drawbacks. The key advantage of the shock excitation is the absence of nonlinear contacts in the source of excitation, while its drawback is related to the impossibility to sustain the same shock force for all runs of the experiment. The second type of excitation (by the LF vibrator) is free from this limitation, because the force acting on the rod from the LF vibrator can be maintained constant during all

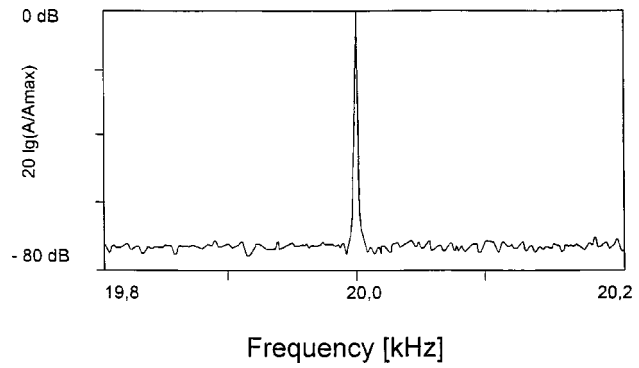


FIG. 2. SVR of two rods: the rod with a cut and the rod with a cut, but without plates in the cut. Torsional waves are excited by the HF vibrator at the frequency 20.0 kHz. Flexural waves are excited by shock.

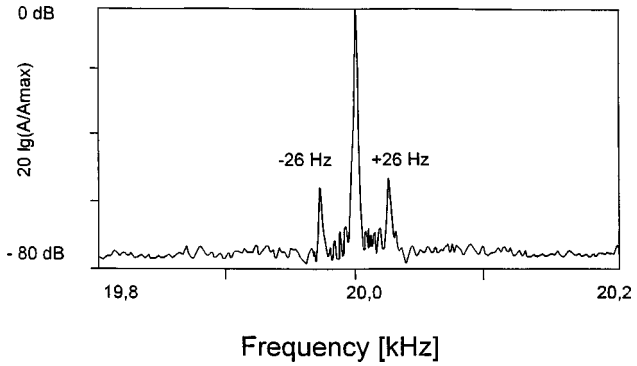


FIG. 3. SVR of a rod with a cut filled with plates. Torsional waves are excited by the HF vibrator at the frequency 20.0 kHz. Flexural waves are excited by shock.

the measurements by feeding the LF vibrator by a controlled electrical voltage (1 V_{rms}) from a power amplifier. The drawbacks of this method are associated with nonlinear phenomena, which can be present in the unit fixing the LF vibrator on the rod and in the LF vibrator itself. These effects are difficult to eliminate completely.

Table I gives four eigenfrequencies of the rod flexural vibrations measured experimentally. The analyzed frequency band (0–25.6 kHz) covers more than 10 eigenfrequencies of torsional rod vibrations.

A typical spectrum of vibrational response (SVR) for the rod without a cut and the rod with a cut but without plates in the cut is plotted in Fig. 2. SVR was measured for simultaneously excited high-frequency harmonic torsional waves ($\omega = 20$ kHz) by the HF vibrator and low-frequency flexural vibrations of the rod induced by the shock. The *y*-axis corresponds to the normalized amplitude of vibroaccelerations in the logarithmic scale: $20 \log A/A_{max}$, where A_{max} is the maximum acceleration at the frequency of torsional oscillations. The *x*-axis corresponds to the frequency band from 19.8 to 20.2 kHz. One can see no modulation in Fig. 2. In this case, the amplitude of flexural vibrations was small and the cut walls did not touch one another, so the system was linear. When cut walls touch one another, the system becomes nonlinear for elastic waves in the rod. This is the case of contact nonlinearity.[13–15]

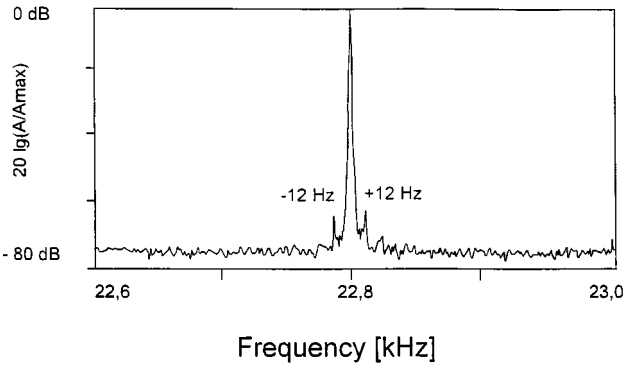


FIG. 4. SVR of a rod with a cut without plates. Torsional waves are excited by the HF vibrator at the frequency 22.8 kHz. Flexural waves are excited by an electrodynamical LF vibrator at the fundamental frequency 12 Hz.

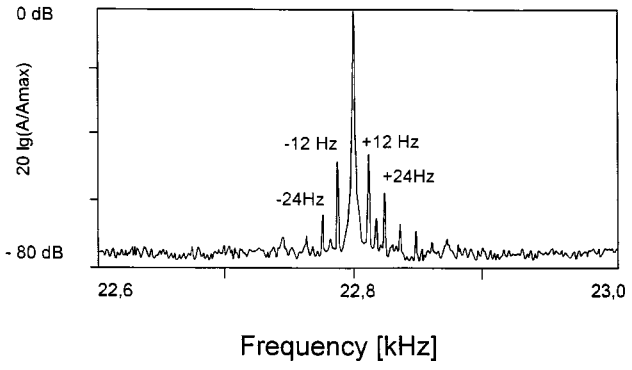


FIG. 5. SVR of a rod with a cut filled with plates. Torsional waves are excited by the HF vibrator at the frequency 22.8 kHz. Flexural waves are excited by an electrodynamical LF vibrator at the fundamental frequency 12 Hz.

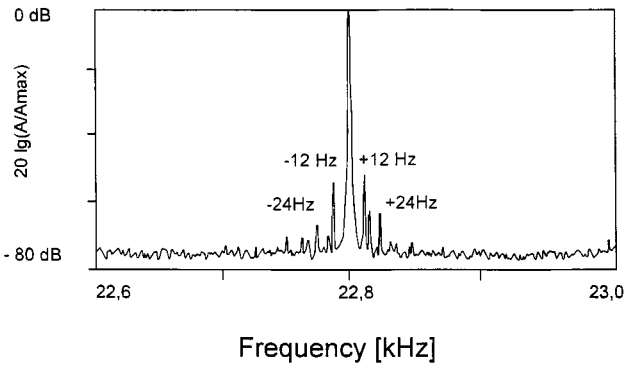


FIG. 6. SVR of a rod with a cut filled with plates lubricated by engine oil. Torsional waves are excited by the HF vibrator at the frequency 22.8 kHz. Flexural waves are excited by an electrodynamical LF vibrator at the fundamental frequency 12 Hz.

Figure 3 gives the SVR of the rod with the cut filled with rigid plates. The excitation conditions were the same as described above. It is evident from Fig. 3 that signals with frequencies $\omega, \omega \pm \Omega$ (where Ω is the frequency of flexural vibrations) are registered in the response. This means that the modulation of high-frequency torsional oscillations by flexural vibrations of the rod at one of the eigenfrequencies takes place. One can conclude that the system has become nonlinear. This phenomenon is associated with the interaction of plates inserted into the cut.

The plate interaction in the cut also was studied. To this end, the plates, inserted into the cut, were lubricated by a small amount of engine oil. Flexural oscillations were excited by the LF vibrator acting on the rod with a constant force. The maximum rod vibrations were registered at the frequency 12 Hz (vertical displacement near the cut being 0.05 mm). There were also the 2nd and 3rd harmonics of the fundamental frequency (12 Hz) in the vibration spectrum due to the electrodynamical LF vibrator. Torsional waves were excited at the frequency 22.8 kHz by the HF vibrator. The results of the measurements are presented in Figs. 4–6. Figure 4 shows the SVR of the rod with the cut but without plates. One can see that there are frequencies $\omega, \omega \pm \Omega$ (where Ω is the frequency of induced flexural vibrations) in the vibration spectrum. This may be associated with nonlinear contact phenomena in the suspension of the LF vibrator at the point where it is attached to the rod, because the effect was not observed in the case when flexural waves in the rod were excited by shock. Figures 5 and 6 give the SVR of the rod with the cut filled with plates without and with lubricant, respectively. Table II gives amplitudes $[V_i = 20 \log(A_i/A_{\max})]$ of the modulation components in the spectrum of response, where the index i ($i=1,2,3$) corresponds to the type of rod. It is evident from Figs. 5–6 and Table II that lubrication of the cut with engine oil decreases the modulation by 5–6 dB.

II. CONCLUSIONS

The results obtained in the experiments point to the existence of modulation of high-frequency torsional waves by low-frequency flexural vibrations in rods with crack-type defects. On the other hand, there is no modulation in the rods with the cut and without the cut. This fact suggests a possibility of using torsional waves for nondestructive nonlinear acoustic testing.

A physical nature of nonlinearity is associated with contact phenomena in the crack. If the walls of the crack do not touch each other during vibrations, nonlinear interaction of waves is not observed.

The value of the modulation also depends of the contact type. When liquid lubricant is used, the modulation decreases by 5–6 dB. This demonstrates a possibility to change the nonlinearity of crack-type defects by filling cracks with liquid.

TABLE II. Amplitudes of the modulation components.

		Frequency (Hz)				
		22 800	22 800	22 800	22 800	
i	Type of rod	+12	+24	−12	−24	V
1	Rod with a cut	−63 dB	⋯	−65 dB	⋯	V_1
2	Rod with a cut and plates	−46 dB	−57 dB	−48 dB	−60 dB	V_2
3	Rod with a cut with inserted plates lubricated with engine oil	−51 dB	−63 dB	−53 dB	−65 dB	V_3
		5 dB	6 dB	5 dB	5 dB	V_2-V_3
		17 dB	⋯	17 dB	⋯	V_2-V_1

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation of Fundamental Research (Grant 97-02-17524) and INCAS, Nizhny Novgorod, Russia. The authors are very grateful to Alexander Sutin (Institute of Applied Physics, Russia) and Paul Johnson (Los Alamos National Laboratory) for fruitful discussions.

[1] V. E. Nazarov, L. A. Ostrovsky, I. A. Soustova, and A. M. Sutin, ''Nonlinear acoustics of microinhomogeneous media,'' Phys. Earth Planet. Inter. **50**, 65–73 (1988).

[2] A. M. Sutin, ''Nonlinear acoustic diagnostics of micro-inhomogeneous media,'' in *Proceedings of the International Symposium on Recent Advances in Surveillance Using Acoustical and Vibrational Methods* (Senlis, France, 1992), pp. 497–506.

[3] V. Yu. Zaitsev, A. M. Sutin, I. Yu. Belyaeva, and V. E. Nazarov, ''Nonlinear interaction of acoustic waves due to cracks and its possible usage for cracks detection,'' J. Vib. Control **1**, 335–344 (1995).

[4] I. Yu. Belyaeva and V. Yu. Zaitsev, ''Nonlinear scattering of acoustic waves by discontinuity-like defects in application to crack detection,'' J. Vib. Control **2**, 465–478 (1996).

[5] V. E. Nazarov, ''Elastic nonlinearity of the solids with partially liquid-filled cracks,'' *Preprint of the Institute of Applied Physics*, Vol. 428 (Nizhny Novgorod, Russia, 1997), p. 10.

[6] G. D. Meegan, P. A. Johnson, R. A. Guer, and K. R. McCall, ''Observations of nonlinear elastic wave behaviour in sandstone,'' J. Acoust. Soc. Am. **94**, 3387–3391 (1993).

[7] P. A. Johnson, B. Zinszner, and P. N. J. Rafolofosaon, ''Resonance and nonlinear elastic phenomena in rock,'' J. Geophys. Res. **101**(B5), 11553–11564 (1996).

[8] A. S. Korotkov and A. M. Sutin, ''Modulation of ultrasound by vibrations in metal constructions with cracks,'' Acoust. Lett. **18**, 59–62 (1994).

[9] A. M. Sutin, A. S. Korotkov, I. N. Didenkulov, E. J. Kim, and S. W. Yoon, ''Nonlinear acoustic methods for crack and fatigue detection,'' in *Proceedings of Physical Acoustics Workshop ''Safety Diagnostics of Underwater Constructions by Using Acoustics''* (KAIST, Seoul Branch, Seoul, Korea, 1995), pp. 43–55.

[10] A. E. Ekimov, A. V. Lebedev, L. A. Ostrovskii, and A. M. Sutin, ''Nonlinear acoustic effects due to cracks in ice cover,'' Sov. Phys. Acoust. **42**, 61–64 (1996).

[11] V. A. Antonets, D. M. Donskoy, and A. M. Sutin, ''Nonlinear vibrodiagnostics of delayer in layer construction,'' J. Mech. Composit. Materials (Russia) **5**, 934–937 (1986).

[12] A. M. Sutin and V. E. Nazarov, ''Nonlinear acoustic methods of crack diagnostics,'' Izvestiya Vysshikh Uchebnykh Zavedenii, Radiofizika **38** (3–4), 169–187 (1995). [Radiophys. Quantum Electron. **38**(3–4), 109–120 (1995)].

[13] M. Richardson, ''Harmonic Generation at an Unbonded Interface. I. Planar interface between semiinfinite elastic media,'' Int. J. Eng. Sci. **17**, 73–75 (1979).

[14] O. Rudenko and Chin An Vu, ''Nonlinear acoustic properties of a rough surface contact and acousto diagnostics of a roughness height distribution,'' Acoust. Phys. **40**(4), 593–596 (1994).

[15] A. M. Sutin and D. M. Donskoy, ''Vibro-acoustic modulation nondestructive evaluation technique,'' in *Proceedings of the International Society for Optical Engineering*, ''Nondestructive Evaluation of Aging Aircraft, Airports, and Aerospace Hardware II'' (San Antonio, TX), **3397**, 226–237 (1998).

Full-wave simulation of the forward scattering of sound in a structured ocean: A comparison with observations

Michael A. Wolfson[a)]
Earth System Science Center, Pennsylvania State University, University Park, Pennsylvania 16802

John L. Spiesberger[b)]
Department of Meteorology, Pennsylvania State University, University Park, Pennsylvania 16802

(Received 20 October 1997; accepted for publication 29 April 1999)

Between 1983 and 1989, acoustic pulse-like signals at 133-Hz, 60-ms resolution, were transmitted from Oahu to Northern California. Analysis of the data indicates that the early arriving, steep paths are stable over basin scales, whereas the late, near-axial paths are sensitive to ocean structure. The late paths undergo vertical scattering on the order of the acoustic waveguide, i.e., 1 km [J. Acoust. Soc. Am. **99**, 173–184 (1996)]. The parabolic approximation is used to simulate pulse propagation over the vertical plane connecting the source and receiver. Several prescriptions are used for the speed of sound: (1) Climatologically averaged sound speed with and without a realization of internal waves superposed; (2) Measured mesoscale structure with and without a realization of internal waves superposed. The spectrum of the internal waves is given by Garrett and Munk. Modeled internal waves and the measured mesoscale structure are sufficient to explain the vertical scattering of sound by 1 km. The mesoscale structure contributes a travel time bias of 0.6 s for the late multipath. This bias is seen to be a relevant contribution in accounting for the travel times of the last arrival. [S0001-4966(99)03108-2]

PACS numbers: 43.30.Dr, 43.30.Rc [DLB]

INTRODUCTION

In a recent paper by Spiesberger and Tappert,[1] geometric acoustics was utilized to interpret long range receptions of acoustic signals transmitted from the Kaneohe source.[2] The experiment involved propagation of 133-Hz, 60-ms resolution pseudo-random signals over a distance of 3709.21 km. Geometric acoustics applied to an environment that included mesoscale structure was able to adequately model the early arriving, resolved stable acoustic multipaths, which were derived by incoherently averaging over the receptions that occurred each day. However, the coda of about one second could not be accounted for with ray or full-wave theory. They hypothesized that forward scattering due to internal gravity waves in addition to mesoscale structure is responsible for vertically scattering the acoustic energy a distance of 1 km from the axis of the sound channel. Recent studies by Colosi and Flatté[3] and Colosi[4] of wave propagation through internal wave fields provide additional credence to this hypothesis, but our studies reveal that the effect of ocean mesoscale structure is important as well. To test the hypothesis, we resort to full-wave modeling, using a parabolic approximation to the one-way Helmholtz wave equation (PE). Numerical simulations based on this model support the hypothesis that acoustic energy interacting with internal waves, mesoscale structure, and the bottom are responsible for vertically scattering the late arriving axial energy by 1 km.

Note that PE simulations at multi-megameter ranges, through ocean environments which include internal waves, have been performed and recently reported in the literature by Colosi and Flatté,[3] and Colosi.[4] Our efforts using these type of simulations differ in that we compare the results to particular observations, more in line with what has been done by Colosi *et al.* at the shorter range of one megameter.[5]

We proceed by the following prescription. Section I describes the Kaneohe experiment. In Sec. II the method used to generate the sound speed fields is described. Much of this section involves describing the method used to model the range dependent sound speed perturbations due to internal waves, although it is not terribly different from what Colosi *et al.*[5] have done. Section III describes the parabolic equation model. Near the source and receiver, sub-bottom interaction is significant, controlling convergence characteristics of the parabolic equation model over other environmental constituents, such as internal waves. To optimize code performance while still satisfying convergence, a variable depth and range step size is used. In Sec. IV the model results are presented and compared with observed impulse response data. Comparisons are made using four different ocean environments: climatological average and measured mesoscale structure, each with and without a realization of an internal wave field superposed. The primary objective being established, Sec. V describes a method for measuring the extent of vertical scattering as a function of range. Finally, Sec. VI discusses the relevance of our findings in regard to tomographic measurements.

Before proceeding to the main development of the paper, let us briefly outline the problem setting. The ocean to first order acts as a waveguide in depth for acoustic pulse transmissions in deep ocean mid-latitude environments. Inhomogeneities in the sound speed field due to oceanic pro-

[a)]Present address: Department of Physics, Washington State University, Pullman, WA 99164-2814.
[b)]Also at: The Applied Research Laboratory, Pennsylvania State University, University Park, PA 16802. Present address: Department of Earth and Environmental Science, University of Pennsylvania, Philadelphia, PA 19104.

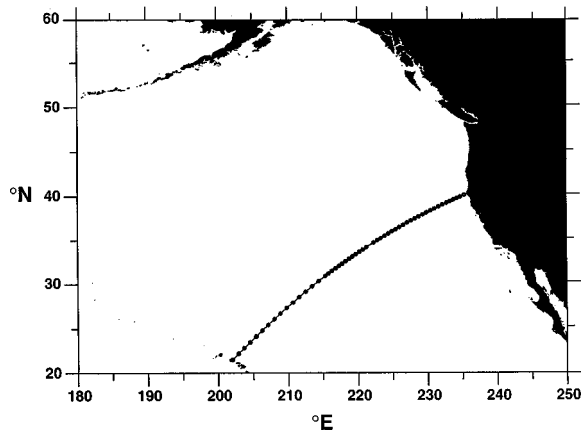


FIG. 1. Plan view of the Kaneohe source experiment. The source is located about five miles north of Kaneohe Bay, Oahu. The receiver near the coast of Northern California, is one of many U.S. Navy SOSUS stations used to receive these transmissions. Circles indicate the positions of 56 stations where conductivity, temperature, and depth were measured with a CTD in July 1988 by the Naval Oceanographic Office. Adapted from Ref. 1.

cesses such as mesoscale energetics and internal waves have scales that are large compared to the acoustic wavelength. They are weak in the sense that the deviation in the rms scattering angle due to interaction with a ''scatterer'' relative to that which would have taken place without the scatterer present is small compared to unity. It is in this sense that we refer to the problem as one of *multiple forward scattering*.

I. KANEOHE EXPERIMENT

The Kaneohe experiment consists of an acoustic source, bottom-mounted off the coast of Kaneohe Bay, Oahu, and a set of U.S. Navy Sound Surveillance Systems (SOSUS) spread out in the eastern North Pacific. The source, located at 21.512 35 °N, 202.228 48 °E, at a depth of 183 m (one meter above the ocean floor), transmitted acoustic pulse-like signals at 133-Hz, 60-ms resolution intermittently over the years 1983 to 1989. The receiver of interest in this study is located off the coast of Northern California (on the ocean floor), at 40.0786 °N, 234.8880 °E at a depth of 1433 m. Geographical locations are given in WGS-84.[6] Conductivity, temperature, and depth were recorded using a CTD to depths of about 2000 m at 56 stations along the geodesic (Fig. 1). Actual CTD depths are shown in Fig. 2. Travel times were measured using clock accuracies of 1 ms. The bathymetry along the geodesic is illustrated in Fig. 3. Further details of the experiment are found elsewhere.[1,2]

Interannual changes in travel time were measured to be about ± 0.2 s. Rossby waves linked to ENSO are the dominant mechanism affecting the observed changes in travel time throughout the six-year experiment.[7] Most of the results of the analyses here do not depend on this change in travel time because the primary interpretation is based on the difference in travel time between the stable multipath and the late arriving energy.

To mitigate the stochastic effects on the received signal due to internal waves and increase the signal-to-noise ratio, daily incoherent averaging was performed. Anywhere from 70 to about 700 pulse-like signals per day were used to perform the average, and the representative data we chose are modeled results (see Fig. 4) using about 700 pulse-like signals transmitted and received on 29 November, 1983. Further details regarding the processing of the data can be found in Ref. 1.

II. MODELS OF OCEAN SOUND SPEED

Four distinct sound speed fields are used for the simulations presented in this paper, and they are designated as: (a) climatology; (b) mesoscale; (c) internal waves; (d) mesoscale and internal waves.

The climatology sound speed field is derived from a historical database of salinity and temperature, and is representative of the average profile which is horizontally smooth. The primary variation seen in this sound speed field is a shoaling of the sound channel axis from the source to receiver (Fig. 4 of Ref. 2). The mesoscale sound speed field contains deterministic mesoscale structure derived from 56 CTDs taken along the source–receiver geodesic in July 1988. We use the word mesoscale to denote sound speed inhomogeneities whose horizontal scales range from 25 km to 500 km, and which evolve on time scales of one month. The sound speed field denoted as ''internal waves'' represents sound speed fluctuations from internal waves superposed onto the climatology sound speed field. The ''mesoscale and internal waves'' sound speed field represents the mesoscale sound speed field with fluctuations due to internal waves superposed.

To describe the ocean sound speed fields, we first define a local Cartesian coordinate system for the vertical plane connecting the source and receiver along the geodesic. Let x measure the horizontal range in kilometers from the source, and z measure the depth in meters positive downward from the ocean surface. The ocean surface is taken to be flat. The sound speed fields (a) through (d) are computed on a discrete grid (x_i, z_j), $i=1,..,i_{\max}$, $j=1,..,j_{\max}$, which is determined by the smallest internal wave scales.

A. Climatology

To compute the climatological sound speed field, salinity and temperature data are extracted from the Levitus database.[8] Bilinear interpolation is then used to compute the temperature and salinity profiles at the range points x_i, $i=1,..,i_{\max}$, at the standard Levitus depths. The range points x_i are determined by $x_i=(i-1)\Delta x$, with Δx chosen to be about 312.5 m. This sets the shortest horizontal internal wave scale to be about 625 m. Del Grosso's formula[9] is then used to compute the sound speed at Levitus' standard depths. Finally, a quadratic spline is used to interpolate sound speeds onto discrete depths z_j. The discrete depths used are every 5 m from the surface down to 500 m, every 20 m from 500 m down to 3020 m, and every 50 m from 3020 m to the bottom of the water column. This interpolation algorithm avoids discontinuities in the sound speed gradient at Levitus' standard depths.[10]

B. Mesoscale

The mesoscale structure is interpolated onto the same (x_i, z_j) grid in a similar manner (Fig. 2). The CTD data are

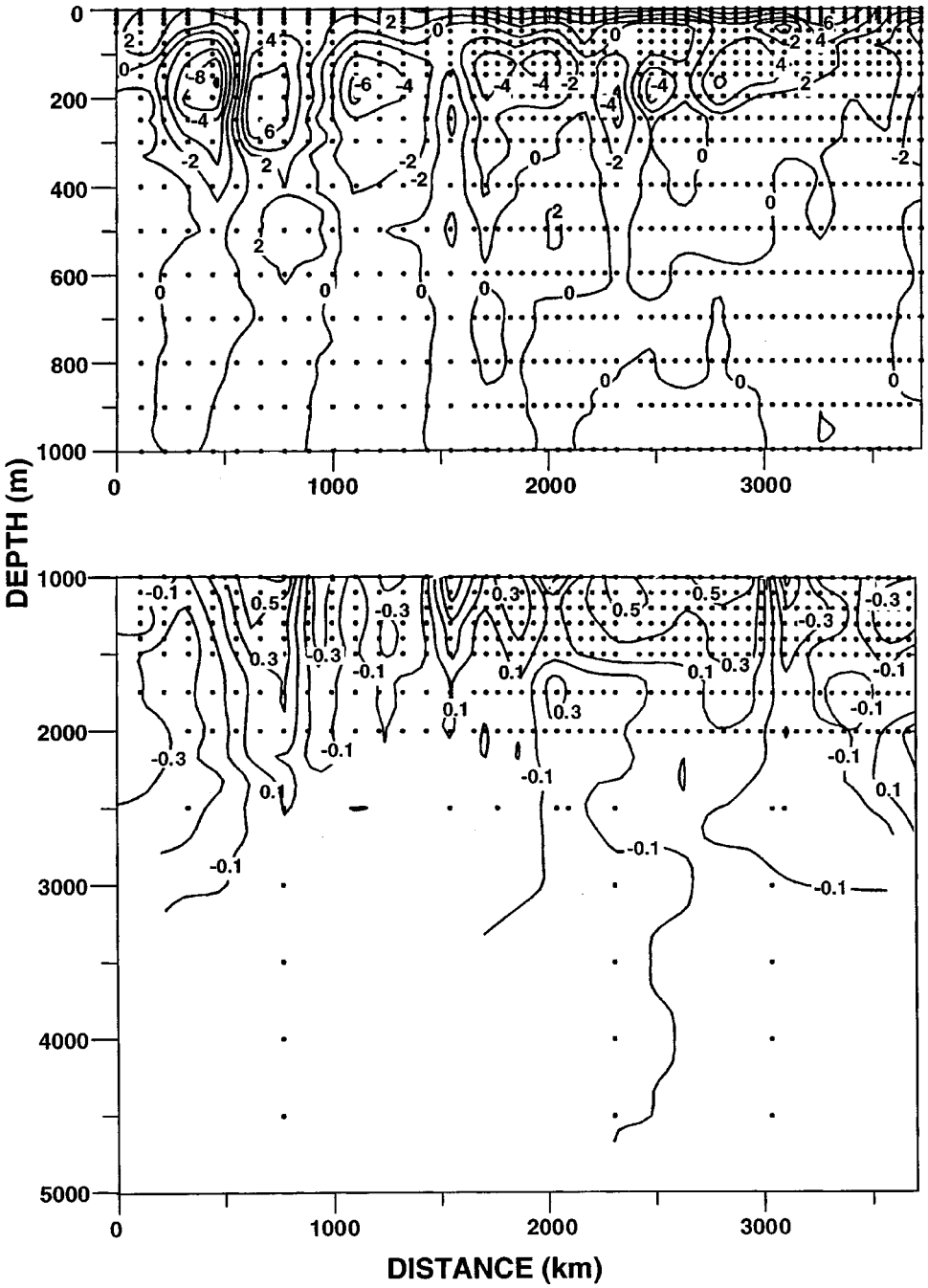


FIG. 2. Sound speeds computed from the July 1988 CTD cruise minus sound speeds computed from Levitus' climatology for the months May, June, and July. Negative values indicate the CTD section is colder and slower than Levitus' data. Contour intervals are 2 m/s and 0.2 m/s in the upper and lower panels, respectively. The dots indicate where the data were taken from the CTD. Adapted from Ref. 1.

linearly interpolated in range onto the points x_i. The same quadratic spline algorithm is then used to interpolate onto the depths z_j. From the source out to 1700 km, the horizontal resolution is about 225 km, and beyond this range to the receiver the horizontal resolution is about 112 km.

C. Internal waves

As mentioned above, the ''internal waves'' and ''mesoscale and internal waves'' sound speed fields involve superposing the sound speed fluctuations due to internal waves onto either the climatology or mesoscale sound speed field. Thus it is only necessary to describe how the sound speed fluctuations due to internal waves are modeled.

The method used to model internal waves is similar to what Colosi *et al.*[5] have done. Differences will be pointed out as the description is elucidated. The basic premise is to partition the gridded vertical plane (x_i, z_j) into M overlapping horizontal ''sections.'' The size of each section is about 80 km, and the extent of overlap is 10 km. Within each section, the ocean is nearly horizontally stratified, so that the buoyancy frequency $N(z_j)$, inertial frequency ω_i, and bottom depth z_{bot} are all range independent. Bathymetry was measured to an accuracy of within 2% of the total depth from the source out to about 130 km, and measured to within 50 m over the last 100 km along the geodesic. Bathymetry was extracted at 9-km intervals from ETOPO5[11] between these two regions. The actual bottom depth used for each section is range averaged from the measured and extracted bathymetry.

Internal waves which are based on the empirical spectrum of Garrett and Munk[12,13] are computed for each section.

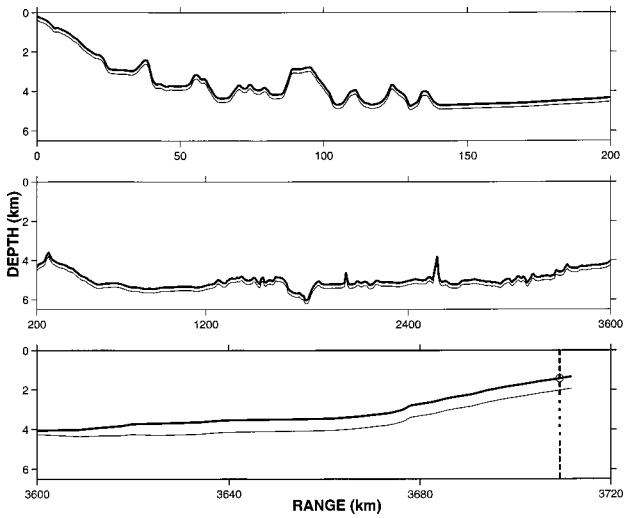


FIG. 3. Bathymetry (thick solid) and sediment-basement interface (thin solid) along Kaneohe geodesic for three different simulation domains: (1) near source bottom-limited, (2) deep ocean waveguide, (3) near receiver bottom-limited. The bathymetry was filtered using a 1-km horizontal length scale, and was extracted from the ETOPO5 database (Ref. 11).

Fully three-dimensional internal wave fields in a local Cartesian coordinate system are constructed, but only the wave field along the section is saved. The sections are then overlapped and smoothly patched together to form the total field of internal waves. The choice of using a three-dimensional internal wave field representation instead of a two-dimensional representation came from our inability to justify an approximation used by Dozier and Tappert.[14] The approximation involves the internal wave eigenfunctions that depend on the magnitude of horizontal wave number. They approximate these eigenfunctions by the eigenfunctions that depend on the component of horizontal wave number parallel to the section. When both components of the horizontal wave number are of the same order, and they are each small enough that the spectral weighting is significant, this approximation breaks down.

The method used to generate each section of internal waves, and the method by which sections are patched together is described in Appendix A. If $\delta c(x_i, z_j)$ denotes the vertical plane of gridded sound speed fluctuations due to internal waves, the total sound speed fields are expressed as

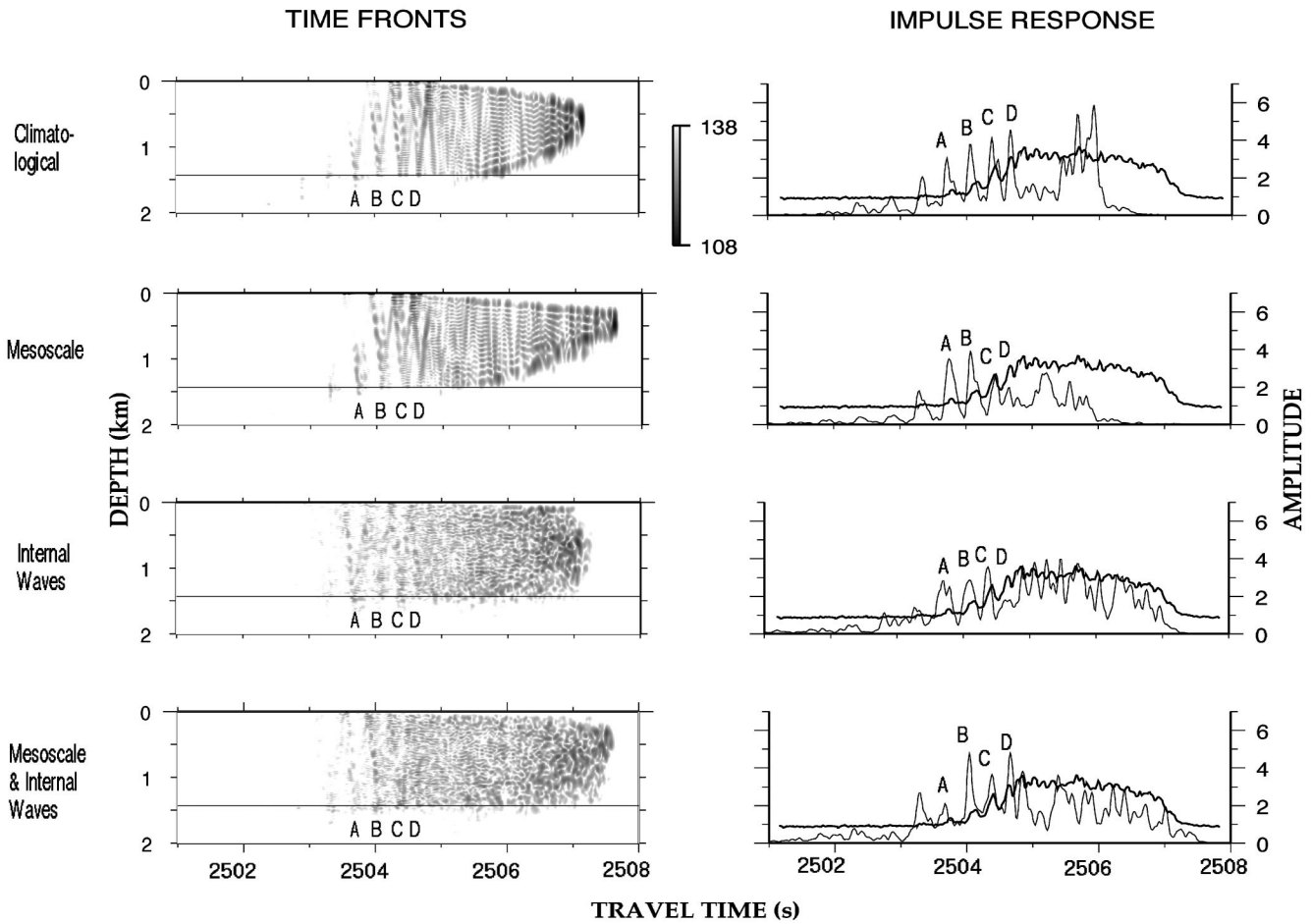


FIG. 4. Composite wavefronts (gray scale with 30-dB dynamic range) and impulse response functions at the receiver range for the different modeling environments used. The computed impulse response functions (thin solid curves—right column) are plotted on an arbitrary scale. The Kaneohe data (solid lines), taken 29 November, 1983, have been incoherently averaged. The importance of scattering due to ocean structure is clearly visible in the lower two panels. The amplitudes of the data were scaled by comparison to the amplitudes of the modeled output arrivals A–D computed by propagation through the climatological sound speed field. Finally, the scaled data were translated along the time axis to give the optimal fit to the arrivals A–D. Note the horizontal line in the left panels is at the hydrophone depth for the impulse response functions shown in the right panels.

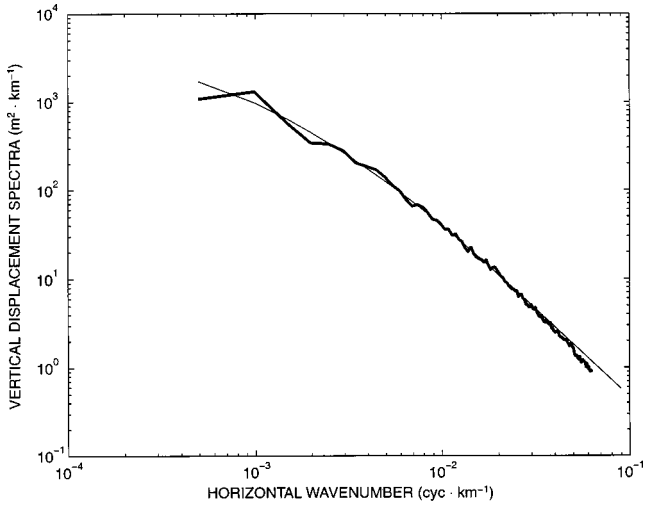


FIG. 5. Power spectral density of vertical displacement of internal waves as a function of horizontal wave number. The thin solid curve is the theoretical estimate given by $\Sigma_{j=1}^{50} F_\zeta(k,j)$ from Eq. (A9), and the thick solid curve is the spectrum calculated from a single realization of a two-dimensional horizontal field 80×80 km in extent, centered at 31.1258 °N, 215.7255 °E, or about 1715 km from the Kaneohe source along the section in Fig. 1.

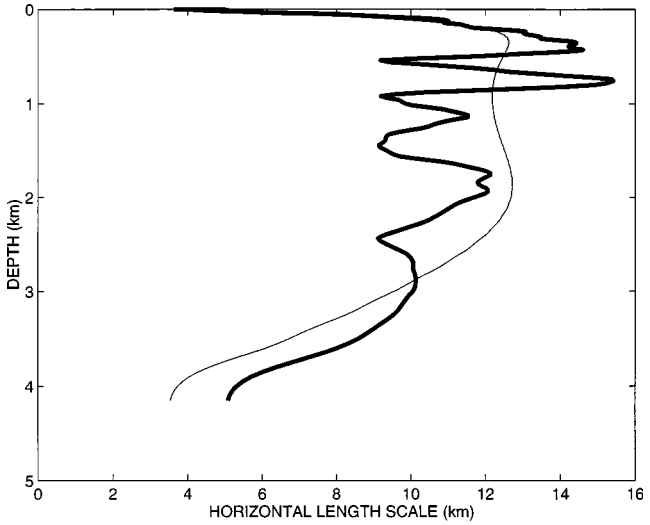


FIG. 7. Horizontal correlation length scale. The thick line is constructed using a single three-dimensional 80×80 km realization of a vertical displacement field, centered at 31.1258 °N and 215.7255 °E, using a discretized form of Eq. (B12) for $\alpha=0$ and $\pi/2$ radians, then taking the average of the results from these two angles. The theoretical expression, Eq. (B13), is used to produce the thin solid line.

$$c(x_i,z_j)=\overline{c}(x_i,z_j)+\delta c(x_i,z_j), \tag{1}$$

where $\overline{c}(x_i,z_j)$ is either the climatological or mesoscale sound speed field.

Various tests assist in validating our internal wave model. Figure 5 illustrates good agreement between simulation and theory [Eq. (A9)] for the power spectral density of vertical displacement as a function of horizontal wave number at a depth of 490 m. The standard deviation of vertical displacement, computed at all depths from a single realization, follows the expected theoretical values [in accordance with Eq. (B7)], except near the surface and bottom (Fig. 6). Because the WKB approximation inherent in the GM spectrum is only applicable in the interior of the turning depths for the internal displacement modes, agreement should not be expected above and below the turning depths. For the realization used, most of the modes have turning depths above 3000 m, and below this depth the zero displacement boundary condition is seen to dominate. The same explanation applies near the surface, except that the upper turning depth is so close to the surface on the depth scale shown that the zero displacement boundary condition and turning depth dependence nearly coincide.

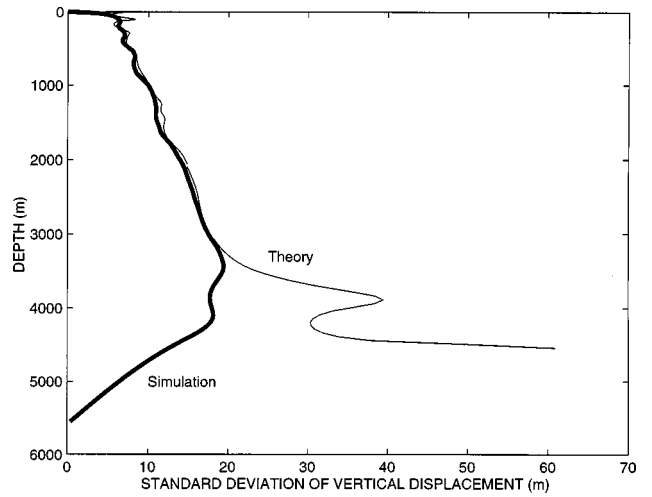


FIG. 6. Standard deviation of the internal wave's vertical displacement as a function of depth. The thin curve is $\langle\zeta^2\rangle^{1/2}=[\int dk\Sigma_j F_\zeta(k,j)]^{1/2}=[b^2E_0N_0/(2N(z))]^{1/2}$ from Eq. (B7). The thick curve is $\langle\zeta^2\rangle^{1/2}$, calculated from a single three-dimensional realization of a vertical displacement field, centered at 31.1258 °N, 215.7255 °E.

The horizontal correlation length scale is also computed, and a formula for this scale is derived in Appendix B 2. Figure 7 illustrates a reasonable fit between this formula and a single three-dimensional realization of an internal wave field. The high amount of variance shown in the simulation is attributed to the fact that the horizontal extent of the realization is only 80 km, less than eight times the internal wave length scale of $O(10)$ km. Smaller variance is expected for larger section sizes, but 80 km appeared to be the appropriate size; small enough to satisfy the slowly varying assumption for the range dependence of $\overline{c}(x_i,z_j)$ required by our model, and large enough to simulate acoustically relevant horizontal scales of internal gravity waves.[15] The qualitative aspects of a single section of sound speed fluctuations are seen in Fig. 8.

This completes the description of how the ocean sound speed fields (a) through (d) are computed. The next section describes the model which simulates acoustic propagation through these four environments.

III. FULL-WAVE MODELING

For simulating two-dimensional full-wave acoustic propagation over long ranges (>1000 km) in weakly range dependent waveguides, models based on ''the parabolic equation method'' (PE) are often desirable.[16,17] Note that by ''weakly'' the problem is assumed to be describable by the physics of multiple forward scattering, as stated in the Intro-

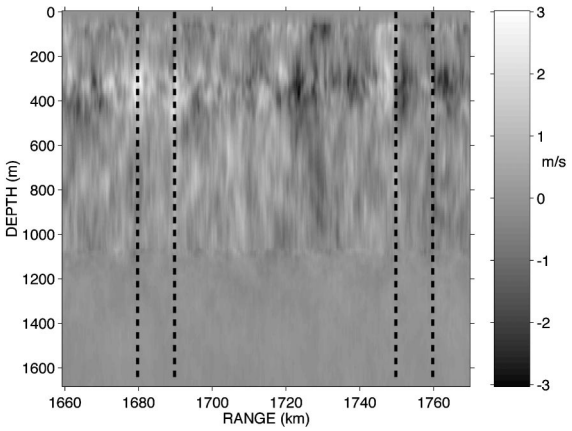


FIG. 8. A realization of a vertical section of sound speed perturbations from internal waves. The value of z_{bot} for this simulation is 4205 m. This field incorporates 50 vertical modes and 128 horizontal wave numbers. The vertical lines indicate the horizontal extent of the section (80 km) and extent of overlap (10 km) as described in Appendix A 2. Only the upper 1640 m are shown.

duction. This desirability is primarily due to the fact that numerical algorithms to solve the wave equation in parabolic form are vastly more efficient than numerical algorithms which solve the Helmholtz wave equation directly (e.g., coupled modes). Further efficiency is obtained by invoking the split-step Fourier algorithm.[18] Most parabolic approximations that implement the split-step algorithm have solutions that depend on a reference sound speed c_0, or are limited to handling acoustic energy confined to small grazing angles ($<15°$). These models have phase errors that accumulate with range, effectively distorting the wavefront and thus yielding errors in the relative and absolute travel times.[19] Our full-wave simulations of the Kaneohe experiment require an algorithm that is insensitive to the reference sound speed and allows for grazing angles up to 90°. The reasoning is the following: The stable early arrivals A–D (see Sec. IV) have travel times that are about 0.4 s later than either what full-wave or ray trace models predict. This discrepancy is probably due to climatic temperature changes in the North Pacific ocean which stem from westward propagating Rossby waves linked to El Niño and the Southern Oscillation.[7] Because these early arrivals are the only stable data, they are used as an anchor for lining up the simulation results and data on a single time scale. The travel time of the late arriving energy, which is attributed to vertical scattering, must then be measured as accurately as possible relative to the early stable arrivals. This dictates that errors in travel times due to sensitivities to grazing angle and/or reference sound speed be minimal. Thus a PE model is used which yields travel times within a few milliseconds accuracy for all the contributing multipaths.[19,20]

A. Pulse synthesis

The method used to synthesize wavefronts and impulse response functions is described in Sec. IV of Tappert *et al.*[20] This method would suffice for comparison with impulse response data for a single pulse transmission, but as mentioned in Sec. I, the data have been incoherently averaged over a day.[1] Besides enhancing the stable arrivals, daily incoherent averaging also broadens these arrivals due to the ''washing out'' of the wavefront time wandering effect from the passage of internal gravity waves.[5] To simulate this broadening of the data, a running average is performed recursively on the impulse response function. If $I(t_n)$ denotes the square modulus of the complex envelope of acoustic pressure at the discrete time t_n, then the averaging technique employed can be expressed as

$$I(t_n) = \tfrac{1}{4} I(t_{n-1}) + \tfrac{1}{2} I(t_n) + \tfrac{1}{4} I(t_{n+1}). \tag{2}$$

The sampling interval was chosen to be $t_n - t_{n-1} = 16$ ms. Running this filter recursively for four iterations suffices to broaden the stable arrivals so as to agree with the widths of the measured stable arrivals, which are about 100 ms.[1]

B. Model input parameters

Some of the more relevant environmental parameters required by the PE model include the computational depth H, number of depth points N, and range step dr. The methods for determining both N and dr are described in Appendix C 1. The computational depth chosen is 8000 m, large enough to allow accurate modeling of sub-bottom interactions, but small enough to minimize computational time.

The shallow bathymetry near the source and receiver allows for significant sub-bottom interaction. For this reason, it is particularly important to model the sub-bottom reasonably well. The PE model used allows for a single sediment layer, modeled as a fluid, with linear sound speed gradient and constant density. Relevant input parameters concerned with sub-bottom interaction of acoustic energy include (1) sediment thickness, (2) sediment density, (3) sound speed ratio at the water–sediment interface, (4) gradient of sound speed in the sediment, and the (5) coefficient of volume attenuation in the sediment.

Because the receiver is at 1433 m depth on a continental rise, some of the acoustic energy contributing to the acoustic multipaths penetrates into the sub-bottom. The steeper ($>10°$), early arriving energy penetrates deeper into the sub-bottom than the later energy, which is in general less steep ($<3°$). The question arises as to how to select the sediment parameters for this situation when only a single sediment layer is modeled. Our solution involves forming a composite wavefront from different simulations that use optimal geoacoustic parameters for energy with both steep and shallow grazing angles (refer to Appendix C 2 for details).

The acoustic propagation model and ocean environments having been described, the next section will discuss and interpret the results of the simulations, as they relate to the data.

IV. MODEL RESULTS AND COMPARISON WITH DATA

Two-dimensional full-wave simulations were performed through the four sound speed fields described in Sec. II. The two-dimensional approximation is valid in this case because a three-dimensional calculation would change the travel time by $O(1)$ ms (Sec. IV of Ref. 1). The synthesized wavefronts and impulse response functions at the receiver depth are shown in Fig. 4. The horizontal line in the left panels of Fig.

4 indicate the receiver depth, so that the impulse responses in the right panels are representative of ''cuts'' of the wavefronts at this depth. The daily incoherently averaged impulse response data, taken 29 November 1983, are superposed for qualitative comparison. The results of the simulations are now presented in the order that they are presented in Fig. 4.

A. Climatology

The modeled wavefront in the top panel of Fig. 4 comes from propagation through a climatological sound speed field. This sound speed field is constructed using Levitus data collected over the months of May, June, and July. The weak range dependence in the environment is revealed in the tail of the wavefront, where distinct modes are evident. Taking into consideration the results from ray tracing,[1] the amplitudes of arrivals A–D are seen to be primarily due to acoustic energy that is initiated in large positive vertical wave numbers. This energy initially is surface reflected, then bottom reflected off the Kaneohe slope, and is finally bottom reflected near the receiver (refer to Fig. 9 of Ref. 2). In the geometric ray representation, the initial launch angle corresponds to about 15°.[7] Positive launch angles and grazing angles are defined to be positive upward in the direction toward the receiver.

Inspection of the modeled impulse response reveals small humps on the right of arrivals A and C. These appear to be remnants of the corresponding negative launch angles near $-15°$, but they are attenuated due to strong bottom interaction near the source. These humps are likely due to acoustic energy that is initially bottom reflected, then surface refracted, and finally bottom reflected near the receiver. To confirm this, a simulation is performed using the same sound speed field, but with the sound speed ratio at the water–sediment interface along the Kaneohe slope set to 1.03, instead of the previous value of 1.02. This gives a critical angle of about 16° instead of the previous value of 11.4°, so that the $-15°$ grazing energy reflects off the bottom, effectively bringing up the hump adjacent to arrival A (not shown).

B. Mesoscale

The wavefront computed from propagation through mesoscale structure indicates strong stability for arrivals A–D (second panel of Fig. 4). Notice also that the envelope of this wavefront has a markedly different shape as compared to that for the pulse propagated through the climatological sound speed field. Especially significant is the time extension of almost 1/2 s in the tail of the wavefront, as compared with the wavefront computed by propagation through the climatology sound speed field (upper panel of Fig. 4).

In geometric acoustics, the travel time change for a single ray that has evolved through unperturbed and perturbed ocean environments can be decomposed into linear and nonlinear components.[22,23] Here, the climatological and mesoscale sound speed fields are considered to be the respective unperturbed and perturbed environments. Denote δT_1 as the total change in travel time, and δT_2 and δT_{NL} as the respective linear and nonlinear components of the change in travel time. Then

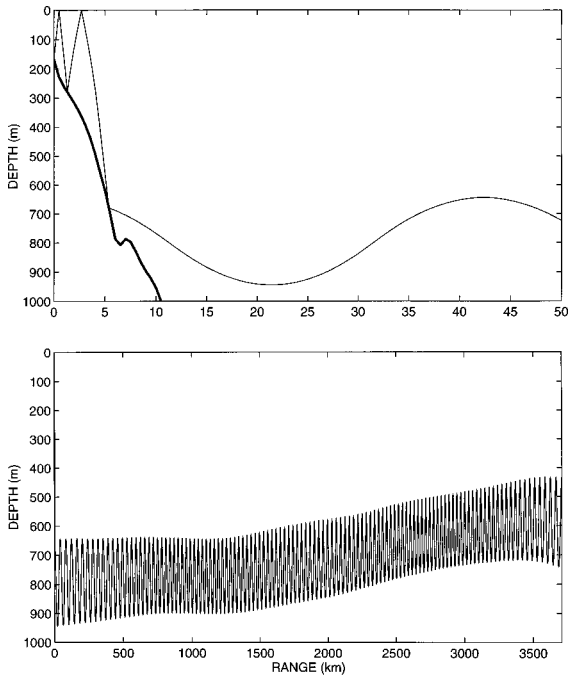


FIG. 9. Path (thin solid line) for the most axial ray through the climatology sound speed field. The upper panel shows the path near the source, and the lower panel shows the entire path. Bathymetry is shown as a thick solid line on the top panel.

$$\delta T_1 = \delta T_2 + \delta T_{NL}. \tag{3}$$

Simulations performed using a ray trace code[1] reveal that the ray with a launch angle of 19.154 degrees has a travel time equal to the cutoff time for the wavefront associated with the climatological environment. This ray's path, shown in Fig. 9, is used to calculate the linear contribution as

$$\delta T_2 \doteq -\int_{\Gamma_0} \frac{\delta c(s)}{c_0^2}\, ds, \tag{4}$$

where Γ_0 is the path through the climatological sound speed field, $c_0(s)$ and $\delta c(s)$ represents the sound speed perturbation due to the mesoscale structure, and ds is an increment of Γ_0.[22,23] We get

$$\delta T_2 = -0.17\,\text{s}. \tag{5}$$

The total travel time change is computed directly from the full-wave simulations. The cutoff time for propagation through the climatological environment is $T_0 = 2507.17$ s (wavefront in first panel of Fig. 4), and the cutoff time for propagation through the mesoscale environment is $T_1 = 2507.61$ s (wavefront in second panel of Fig. 4), so that

$$\delta T_1 \doteq T_1 - T_0 = +0.44\,\text{s}. \tag{6}$$

Then

$$\delta T_{NL} = +0.61\,\text{s}. \tag{7}$$

This leads us to conclude that the total travel time difference seen in the tail of these wavefronts is due primarily to a nonlinear change in travel time. The nonlinear change is an order of magnitude greater than what other simulations report at ranges of about 3000 km for steep energy.[22] This is intriguing, especially since the observed mesoscale structure

TABLE I. Modeled travel times to the nearest 10 ms for peaks A through D through the indicated ocean environments. The travel times from a ray trace model through the mesoscale environment are indicated for comparison (Ref. 1). The climatological average environment is from the spring season of Levitus, and the mesoscale environment is derived from a July 1988 CTD survey.

Ocean environment	Model	Peak travel times (s)			
		A	B	C	D
Climatological	PE	2503.70	2504.06	2504.38	2504.67
Mesoscale	PE	2503.74	2504.07	2504.45	2504.65
	ray	2503.75	2504.09	2504.41	2504.68
Internal waves	PE	2503.74	2504.09	2504.37	2504.66
Mesoscale and internal waves	PE	2503.67	2504.03	2504.38	2504.66

is weak, i.e., $\delta c_{\mathrm{rms}} = 0.6$ m/s. The nonlinear term is only computed for a single realization of mesoscale structure. If one investigates the month-to-month data and measures the time between the multipath arrival A and the latest arriving acoustic energy, it is seen that the cutoff times are not sensitive to changes in the mesoscale structure. For example, Fig. 3 of Ref. 2 illustrates the cutoff times for data recorded over consecutive days in 1983 and one day in 1987. This supports the interpretation that the nonlinear component of the total travel time change is a bias. However, it must be recognized that the impulse response data were recorded at a depth well below the axis of the sound channel. It is not certain how the month-to-month cutoff time variability at a depth of 1433 m correlates with the cutoff time variability at the axis depth (which is about 600 m at the receiver range). This adds some uncertainty to our interpretation.

This interpretation of travel time bias has been investigated for steeper ray paths at ranges up to 3000 km,[22] where it was first shown that the magnitude of the bias is insensitive to the location of the eddies. The bias was shown to have either a plus or minus sign, and had magnitudes of up to about 50 ms, much less than what we observe for the latest arrivals.

C. Internal waves

The two lower panels of Fig. 4 illustrate the wavefronts associated with pulse propagation through internal waves superposed with the climatology and mesoscale environments. Notice that the ''accordionlike'' behavior of the early portion of the wavefronts that is so readily apparent in the upper panels is still observed in the lower panels. The lower caustics of this accordion contribute to the arrivals A–D, as seen in the adjacent impulse response curves. So, it can be concluded that internal waves of nominal strength do not destroy the inherent stability in arrivals A–D (see Table I). Of course, that is to be expected, since this stability is noticed in the data.

The next detail to observe is the strong vertical scattering in the tail portion of the wavefronts in the lower two panels. Internal waves of nominal strength are seen to be a likely candidate for contributing to the vertical scattering the axial energy over one kilometer in extent. The final cutoff in the impulse response curves (right lower panels of Fig. 4) is seen to better match the data, as compared to the wavefronts resulting from propagation through an environment without internal waves.

The lower two panels of Fig. 4 indicate that the positive nonlinear change in travel time due to the observed mesoscale structure swamps any travel time bias induced from internal waves. According to Dashen *et al.*,[24] the internal wave induced bias should be proportional to the logarithm of the acoustic frequency times the square of the propagation range times the strength of the internal waves. Results reported from Colosi *et al.* (refer to Table IV of Ref. 5) at 1000-km range, using twice the nominal GM internal wave strength and an acoustic center frequency of 250 Hz, indicate an internal wave induced bias in the tail of the wavefront of -4 ms. Extrapolating these results to our simulations, one expects an internal wave induced bias of about -24 ms. This is indeed much smaller than the total travel time change of 440 ms due to the mesoscale structure. Notice also that the cutoff time is extended by an additional 1/4 s in the impulse response that is derived from propagation through an environment composed of mesoscale structure superposed with internal waves (lowest panel of Fig. 4), as compared with propagation through an environment composed of a climatological sound speed field superposed with internal waves (third panel of Fig. 4).

All the impulse response curves in Fig. 4 show that absolute travel times of arrivals A–D are similar, within about 40 ms in all simulations (Table I). Simulations performed using different realizations of internal wave fields yielded similar results (not shown). Arrivals A–D changed their travel time by less than 10 ms, although their amplitudes varied by as much as 50%. This corroborates the observation that steep energy is less affected by sound speed fluctuations than energy with shallow grazing angles.[5]

V. VERTICAL SCATTERING

When the horizontal stratification of the ocean waveguide is broken by weak volume inhomogeneities, due to, for example, internal waves, an acoustic pulse undergoes multiple forward scattering. A consequence of this is the spreading in depth of acoustic energy, which we refer to as vertical scattering. Recently, an ''outer intensity envelope'' of the wavefront has been used to illustrate the extent of vertical scattering due to propagation through volume inhomogeneities, such as internal waves.[5] This intensity envelope is constructed from a wavefront of acoustic transmission loss by interpolating the minimum and maximum depth where the transmission loss assumes a predefined cutoff value. This is

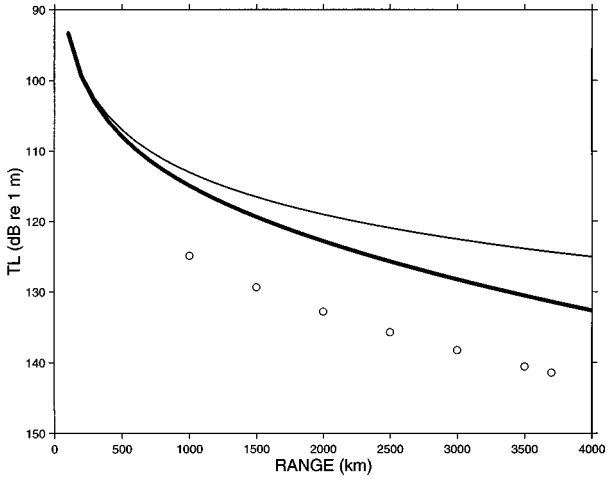


FIG. 10. Transmission loss as a function of range for geometric spreading (thin solid), geometric spreading plus absorption (thick solid), and the intensity cutoff values (circles) used for computing the outer intensity envelopes of Fig. 11.

done throughout the time domain of the wavefront, which depends on the dynamic range of intensity chosen *a priori*. Comparing the envelopes constructed from simulations with and without internal waves gives a measure of the extent of vertical scattering. Here, we take this idea one step further, inquiring how the extent of vertical scattering changes as a function of range.

The measure of vertical scattering as a function of range depends on the dynamic range and reference level of intensity chosen. It is also dependent on which portion of the pulse is chosen. This paper is primarily interested in the vertical scattering of the late arriving ''axial'' acoustic energy. Other parameters affecting the measure of vertical scattering may include the source bandwidth and center frequency, pulse length, and integration time. To clarify the relevant issues, the following thought-experiment is posed: The same Kaneohe experiment is performed, but instead of the single hydrophone, vertical arrays of hydrophones are arranged along the geodesic every 500 km in range. Signal enhancement is performed via incoherent averaging, just as described by Spiesberger and Tappert.[1] If the ocean structure could be turned off and on, what would be the amount of vertical scattering measured due to ocean structure at the range of each vertical array?

If the acoustic energy is contained within the waveguide (i.e., bottom interactions are assumed negligible), then, in an adiabatic environment, the total transmission loss at any field point is due solely to spreading loss and volume absorption. Sound interacts with the bottom near the source and receiver, but away from these regions, bottom interactions are negligible. In a nonadiabatic environment, the principal additional mechanism that affects transmission loss is multiple forward scattering.

As usual for adiabatic waveguide environments, the spreading is assumed spherical out to a range of about 5 km, then cylindrical beyond.[25] Taking into account the linear growth of the number of arrivals at any particular depth (we assume a growth rate of 0.02 arrivals per km,[25] or an average ray loop distance of 50 km), the transmission loss due to spreading is slightly less than, but proportional to, spherical. The transmission loss due to absorption for sub-kilohertz frequencies is predominantly due to boric acid relaxation, and is linear with range (the absorption coefficient is about 1.2 dB/1000 km, see Refs. 2, 26). The cutoff value chosen to construct the envelopes for the wavefronts is 10 dB greater than the transmission loss due to spreading and absorption (Fig. 10).

The tails of the outer envelopes are shown in Fig. 11 at 500-km intervals, starting at 1000 km. The environments used for these simulations are the climatological sound speed field and a structured ocean consisting of internal waves superposed with the mesoscale sound speed field. Two observations are worth mentioning: (1) The envelopes shoal toward the surface with range, following the trend of the sound channel (Fig. 4 of Ref. 1); (2) The extent of vertical scattering is seen to markedly increase beyond the range of 2000 km.

VI. DISCUSSION

Full-wave simulations using a parabolic equation model demonstrate that internal waves, eddies, and bottom characteristics can account for the strong vertical scattering ob-

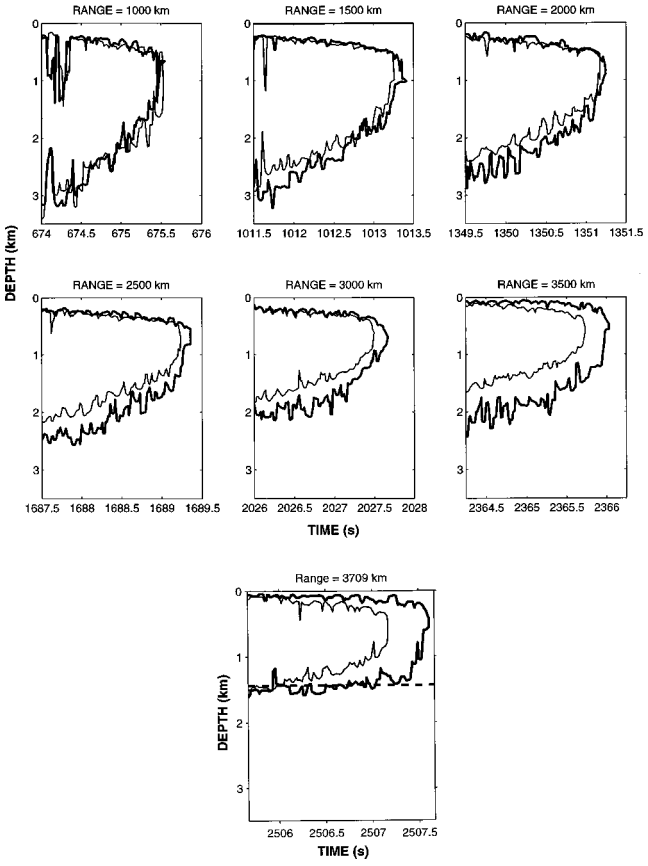


FIG. 11. Outer intensity envelopes of wavefronts at various ranges along the section indicated in Fig. 1. The thin solid curves represent the outer intensity envelopes derived from propagation through the climatological environment, and the thick solid curves represent the outer intensity envelopes derived from propagation through an environment composed of mesoscale with a realization of internal waves superposed. Only the final 1.5–1.75 s of the tail are shown at the indicated ranges, each in a 2-s window. The thick dashed line in the last panel indicates the bottom depth at the receiver range.

served for acoustic pulse-like signals of 133-Hz, 60-ms resolution at the range of 3709.21 km. The internal wave model is range dependent and fully three dimensional, based on the internal wave spectrum of Garrett and Munk.[12,13] The effects of internal waves, mesoscale structure, and bottom interactions are adequate to explain the Kaneohe data.

In our simulations, we strived to avoid ''tuning'' the model to fit the data. The sub-bottom and bathymetry near the source and receiver need to be modeled reasonably well to account for the relative differences in amplitudes seen in the data, but as one would expect, arrival times were not sensitive to how the sub-bottom is modeled. However, the simulations performed did not include the effects of rough bottom scattering near the source. It is not known to what extent the final cutoff time seen in the impulse response data is sensitive to bathymetric features with range scales below one kilometer. Thus our results regarding scattering are consistent with the data, but the scattering of sound from a rough bottom has not been explored in a quantitative manner. It seems that the actual effects of the rough bottom would be suppressed because sound does not actually travel along thin rays. Instead, sound propagates within eigentubes[21] whose broader scales would insonify larger swaths of the bottom.

Oceanic mesoscale structure significantly extend the cutoff in the tail of the wavefront, and it was demonstrated that this wavefront tail extension is attributed to a nonlinear change in travel time. The observations are consistent with a bias interpretation for this nonlinear component of travel time change, and this effect will have to be taken into account if these data are used in tomographic inversions. The stable arrivals A–D, however, are insignificantly affected in regards to travel time (see Table I and Fig. 4). This is in agreement with the relatively small bias [$O(50)$ ms] reported for steep, early arrivals at ranges of 3000 km.[22] This supports previous evidence that arrivals A–D are useful for measuring temperature.[1,2,7]

ACKNOWLEDGMENTS

This research was supported by the Office of Naval Research, Grant No. N00014-95-1-0755. We express thanks to Dr. Fred Tappert and Dr. Diana McCammon for fruitful consultations.

APPENDIX A: MODEL FOR SIMULATING INTERNAL WAVE FIELDS

This Appendix provides a detailed prescription as to how the sound speed fluctuations due to internal waves are numerically generated.

1. Simulating a single section

Let ζ denote the vertical displacement of a water parcel due to small amplitude internal waves, with positive ζ indicating downward displacement. Recall that the depth z is defined as positive downward, with $z=0$ at the surface. Then the contribution to the total sound speed field due to internal waves is

$$\delta c(x,z,t) = -\left(\frac{\partial \overline{c}(z)}{\partial z}\right)_{\text{pot}} \zeta(x,z,t), \tag{A1}$$

where $(\partial \overline{c}(z)/\partial z)_{\text{pot}}$ is the climatologically averaged potential sound speed gradient computed at the middle of a section. The potential sound speed gradient is defined as

$$\left(\frac{\partial \overline{c}}{\partial z}\right)_{\text{pot}} = \left(\frac{\partial \overline{c}}{\partial z}\right) - \left(\frac{\partial \overline{c}}{\partial z}\right)_a, \tag{A2}$$

where the second term on the right-hand side is the adiabatic sound speed gradient. In the ocean, the adiabatic sound speed gradient is expressed in terms of temperature, T, pressure, P, and salinity, S, as

$$\left(\frac{\partial \overline{c}}{\partial z}\right)_a = \left(\frac{\partial \overline{c}}{\partial P}\right)\left(\frac{\partial P}{\partial z}\right)_a + \left(\frac{\partial \overline{c}}{\partial T}\right)\left(\frac{\partial T}{\partial z}\right)_a, \tag{A3}$$

where the second term on the right-hand side is negligible. Both $\partial \overline{c}/\partial z$ and $\partial \overline{c}/\partial P$ are calculated using Levitus' database and Del Grosso's sound speed formula.[8,9] Linear interpolation is used to provide the potential sound speed gradient at any depth of interest.

The field of vertical displacements due to internal waves is described statistically as a sum over internal wave vertical displacements modes $W(j,k,z)$,[15] where j is the mode number index and k is the magnitude of the horizontal wave number vector. The vertical modes and corresponding eigenfrequencies, $\omega(j,k)$, satisfy the Sturm–Liouville problem[27,28]

$$\frac{d^2W}{dz^2} + \left(\frac{N^2(z)-\omega^2}{\omega^2-\omega_i^2}\right)k^2 W = 0, \tag{A4}$$

with boundary conditions

$$W(z=0) = W(z=z_{\text{bot}}) = 0, \tag{A5}$$

where z_{bot} is the range averaged ocean bottom depth, $N(z)$ the buoyancy frequency, and ω_i the inertial frequency, all determined at the center of a section. The eigenfunctions are orthonormal over their natural weighting $N^2(z)-\omega_i^2$. The buoyancy profile is determined from

$$N^2(z) = -g\rho^{-1}\partial_z\rho - \frac{g^2}{\overline{c(z)}^2}, \tag{A6}$$

where g is gravity, and ρ is the density of sea-water, which is computed according to Fofonoff.[29]

Let y be the horizontal spatial coordinate perpendicular to x, and let k_1 and k_2 be the respective x and y components of the horizontal wave number vector; i.e., $k=\sqrt{k_1^2+k_2^2}$. Then, without any further approximations, a three-dimensional, time dependent field of vertical displacements due to small amplitude internal wave motions is described by

$$\hat{\zeta}(x,y,z,t)=\mathcal{D}^{-1}\,\mathrm{Re}\left\{\int_{-\infty}^{\infty}dk_1\int_{-\infty}^{\infty}dk_2 Q(k_1,k_2,z,t)\times\exp(ik_1x)\exp(ik_2y)\right\},\tag{A7}$$

where

$$Q(k_1,k_2,z,t)=(2\pi k)^{-1}\sum_j\sqrt{F_\zeta(k,j)}[A(j,k_1,k_2)+iB(j,k_1,k_2)]W(j,k,z)\exp[i\omega(j,k)t].\tag{A8}$$

Choosing $A(j,k_1,k_2)$ and $B(j,k_1,k_2)$ as independent Gaussian random numbers with zero mean and unit variance, the field of vertical displacements will then be a Gaussian random process with zero mean and variance given by the power spectrum $F_\zeta(k,j)$. This spectrum is taken to be consistent with the empirical spectrum of Garrett and Munk,[12,13] given by

$$F_\zeta(k,j)=\zeta_{\mathrm{ref}}^2\,\frac{8}{\pi^2}\,\frac{k^2k_j}{(k^2+k_j^2)^2}\,\frac{j_*}{j^2+j_*^2},\tag{A9}$$

where $k_j=j\pi\omega_i/(N_0b)$, $N_0=3$ cph, $\zeta_{\mathrm{ref}}=7.3$ m, $j_*=3$. The parameter b is the buoyancy scale depth in km. The normalization constant $\mathcal{D}$ in Eq. (A7) is set according to

$$\mathcal{D}^2=(2\pi k)^{-1}\sum_j\int dkF_\zeta(k,j)W^2(j,k,z_{\mathrm{ref}}),\tag{A10}$$

where z_{ref} is the depth where $N(z_{\mathrm{ref}})=N_0$. Thus the field will be insured to statistically have the correct variance at every depth. The derivation of Eq. (A9) is provided in Appendix B 1. The two-dimensional field along the geodesic is then simply

$$\zeta(x,z,t)=\hat{\zeta}(x,y=0,z,t).\tag{A11}$$

The sound speed fluctuations for a section are computed according to Eq. (A1).

2. Section patching

Let the subscript m denote the mth section of sound speed fluctuations, as in $\delta c_m(x,z,t)$. Also, denote the horizontal extent of each section by X and the extent of overlap for adjacent sections be X_p. The sound speed fluctuation values inside the patching region are computed according to

$$\delta c(x,z,t)=S_m(x)\,\delta c_m(x,z,t)+S_{m+1}(x)\,\delta c_{m+1}(x,z,t),\tag{A12}$$

where

$$S_m(x)=\begin{cases}\cos^2\left[\left(\dfrac{x-m(X-X_p)}{X_p}\right)\dfrac{\pi}{2}\right];\\ \quad\text{if } m(X-X_p)\leqslant x\leqslant m(X-X_p)+X_p,\\ 0;\quad\text{otherwise,}\end{cases}\tag{A13}$$

$$S_{m+1}(x)=\begin{cases}\sin^2\left[\left(\dfrac{x-m(X-X_p)}{X_p}\right)\dfrac{\pi}{2}\right];\\ \quad\text{if } m(X-X_p)\leqslant x\leqslant m(X-X_p)+X_p.\\ 0;\quad\text{otherwise.}\end{cases}\tag{A14}$$

The results reported here use $X=80$ km and $X_p=10$ km. After the sections of sound speed fluctuations are patched together, the total sound speed field is computed from Eq. (1).

APPENDIX B: DERIVATIONS OF USEFUL FORMULAS

This Appendix provides derivations of formulas used for simulating internal wave sound speed fluctuations and verifying that the simulations follow the chosen theoretical model.

1. Derivation of power spectrum of vertical displacement

The vertical displacement spectrum, Eq. (A9), follows directly from the Garrett and Munk form expounded on page 285, Eq. 9.19 of Ref. 30. It is expressed in terms of frequency and vertical mode number as

$$f_\zeta(\omega,j)=b^2\frac{N_0}{N}\left(\frac{\omega^2-\omega_i^2}{\omega^2}\right)E(\omega,j),\tag{B1}$$

where the energy density $E(\omega,j)$ is

$$E(\omega,j)=E_0B(\omega)H(j),\tag{B2}$$

$$B(\omega)=\frac{2}{\pi}\,\frac{\omega_i}{\omega(\omega^2-\omega_i^2)^{1/2}},\tag{B3}$$

$$H(j)=\frac{(j^2+j_*^2)^{-1}}{\Sigma_j(j^2+j_*^2)^{-1}}.\tag{B4}$$

The dimensionless internal wave energy parameter E_0 is set to 6.3×10^{-5}, and b is the buoyancy scale depth. The vertical displacement spectrum as a function of magnitude of horizontal wave number vector and vertical mode number is

$$F_\zeta(k,j)=f_\zeta(\omega=\omega(k),j)\,\frac{d\omega}{dk},\tag{B5}$$

where the Jacobian of transformation, $d\omega/dk$, is obtained from the WKB relation $\omega=[(N_0b/j\pi)^2k^2+\omega_i^2]^{1/2}$:[30]

$$\frac{d\omega}{dk}=\left(\frac{N_0b}{j\pi}\right)^2\frac{k}{\left[\left(\dfrac{N_0b}{j\pi}\right)^2k^2+\omega_i^2\right]^{1/2}}.\tag{B6}$$

Substituting this expression into Eq. (B5), defining $k_j=j\pi\omega_i/N_0b$, yields

$$F_\zeta(k,j)=\frac{4}{\pi}\,\frac{b^2N_0E_0}{2N(z)}H(j)\,\frac{k^2k_j}{(k^2+k_j^2)^2}.\tag{B7}$$

Defining $\zeta_{\mathrm{ref}}^2=b^2E_0/2$ to be the reference vertical displacement at the depth $z=z_{\mathrm{ref}}$ where $N(z_{\mathrm{ref}})=N_0$ eliminates the vertical displacement spectrum's dependence on depth, giving

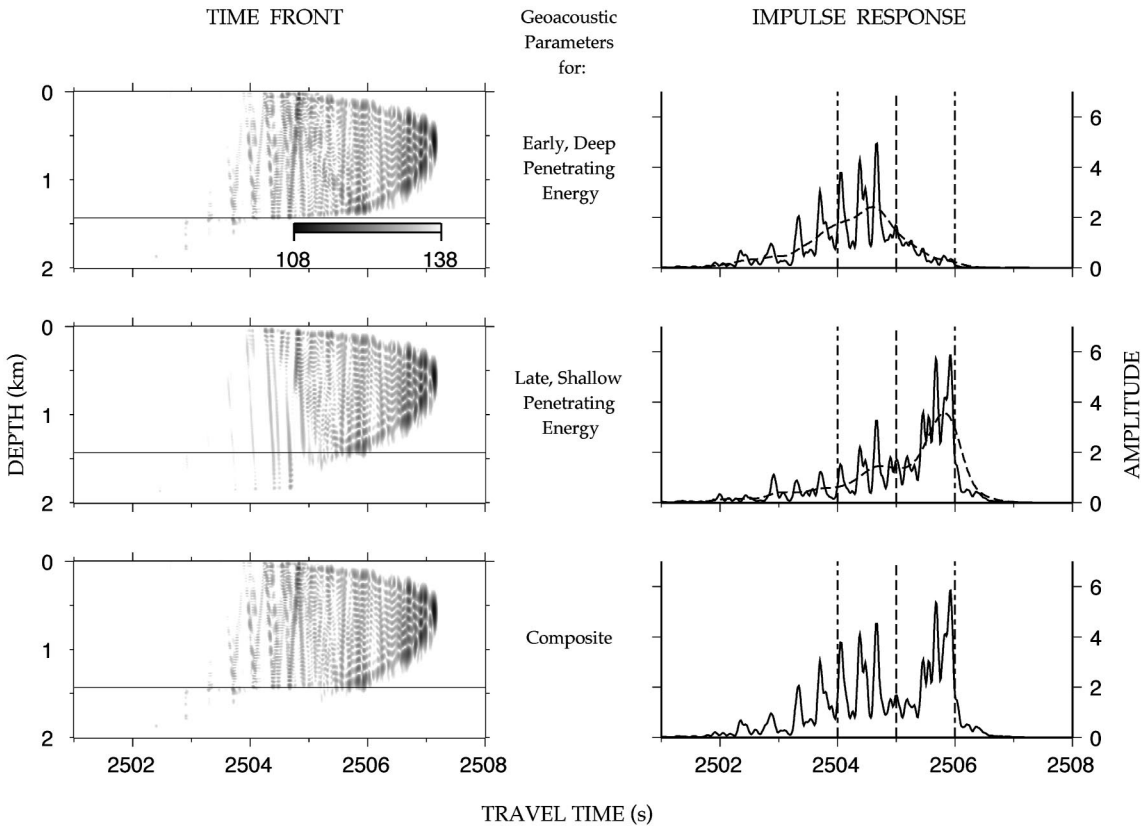


FIG. C1. Wavefronts and associated impulse response functions at the receiver range for the Levitus environment using bottom parameters that were chosen to model: (1) the early arriving, steep penetrating energy (upper panel); (2) the late, shallow penetrating energy (middle panel). The lower panel shows the composite wavefront and impulse response function constructed by smoothly joining the amplitude of the wavefront after it was Pascal filtered in range over all depths on the PE depth mesh. The smoothing was done using one quarter cycle of a cosine squared window applied over the 2-s record bounded by the vertical long, dash-dotted lines. The short, dash-dotted lines indicate the time where the two different wavefronts were joined. The wavefronts are shown using a gray scale with 30-dB dynamic range and the impulse response functions are shown with an arbitrary amplitude scale. Note the horizontal line in the left panels is at the hydrophone depth for the impulse response functions shown in the right panels.

$$F_\zeta(k,j)=\frac{4}{\pi}\zeta_{\mathrm{ref}}^2 H(j)\frac{k^2 k_j}{(k^2+k_j^2)^2}. \tag{B8}$$

Equation (A9) is obtained by applying the Poisson summation formula to the normalization of $H(j)$.

2. Derivation of horizontal length scale

Starting from Eq. (A7), without loss of generality, set $t=0$, and instead of taking the real part of a complex random process, choose the random variables such that $Q(-k_1,-k_2,z,0)=Q^*(k_1,k_2,z,0)$. Then, since for each depth the process ζ is spatially homogeneous, aside from the normalization constant, $\mathcal{D}$, the autocorrelation function is related to the power spectral density of $Q(k_1,k_2,z,0)$ by a Fourier transform,

$$\langle|Q(k_1,k_2,z,0)|^2\rangle=\left(\frac{1}{2\pi}\right)^2\int_{-\infty}^{\infty}dx'\int_{-\infty}^{\infty}dy'\,C(x',y',z)\times\exp[-i(k_1x'+k_2y')], \tag{B9}$$

where $C(x',y',z)=\langle\hat{\zeta}(x,y,z,0)\hat{\zeta}(x+x',y+y',z,0)\rangle$ is the autocorrelation function (the angled brackets denote the ensemble average operator). The inverse relation to this gives

$$C(x',y',z)=\int_{-\infty}^{\infty}dk_1\int_{-\infty}^{\infty}dk_2\langle|Q(k_1,k_2,z,0)|^2\rangle\times\exp[i(k_1x'+k_2y')]. \tag{B10}$$

Equation (B9) is the two-dimensional form of the Wiener–Kinchine theorem.[31] Since the random variables, $A(j,k_1,k_2)$ and $B(j,k_1,k_2)$, are uncorrelated for different vertical modes, j,

$$\langle|Q(k_1,k_2,z,0)|^2\rangle=\sum_j W^2(j,k,z)F_\zeta(k,j). \tag{B11}$$

The horizontal integral scale length is defined as

$$L_H(z)\doteq\frac{\int_0^\infty drC(r\cos\alpha,r\sin\alpha,z)}{C(0,0,z)}, \tag{B12}$$

where $(x',y')=(r\cos\alpha,r\sin\alpha)$, α representing the angle r makes with the x-axis.[32] Substituting Eq. (B11) into Eq. (B10), converting the denominator of Eq. (B12) to polar coordinates, and using the fact that $\langle|Q(k_1,k_2,z,0)|^2\rangle$ is an even function of k_1 and k_2, we obtain

$$L_H(z)=\frac{\int_0^\infty dkk^{-1}\Sigma_j W^2(j,k,z)F_\zeta(j,k)}{\int_0^\infty dk\Sigma_j W^2(j,k,z)F_\zeta(j,k)}. \tag{B13}$$

TABLE CI. Range step and number of depth points used in PE simulations for the three regions delineated in Fig. 2.

Region	1	2	3
Range (km)	0–200	200–3600	3600–3709.21
N	2048	1024	2048
dr (m)	25	100	25

TABLE CII. Geoacoustic parameters used for generating composite time fronts. The parameter σ denotes the ratio of sound speed at the bottom of the water to that at the top of the sediment.

	σ	ρ_s (kg/m^3)	α (dB/km/kHz)
A–D	1.020	1.8	0.02
Tail	1.002	1.5	0.02

APPENDIX C: SYNTHESIS OF ACOUSTIC WAVE FIELDS

1. Convergence

This Appendix describes the method used to insure convergence of the PE simulation. Additionally, it explains our hybrid method used to synthesize wavefronts when realistic modeling of acoustic propagation in the sub-bottom is required within the constraint of using a PE model. Both of these methods were developed with the primary goal of retaining numerical efficiency as much as possible.

The ocean environment is separated into three regions: (1) a downslope region covering the first 200 km in range; (2) a deep, waveguide region extending from 200 km to 3600 km in range; (3) an upslope region extending from 3600 km to 3709.21 km, the receiver range. Our criterion for convergence is to compare travel times and amplitudes at selected depths at the end of each region for two runs that differ only in either the number of depth points N or range step dr. The selected depths were chosen to be 200, 400, 800, 1000, and 1600 m. The run with the coarser mesh is said to have converged if the travel times and amplitudes, taken over a 40-dB dynamic range (measured down from the maximum intensity) differ by less than 3 ms and 1 dB, respectively.

A low pass filter is applied in the wave number domain of the PE field for each frequency when a smaller number of depth points N is desired; e.g., in going from region 1 to region 2. Similarly, zero padding in the wave number domain is used when a larger number of depth points is desired. The values of dr and N used in each region are summarized in Table CI.

2. Selection of geoacoustic parameters

Because the acoustic propagation model assumes a constant density in a single sediment layer at any given range, and in reality the density is depth dependent, the issue of how to choose the appropriate sediment density arises. There are many ways to resolve this problem. The most obvious way is to allow for a density profile in the sediment, while still incorporating the range dependent c_0-insensitive transformation. Since a model of this form does not exist, our solution was to form a composite of two separate runs, each run using a density appropriate for the two extreme cases; (1) deep penetrating, earlier arriving energy, and (2) shallow, late arriving energy. The wavefronts from each simulation are smoothly joined to form a composite wavefront and associated impulse response curve at the receiver range. Figure C1 summarizes the composition method used synthesizes the presented wavefronts and impulse functions. The relevant geoacoustic parameters used in each case are now described. As in Fig. 8, the horizontal line in the left panels at the receiver depth reveal that the impulse responses in the right panels are ''cuts'' of the wavefront at this depth.

Using the NGDC database,[33] it was determined that the sub-bottom near the receiver is composed predominantly of a 600-m layer of ooze and silt. Near the water–sediment interface, the density expected to be near 1.5 g/cm^3, with a sound speed gradient close to 1 s^{-1}.[34,35] The ratio of sediment sound speed to water sound speed at the water–sediment interface is about 1.002, yielding a critical grazing angle of 3.6° (Diana McCammon, personal communication). The attenuation coefficient of 0.02 dB/m/kHz is appropriate for this type of sediment and acoustic frequency.[36]

The late arriving, small grazing angle energy in the tail of the pulse only significantly interacts with a small region in depth into the sediment; it is refracted up and out before deep penetration. Here, the values determined above near the water–sediment interface are used. For the steeper arriving energy, deeper bottom penetration is expected and an integrated value of density of 1.8 g/cm^3 is used. To be consistent, a correspondingly larger sound speed ratio of 1.02 is used, yielding a critical angle of 11.4°. Table CII summarizes the geoacoustic parameters in the simulations.

[1] J. L. Spiesberger and F. D. Tappert, ''Kaneohe acoustic thermometer further validated with rays over 3700 km and the demise of the idea of axially trapped energy,'' J. Acoust. Soc. Am. **99**, 173–184 (1996).
[2] J. L. Spiesberger, K. Metzger, and J. A. Furgerson, ''Listening for climatic temperature change in the northeast pacific: 1983–1989,'' J. Acoust. Soc. Am. **92**, 384–396 (1992).
[3] J. A. Colosi and S. M. Flatté, ''Mode coupling by internal waves for multimegameter acoustic propagation in the ocean,'' J. Acoust. Soc. Am. **100**, 3607–3620 (1996).
[4] J. A. Colosi, ''Random media effects in basin-scale acoustic transmissions,'' in *Monte Carlo Simulations in Oceanography*, pp. 157–166. A'ha Huliko'a Hawaiian Winter Workshop, University of Hawaii at Manoa, 1997.
[5] J. A. Colosi, S. M. Flatté, and C. Bracher, ''Internal-wave effects on 1000-km oceanic acoustic pulse propagation: Simulation and comparison with experiment,'' J. Acoust. Soc. Am. **96**, 452–468 (1994).
[6] T. Vincenty, ''Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations,'' Surv. Rev. **176**, 88–94 (1975).
[7] J. L. Spiesberger, H. E. Hurlburt, M. Johnson, M. Keller, S. Meyers, and J. J. O'Brien, ''Acoustic thermometry data compared with two ocean models: The importance of Rossby waves and ENSO in modifying the ocean interior,'' Dyn. Atmos. Oceans **26**, 209–240 (1998).
[8] S. Levitus, ''Climatological atlas of the world ocean,'' in NOAA Prof. Pap. 13. U.S. Government Printing Office, Washington, DC, 1982.
[9] V. A. Del Grosso, ''New equation for the speed of sound in natural waters with comparisons to other equations,'' J. Acoust. Soc. Am. **56**, 1084–1091 (1974).
[10] J. B. Bowlin, J. L. Spiesberger, T. F. Duda, and L. F. Freitag, ''Ocean acoustical ray-tracing software ray,'' Technical report, Woods Hole Oceanographic Institution, 1992.
[11] National Geophysical Data Center, Boulder, CO, ''5 minute gridded world elevations and bathymetry—a digital database,'' 1987.
[12] C. Garrett and W. Munk, ''Space-time scales of internal waves,'' Geophys. Fluid Dyn. **2**, 225–264 (1972).

[13] C. Garrett and W. Munk, ''Space-time scales of internal waves: a progress report,'' J. Geophys. Res. **80**, 291–297 (1975).
[14] L. B. Dozier and F. D. Tappert, ''Statistics of normal mode amplitudes in a random ocean,'' J. Acoust. Soc. Am. **63**, 353–365 (1978).
[15] S. Flatté, R. Dashen, W. Munk, K. Watson, and F. Zachariasen, *Sound Transmission Through a Fluctuating Ocean* (Cambridge University Press, Cambridge, 1979).
[16] M. Leontovich and V. Fock, ''Solution of the problem of propagation of electromagnetic waves along the earth's surface by the parabolic equation method,'' Zh. Eksp. Teor. Fiz. **16**, 557–573 (1946).
[17] Frederick D. Tappert, *The Parabolic Approximation Method*, edited by J. B. Keller and J. S. Papadakis (Springer-Verlag, New York, 1977), pp. 224–287.
[18] F. D. Tappert, K. B. Smith, and M. A. Wolfson, ''Analysis of the split-step Fourier algorithm for the solution of parabolic wave equations,'' Math. Modeling Sci. Comp. (in press).
[19] F. D. Tappert and M. G. Brown, ''Asymptotic phase errors in parabolic approximations to the one-way Helmholtz equation,'' J. Acoust. Soc. Am. **99**, 1405–1413 (1996).
[20] F. D. Tappert, J. L. Spiesberger, and Linda Boden, ''New full-wave approximation for ocean acoustic travel time predictions,'' J. Acoust. Soc. Am. **97**, 2771–2782 (1995).
[21] J. Bowlin, ''Generating eigenray tubes from two solutions of the wave equation,'' J. Acoust. Soc. Am. **89**, 2663–2669 (1997).
[22] J. L. Spiesberger, ''Ocean acoustic tomography: Travel time biases,'' J. Acoust. Soc. Am. **77**, 83–100 (1985).
[23] W. Munk and C. Wunsch, ''Biases and caustic in long-range acoustic tomography,'' Deep-Sea Res. **32**, 1317–1346 (1985).
[24] R. Dashen, S. M. Flatté, and S. A. Reynolds, ''Path-integral treatment of acoustic mutual coherence functions for ray in a sound channel,'' J. Acoust. Soc. Am. **77**, 1716–1722 (1985).
[25] W. Munk, P. Worcester, and C. Wunsch, *Ocean Acoustic Tomography* (Cambridge University Press, Cambridge, 1995).
[26] C. S. Clay and H. Medwin, *Acoustical Oceanography: Principles and Applications* (Wiley, New York, 1977).
[27] C. Eckart, *Hydrodynamics of the Oceans and Atmospheres* (Pergamon, New York, 1960).
[28] O. M. Phillips, *The Dynamics of the Upper Ocean* (Cambridge University Press, Cambridge, 1977).
[29] N. P. Fofonoff, ''Physical properties of sea-water,'' in *The Sea*, Volume 1, edited by M. N. Hill (Interscience Publishers, New York, 1962), Chap. 1.
[30] C. Wunsch and B. Warren, *The Evolution of Physical Oceanography* (MIT Press, Cambridge, 1981).
[31] R. L. Stratanovich, *Topics in the Theory of Random Noise*, Vol. 1 (Gordon and Breach, New York, 1963).
[32] V. I. Tatarski, *Wave Propagation in a Turbulent Medium* (Dover, New York, 1961).
[33] D. L. Divins and B. Eakins, ''Total sediment thickness map for the Southeast Pacific Ocean,'' in Intergovernmental Oceanographic Commission, edited by G. B. Udintsev, International Geological-Geophysical Atlas of the Pacific Ocean (in press).
[34] E. L. Hamilton, ''Sound velocity-density relations in sea-floor sediments and rocks,'' J. Acoust. Soc. Am. **63**, 366–377 (1978).
[35] E. L. Hamilton, ''Sound velocity gradients in marine sediments,'' J. Acoust. Soc. Am. **65**, 909–922 (1979).
[36] E. L. Hamilton, ''Sound attenuation as a function of depth in the sea floor,'' J. Acoust. Soc. Am. **59**, 528–536 (1976).

Statistical characterization of high-frequency shallow-water seafloor backscatter

Anthony P. Lyons[a] and Douglas A. Abraham[b]
SACLANT Undersea Research Centre, APO AE 09613-5000

(Received 11 September 1997; accepted for publication 6 May 1999)

Knowledge of the background reverberation environment is a prerequisite for the design of any target detection scheme. While the problem of understanding and predicting high-frequency background seafloor reverberation level or mean energy scattered per unit area of the seabed has received considerable attention, studies of high-frequency reverberation amplitude statistics are relatively scarce. Of these studies, many have dealt with scattering from more or less homogeneous seafloors in terms of bottom type, whereas most shallow-water areas will not be homogeneous but will have patchiness in space and time which will have a strong influence on scattered amplitude statistics. In this work, a comparison is presented between 80-kHz seafloor backscatter statistics obtained at shallow-water sites around Sardinia and Sicily. The data include measurements from several distinct bottom provinces, including sites with *Posidonia Oceanica* sea grass and sites covered with live shellfish. Results of the analysis are cast both in terms of mean power level or backscattering coefficient as well as of the amplitude statistics. The reverberation statistics did not always exhibit a Rayleigh probability distribution function (PDF), but exhibited statistical distributions with heavier tails. Several more appropriate models of reverberation PDF were examined in order to better describe the measured amplitude distributions. The Rayleigh mixture and the *K* models were found to be the most robust in describing the observed data.
[S0001-4966(99)06008-7]

PACS numbers: 43.30.Gr, 43.30.Re, 43.30.Hw [DLB]

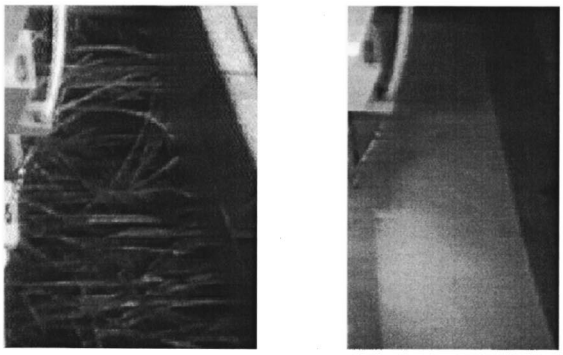
INTRODUCTION

The acoustic detection and identification of objects on the seafloor is hindered by seafloor reverberation. Although the problem of understanding and predicting high-frequency background reverberation level or mean energy scattered per unit area of the seabed has received considerable attention, studies of high-frequency reverberation statistics are relatively scarce. Of these studies, many have dealt with scattering from more or less homogeneous seafloors in terms of bottom type or scatterer distribution.[1–3] Most shallow-water areas, however, will not be homogeneous but will have patchiness in space and time, which may be a result of the local biology. Shellfish are an example of spatial inhomogeneity which often do not exist uniformly on the seabed but are distributed in clumps of varying density with distance scales of decimeters to meters. The motion of seagrass in swell or currents causes a constantly changing number of scattering sites which can be thought of as time-varying patchiness on a scale of several seconds.

Clutter is the acoustic expression of the nonuniformity of these types of seafloor environments. When the effective number of scatterers in the resolution cell of a sonar is large enough, the amplitude distribution is expected to be Rayleigh as the central-limit theorem holds resulting in Gaussian reverberation. The changes in density of scatterers commonly found in shallow water suggests that this model might not always be appropriate, especially when the area ensonified by the transmit and receive beam patterns is not large enough to encompass enough of the patches of differing scatterer density. The non-Rayleigh nature of the amplitude statistics for seafloor scattering has been noted before,[2,3] and attempts have been made to fit various models to the observed amplitude distributions including Weibull, log-normal, and Gaussian probability distributions functions (PDFs). Several distributional models with a somewhat more physical basis than the log-normal or the Weibull can be described as being the product of two components. These include the *K* distribution,[4,5] and models developed by McDaniel[6] and Crowther.[7] While these models have been successful in describing various characteristics of the amplitude PDFs, such as the beamwidth dependence, difficulty in parameter estimation and lack of generality could be considered drawbacks. Another alternative PDF model is one composed of a mixture of Rayleigh distributions.[8]

This paper presents a statistical analysis of 80-kHz acoustic seafloor backscatter data collected at 14 shallow-water sites (17 – 90 m depth) around Sardinia and Sicily. The study sites consisted of areas with coarse sediments, fine sediments, and also areas with a variety of more inhomogeneous (i.e., patchy) bottom types, including sites with *Posidonia Oceanica* sea grass and sites covered with live shellfish. Examples of three of the types of seafloor studied are shown in the video stills seen in Fig. 1. The diversity of sites studied allowed an excellent opportunity to examine the statistics of reverberation from a wide variety of seafloor environments in order to determine more appropriate statistical models for amplitude PDFs than the Rayleigh. Results of statistical analysis are cast in terms of mean power level or backscattering coefficient as well as analysis of the ampli-

[a]Electronic mail: lyons@saclantc.nato.int
[b]Present address: University of Connecticut, Electrical & Systems Engineering Department, 260 Glenbrook Road, U-157, Storrs, CT 06269-2157. Electronic mail: d.a.abraham@ieee.org



(a) Posidonia covered sand bottom - station RR06

(b) Sand bottom - station RR12

(c) Shell covered sand bottom - station RR13

FIG. 1. Examples of the diversity of sand seafloors: the potential for inhomogeneity is observed in the shell and posidonia seagrass examples.

tude statistics. Rayleigh, Weibull, *K*, and 3-component Rayleigh mixture probability distribution functions are compared to measured data, and a nonparametric test is used to describe the goodness of fit between modeled and measured amplitude distributions. The relationship between the mean scattered level and bottom type is also examined.

I. EXPERIMENTAL MEASUREMENTS AND DATA ANALYSIS

The 80-kHz acoustic system used in this study has been fully calibrated against reference hydrophones at the SACLANT Undersea Research Center to quantify source level and beam pattern. The configuration of the acoustic system used in this study is displayed in Fig. 2 for two different tilt angles of the transducer beam axis. The dark

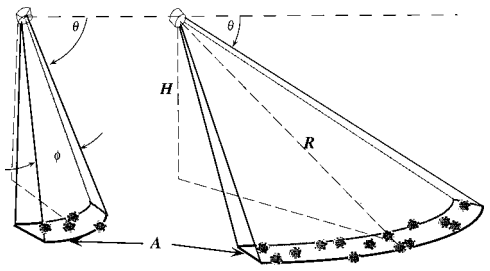


FIG. 2. The experimental configuration displaying the relationship between resolution cell size (a function of the system parameters of height, grazing angle, transmit and receive beam patterns, and pulse length) and the patchiness of the complex shallow-water seafloor environment.

patches in this figure are illustrative of patches of different scattering strength within the resolution cell of the sonar. In this figure, ϕ is the equivalent two-way horizontal beam width of a line array,[9] H is the height of the transducer, and R is the slant range to the ensonified area. The transducer height above the seafloor, H, was between 5 and 9 m for all measurements. The beam pattern of the sonar used in this study was rotationally symmetric with a beamwidth of 20 deg. From each experimental site, 200 1-ms pulses were transmitted at a rate of 1 per s and the returns at each of four transducer tilt angles were analyzed. This arrangement, along with the 20 deg vertical beamwidth of the transducer, allowed measurements to be taken over a large range of grazing angles. The received reverberation data were matched filtered using the transmitted waveform (1 ms sinusoid at 80 kHz, which has an effective bandwidth of 1 kHz), baseband ed to zero Hertz and resampled at a frequency of 20 kHz. Using the transducer calibration value (pressure to voltage transfer function), processing gain and system gains, the absolute received levels at the hydrophone were recovered from the recorded data. System gains were measured while at sea. In order to eliminate the influence of noise on the the analysis of scattering strength and amplitude statistics, only the portions of the echo for which the two-way beam pattern loss was less than 6 dB and for which the noise intensity was at least 10 dB below the signal intensity were used.

A. Scattering strength

In order to obtain quantitative seafloor information from the received level, three effects that modify the sound-pressure level as the pulse travels from the source to the seafloor and back to the receiver have been taken into account. These three factors are the effects of the beam pattern (BP), transmission loss (TL), including both spherical spreading and absorption (α), and equivalent ensonified area

$$A = \phi R(\sqrt{(R + c\tau/2)^2 - H^2} - \sqrt{R^2 - H^2}). \tag{1}$$

With the height of the transducer held constant, the ensonified area (resolution cell) increases with range (or conversely decreases with grazing angle, θ) which will be seen to have a large effect on the scattering amplitude statistics. With the above-described components, seafloor backscattering strength as a function of time can readily be calculated using an inverted form of the sonar equation with a knowledge of the source level, transducer calibration, and logging calibration. This is accomplished with the equation

$$\mathrm{BSS} = \mathrm{RL} - \mathrm{SL} - 2\mathrm{BP} + 2\alpha R + 2\mathrm{TL} - 10\log A, \tag{2}$$

where BSS is mean backscattering strength in dB/m^2, RL is the received level on the hydrophone in dB $re{:}1\ \mu$Pa, and SL is the transducer on-axis source level in dB $re{:}1\ \mu$Pa at 1 m. The backscattering strength as a function of grazing angle can then be obtained from knowledge of the transducer height and the sound speed of the sea water. In data processing, only grazing angles within the 3-dB down points of the one-way vertical-beam pattern were considered (20 deg for the transducer used for this experiment). The 20 deg vertical beamwidth of the transducer allowed scattering-strength measurements to be taken with four tilt angles over a large

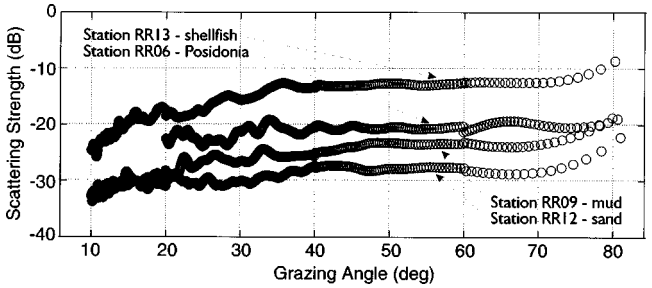


FIG. 3. Bottom backscatter strength as a function of grazing angle for four different bottom types: shellfish covered sand, *Posidonia Oceanica* seagrass covered sand, medium sand, and mud.

range of grazing angles (for most sites, the range of grazing angles from which data were obtained was from 10 deg to 80 deg).

Examples of backscattering strength versus grazing angle resulting from the inversion described above for four of the sites studied are shown in Fig. 3. Because all system-dependent factors, as well as measurement geometry effects (including spreading loss and absorption) and the ensonified area contribution have been removed from the original raw data, the resulting inverted values represent the true quantitative acoustic response of the seafloor (backscattering strength.) Thus, in this graph, quantitatively correct values indicate the different scattering properties of the four sites. The general patterns of the curves for the measurement sites are consistent with values reported near this frequency in that they approach a maximum as they near normal incidence, fall off to a nearly constant level over a wide range of grazing angles, then decrease at low grazing angles. In some of the study sites, the reverberation-to-noise ratio is too low to give reliable scattering-strength estimates at the smallest grazing angles as seen in the posidonia example in Fig. 3, where the minimum grazing angle for which data is presented is 20 deg. The shellfish-covered site had extremely high levels of backscatter at 80 kHz, as did the posidonia-covered site. Surprisingly, the mud site shown in these examples had slightly higher levels of backscattering strength, suggesting that absolute level is not sufficient to separate different bottom types. The next section will present a quantitative comparison between the different bottom types and their respective backscatter strength for different grazing regimes.

B. Amplitude statistics

Acoustic reverberation typically exhibits a decay in the power level with range as the transmission loss increases. However, the statistical distribution of the reverberation may also vary. Statistical analysis of reverberation requires independent and identically distributed data. In order to obtain independent data, the data contained within an individual ping may be decimated. The decimation is accomplished by taking samples separated by the correlation width of the data. It was confirmed by correlation analysis that the decimation (which amounted to taking every 15th point) adequately decorrelated the data. The resulting data are statistically independent and indexed by grazing angle and ping number. Because the portions of the signal within the two-way beam pattern are shorter in time for steeper angles, higher grazing angles will have fewer data points than lower grazing angles for a given sampling frequency. To analyze the change in the statistical distribution of the reverberation and not the change in the power level, the data for each individual grazing angle are normalized by the power level as estimated from the ensemble of 200 pings.

Ideally, a statistical analysis would be performed for each grazing angle from 10 deg–80 deg as obtained from the four tilt angles (10 deg, 30 deg, 50 deg, and 70 deg). However, as an ensemble of only 200 pings is available and because it is expected that the reverberation statistics change slowly with grazing angle, grouping data across grazing angles is necessary and feasible. To avoid the issues of nonstationarity introduced by acquiring the data for different tilt angles at different times, the data from each tilt angle are grouped. To ensure that the grouped data are identically distributed (i.e., the normalized amplitude statistics are stationary), the Mann–Whitney test was applied as in Stanic and Kennedy[2,3] across grazing angle and ping number. Data that failed the Mann–Whitney test with a 5% probability of type I error (i.e., the probability that identically distributed data fail the Mann–Whitney test is 0.05) were excluded from the analysis.

Rayleigh, Weibull, *K*, and three-component Rayleigh mixture distributions were compared to the experimental PDFs. The Weibull, *K*, and Rayleigh mixture distributions have the Rayleigh distribution as a submember. The Weibull distribution has traditionally been used to describe amplitude distributions and often does a better job than the Rayleigh. The *K* distribution, successfully used to describe the statistics of radar sea surface clutter,[4,5] is the product of two components: a rapidly fluctuating Rayleigh (or ''speckle'') distributed component and a chi-distributed component. The physical interpretation is that the Rayleigh-distributed component is from many scatterers that are modulated by large scale (time-varying) structure. The Rayleigh-mixture model[8] is a combination of Rayleigh random variables with each component having its own power. This distribution can be thought of as describing scattering from two (or more) different types of materials in a manner similar to that put forward by Crowther.[7] The PDFs used in this report and their associated cumulative distribution functions are given in Table I.

Fitting the model distributions to the experimental results entailed estimating the parameters of each of the candidate cumulative distribution functions (CDFs) which are the areas under the PDF up to and including a given amplitude. Details on the estimation of parameters for each of the distributions used in this analysis can be found in Abraham.[10] Maximum-likelihood estimates obtained using an iterative algorithm[11] were used for the parameters of the Weibull distribution, method of moments estimates were used for the parameters of the *K* distribution, matching the mean and variance, and for the Rayleigh-mixture model the maximum-likelihood parameter estimates were obtained using the expectation-maximization (EM) algorithm.[10]

Figures 4–7 show graphical examples of the experimen-

TABLE I. Distribution functions.

Function	PDF	CDF
Rayleigh	$(2x/\lambda)e^{-x^2/\lambda}$	$1-e^{-x^2/\lambda}$
Weibull	$\alpha\beta x^{\beta-1}e^{-\alpha x^\beta}$	$1-e^{-\alpha x^\beta}$
K	$4/\sqrt{\alpha}\Gamma(\nu)(x/\sqrt{\alpha})^\nu K_{\nu-1}(2x/\sqrt{\alpha})$	$1-1/\Gamma(\nu)2^{\nu-1}(2x/\sqrt{\alpha})^\nu K_\nu(2x/\sqrt{\alpha})$
Rayleigh mixture	$\Sigma_{i=1}^m \epsilon_i(2x/\lambda_i)e^{-x^2/\lambda_i}$.	$1-\Sigma_{i=1}^m \epsilon_i e^{-x^2/\lambda_i}$

tally observed backscattered amplitude distributions along with the fitted models. In order to highlight the high amplitude or tail portion of the distributions, both the experimental data and model fits are displayed here as probability of false alarm (PFA$=1-$CDF) or probability that the amplitude will be higher or equal to a given amplitude. The non-Rayleigh nature of the reverberation distributions is readily seen in the high grazing angle shellfish, posidonia, and sand data. The distributions tend to Rayleigh at lower grazing angles. A simple explanation for this effect is that as a consequence of the height of the transducer remaining constant, the resolution cell of the sonar will increase at the smaller grazing angles. More patches of seafloor are included in the beam at low grazing angle, which drives the amplitude distribution toward Rayleigh as the central limit becomes valid. Chotiros[1] has observed similar effects of changes in the receiver horizontal beamwidth (resolution cell size) on reverberation statistics with the widebeam results tending toward Rayleigh. A quantitative table of goodness of fit of the observed data to each of the model distributions will be shown in the next section.

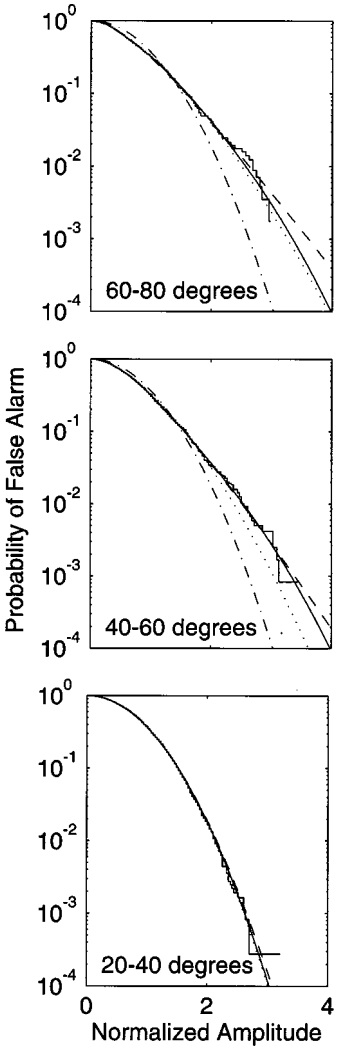


FIG. 4. Scattered amplitude distribution displayed as probability of false alarm for the posidonia-covered seafloor (station RR06) for three different grazing-angle regimes. The dashed–dotted line is the Rayleigh distribution fit to the experimental data, the dotted line is the Weibull fit, the dashed is the K fit, and the solid is the Rayleigh-mixture fit.

II. STATISTICAL RESULTS

To evaluate the flexibility and accuracy of the models in representing the reverberation from the different seafloor types, the Kolmogorov–Smirnoff (KS) test statistic p-values

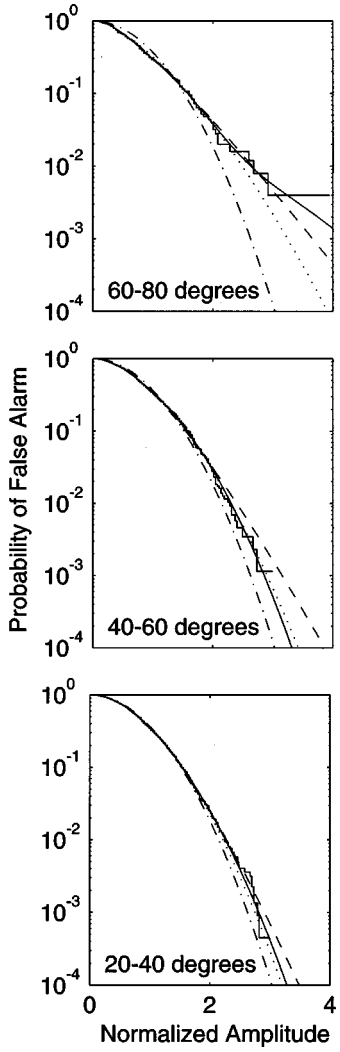


FIG. 5. Probability of false alarm for the shellfish-covered seafloor (station RR13). Line types for the model distribution fits are as in Fig. 3.

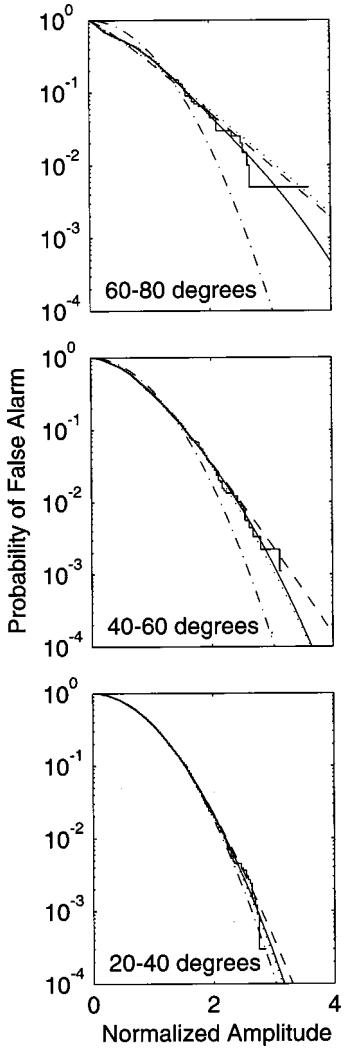


FIG. 6. Probability of false alarm for a sand seafloor (station RR12). Line types for the model distribution fits are as in Fig. 3.

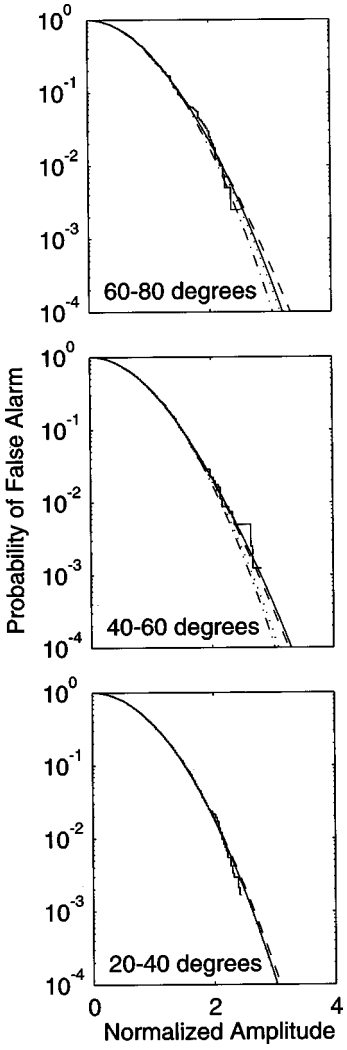


FIG. 7. Probability of false alarm for a mud seafloor (station RR09). Line types for the model distribution fits are as in Fig. 3.

were used to compare real data to model distributions. The KS test statistic[12] is a nonparametric measure that describes how well a theoretical CDF fits the CDF estimated from observed data. If $F(x)$ is a theoretical CDF and $S_n(x)$ is the sample CDF formed from n samples of observed data, the KS test statistic is the maximum absolute difference

$$t = \max_{-\infty<x<\infty} |F(x) - S_n(x)|. \tag{3}$$

When n is large enough, t becomes independent of the distribution $F(x)$ under the null hypothesis that the data are distributed according to $F(x)$. To provide a concrete and familiar scale for comparing the results of the KS test, the p-value of t is utilized. The p-value is the probability of observing a value greater than t under the null hypothesis; that is, if t is the observed KS test statistic and T is the random variable formed by Eq. (3) using the random variable X distributed according to $F(x)$, then the p-value is

$$p = \Pr\{T \geqslant t\}. \tag{4}$$

When the p-value is small, we are led to believe that observing a more extreme value for the KS test statistic is not likely under the null hypothesis, and subsequently decide that the data do not follow the hypothesized CDF. If, however, the maximum absolute difference between the hypothesized CDF and the estimated CDF is small, then the p-value is near one. Thus, we use the p-value of the KS test statistic to distinguish between good and bad fitting of the observed data by our model CDFs. It should be noted that Kolmogorov's theorem does not hold when parameters estimated from the data are used to form the theoretical CDF.[12] Nevertheless, the p-value formed from the KS statistic gives a sufficient measure of the goodness of fit between the model and estimated CDFs.

Results are presented for each of 14 sites and are grouped in terms of grazing angle. Table II shows bottom type, average scattering level, and KS p-values for the 60 deg to 80 deg grazing angle, Table III for 40 deg to 60 deg grazing angle, Table IV for 20 deg to 40 deg grazing angle, and Table V for 10 deg to 20 deg grazing angle. Any p-values above 0.8 are shown in bold to highlight the best fits to observed data. Table VI gives the parameter values estimated in fitting the model distributions to the experimental data.

Quantitative agreement is seen with the qualitative assessment of the last section. The highest grazing angles (40

TABLE II. Results for selected sites and for 60 deg–80 deg grazing angle.

Station	Bottom type	Level (dB)	Rayleigh	Weibull	Mixture	*K*
RR04	sand/shell	−19.2	0.215	0.590	0.763	0.362
RR05	coarse sand	−23.1	0.745	**0.998**	**0.980**	**0.980**
RR06	posidonia	−20.0	1.04×10^{-9}	0.667	**0.983**	**0.933**
RR07	mud	−21.7	0.161	0.667	**0.834**	**0.915**
RR08	mud	−19.4	0.6309	0.758	**0.963**	**0.913**
RR09	mud	−23.5	**0.902**	**0.986**	**0.998**	**0.999**
RR10	sand/shell	−20.4	**0.818**	**0.843**	**0.818**	0.747
RR11	sand/shell	23.0	**0.992**	**0.999**	**0.997**	**0.999**
RR12	med. sand	−28.9	3.42×10^{-12}	0.543	**0.977**	0.123
RR13	shells	−13.8	1.57×10^{-4}	0.440	**0.850**	**0.802**
RR14	sand/shell	−23.2	1.44×10^{-15}	0.090	**0.991**	0.098
RR15	sand/shell	−22.6	7.95×10^{-13}	**0.891**	0.739	0.451
RR16	mud	−25.7	0.026	0.473	**0.999**	**0.875**
RR17	mud	−22.1	0.071	0.492	**0.865**	**0.816**

deg–80 deg) are usually not well described by the Rayleigh distribution. The ensonified area at any given time at high grazing angles is small, so that only a few patches of differing scattering strength are included in the beam. In this case, the central-limit theorem becomes invalid resulting in non-Rayleigh reverberation. In most cases, the Weibull, Rayleigh mixture, and *K* distributions all match the observed distribution better than a standard Rayleigh. This results from the fact that the these higher-order model distributions have more parameters to ''tweak'' (see Table I) when trying to fit the distributions estimated from the observed data. The Rayleigh describes the amplitude distribution well at the lowest grazing angles but, as the other candidate PDFs contain the Rayleigh as a subset, they also perform adequately. The Rayleigh mixture and *K* distributional models are the most robust of the distributional models in fitting the observed data, working over the entire range of grazing angles (or resolution cell size). For the non-Rayleigh distributed data, a closer look at the estimated Rayleigh mixture-model parameters shows that in most cases two components are dominating the description, even when three components were used in the model fit. This result strongly suggests that most of the sites studied consist of a bimodal distribution of scattering patches, i.e., only two patches of differing scattering strength.

While the limited amount of data necessitated the rather coarse grouping of backscatter data into 20 deg grazing-angle bins in order to make comparison possible between the experimental and model amplitude distributions, it is also desirable to know the effect of the ensonified area on the amplitude statistics on a finer scale. The variance of the intensity fluctuation, or scintillation index, which is given by

$$\sigma_I^2=\frac{\langle(I-\langle I\rangle)^2\rangle}{\langle I\rangle^2}, \tag{5}$$

can be estimated with fewer samples than required for PDFs so that the coarse groupings are not needed, providing a more precise indication of the size of the resolution cell at which the amplitude statistics depart from Rayleigh and also how quickly this departure takes place. Estimates of scintillation index for four sites with non-Rayleigh statistics are shown in the top graph of Fig. 8. The tendency away from a Rayleigh (Rayleigh-distributed amplitudes result in a scintillation index of 1) is evident for these four sites as the ensonified area decreases. The point at which the statistics depart from Rayleigh is a function of the patch size of the bottom inhomogeneities. The scintillation index for four sites that display Rayleigh statistics for all ensonified areas are also shown in the bottom graph of Fig. 8.

Tables II–V also display the mean scattered level (averaged over the 20 deg grazing-angle bin) for each station.

TABLE III. Results for selected sites and for 40 deg–60 deg grazing angle.

Station	Bottom type	Level (dB)	Rayleigh	Weibull	Mixture	*K*
RR04	sand/shell	−21.8	0.652	**0.929**	**0.961**	**0.962**
RR05	coarse sand	−26.4	**0.864**	**0.905**	**0.864**	0.797
RR06	posidonia	−21.0	1.28×10^{-7}	0.367	**0.989**	**0.996**
RR07	mud	−21.9	0.487	0.691	0.784	**0.874**
RR08	mud	−23.5	0.762	**0.877**	**0.963**	**0.975**
RR09	mud	−23.8	**0.827**	**0.979**	**0.999**	**0.999**
RR10	sand/shell	−21.2	**0.978**	**0.853**	**0.978**	**0.984**
RR11	sand/shell	−24.4	**0.953**	**0.941**	**0.953**	**0.969**
RR12	medium sand	−28.2	4.25×10^{-4}	0.716	**0.881**	0.615
RR13	shells	−13.5	0.002	0.508	**0.963**	**0.866**
RR14	sand/shell	−26.1	1.36×10^{-7}	0.515	**0.968**	0.179
RR15	sand/shell	−22.6	0.151	**0.967**	**0.988**	**0.880**
RR16	mud	−25.4	0.246	0.785	**0.951**	**0.965**
RR17	mud	24.6	0.118	**0.947**	**0.878**	**0.931**

TABLE IV. Results for selected sites and for 20 deg–40 deg grazing angle.

Station	Bottom type	Level (dB)	Rayleigh	Weibull	Mixture	K
RR04	sand/shell	−22.6	**0.920**	**0.967**	**0.919**	**0.811**
RR05	coarse sand	−26.6	0.112	0.799	**0.918**	**0.973**
RR06	posidonia	−21.8	**0.955**	**0.982**	**0.955**	0.794
RR07	mud	−27.1	0.746	**0.942**	**0.941**	**0.840**
RR08	mud	−27.0	0.453	0.799	0.725	**0.878**
RR09	mud	−27.0	**0.941**	**0.886**	**0.941**	**0.974**
RR10	sand/shell	−23.7	**0.803**	**0.983**	**0.952**	**0.980**
RR11	sand/shell	−26.1	**0.812**	**0.871**	**0.986**	**0.885**
RR12	medium sand	−29.5	0.070	**0.927**	0.616	0.749
RR13	shells	−16.2	0.014	0.328	**0.888**	**0.863**
RR14	sand/shell	−26.7	0.005	0.699	0.746	0.637
RR15	sand/shell	−22.8	0.022	**0.903**	**0.847**	**0.900**
RR16	mud	−26.2	**0.888**	**0.924**	**0.888**	**0.877**
RR17	mud	−27.7	0.288	**0.836**	**0.962**	**0.910**

Although the bottom types were not always separable using the mean scattering level, in general the muds displayed the lowest scattering levels, the sands gave medium levels, and the shellfish- and posidonia-covered bottoms showed the highest scattering levels. The many exceptions to this general rule on mean scattering strength suggests that scattering strength is not an accurate descriptor of sediment type. Any seafloor-classification scheme based on mean scattering strength should include other parameters affecting the scattered level such as the interface roughness spectrum or distributions of discrete scatterers on the seafloor.

III. DISCUSSION

In the previous section, it was observed that the K distribution fit the amplitude statistics of the reverberation data well for most of the sites considered and that the Rayleigh-mixture distribution fit the data well for all of them. To explain these results, we look to the ''compound'' representation of the K random variable.[5] This representation describes the random variable as the product of two independent random variables, an exponential or speckle variable and a variable which modulates this speckle, leading to the PDF

$$f(x)=\int_{\lambda=0}^{\infty} f_{\lambda}(\lambda) f(x|\lambda)\, d\lambda, \tag{6}$$

where $f(x|\lambda)$ is a Rayleigh PDF with power λ and $f_{\lambda}(\lambda)$ describes a slow fluctuation of the power and is chi-squared (more precisely, a gamma random variable is used to allow noninteger values for the degrees of freedom parameter) distributed with ν degrees of freedom. However, if the probability distribution for λ is chosen appropriately, one can obtain Crowther's model[7] for a seabed composed of two different types of patches distributed spatially according to a Markov process. Furthermore, allowing λ to be chi-squared distributed with n_{λ} degrees of freedom and x given λ to be chi distributed (i.e., the square root of a chi-squared distributed random variable) with n_x degrees of freedom and scale parameter λ, one obtains McDaniel's non-Rayleigh reverberation model[6] for backscatter from a rough surface where the finite size of the resolution cell is considered in both range and bearing.

As the K distribution and Crowther's model are based on physical processes, and McDaniel's model accounts for the effects of the transmit waveform and beamforming, it is important to notice that they are all described by one unified model, and we look to relate this model to the Rayleigh mixture model. First, we note that when the resolution cell size is much smaller than the patch size, Crowther's model simplifies to a two-component Rayleigh mixture. In this case, the random variable λ has essentially been quantized to two discrete values representing the backscatter from the two different patch types. Suppose that λ is quantized to the p values $\lambda_1,\lambda_2,\ldots,\lambda_p$ with corresponding probabilities $\pi_1,\pi_2,\ldots,\pi_p$. Then, the PDF of Eq. (6) results in the mixture distribution

$$f_m(x)=\sum_{i=1}^{p} \pi_i f(x|\lambda_i). \tag{7}$$

This result is also arrived at by approximating the integral of Eq. (6) by a finite summation. Now, if $f(x|\lambda)$ is Rayleigh, as for the K distribution and Crowther's model, Eq. (7) describes a Rayleigh mixture.

TABLE V. Results for selected sites and for 10 deg–20 deg grazing angle.

Station	Bottom type	Level (dB)	Rayleigh	Weibull	Mixture	K
RR05	coarse sand	−28.9	**0.922**	**0.871**	0.797	0.782
RR07	mud	−30.3	**0.897**	**0.995**	**0.897**	0.744
RR08	mud	−28.4	0.671	**0.958**	**0.909**	0.759
RR09	mud	−30.1	**0.995**	**0.999**	**0.995**	**0.875**
RR10	sand/shell	−24.7	**0.970**	**0.997**	**0.999**	**0.998**
RR12	medium sand	−29.0	0.204	**0.915**	**0.902**	**0.912**
RR13	shells	−21.1	**0.939**	**0.999**	**0.968**	**0.992**

TABLE VI. Parameter estimates used in fitting the model CDFs to experimental data.

Station	Grazing angle (deg)	Weibull (α,β)	Mixture (ϵ_i,λ_i)	K (α,ν)
RR04	60–80	(1.028, 1.873)	(0.093 0.448 0.459, 0.178 1.084 1.084)	(0.097, 10.35)
RR04	40–60	(1.009, 1.955)	(0.369 0.262 0.369, 1.101 0.717 1.101)	(0.033, 30.20)
RR04	20–40	(0.998, 2.010)	(0.334 0.333 0.333, 0.997 0.994 1.009)	(0.010, 100.0)
RR05	60–80	(1.031, 1.862)	(0.353 0.375 0.272, 1.172 1.172 0.542)	(0.108, 9.224)
RR05	40–60	(0.978, 2.110)	(0.333 0.333 0.334, 1.000 1.000 1.000)	(0.010, 100.0)
RR05	20–40	(1.024, 1.891)	(0.325 0.326 0.349, 0.996 0.653 1.328)	(0.092, 10.86)
RR05	10–20	(0.999, 2.004)	(0.336 0.334 0.330, 0.928 0.938 1.136)	(0.010, 100.0)
RR06	60–80	(1.138, 1.514)	(0.273 0.491 0.236, 0.199 0.920 2.093)	(0.557, 1.797)
RR06	40–60	(1.092, 1.656)	(0.228 0.588 0.185, 2.049 0.823 0.269)	(0.377, 2.654)
RR06	20–40	(0.997, 2.016)	(0.333 0.333 0.334, 1.000 1.000 1.000)	(0.010, 100.0)
RR07	60–80	(1.044, 1.810)	(0.300 0.298 0.402, 1.235 0.447 1.235)	(0.164, 6.086)
RR07	40–60	(1.018, 1.918)	(0.316 0.338 0.346, 0.754 1.112 1.115)	(0.042, 24.03)
RR07	20–40	(1.017, 1.923)	(0.332 0.333 0.335, 0.804 0.995 1.198)	(0.050, 20.09)
RR07	10–20	(0.996, 2.021)	(0.334 0.333 0.333, 1.000 1.000 1.000)	(0.010, 100.0)
RR08	60–80	(1.008, 1.964)	(0.419 0.422 0.159, 1.101 1.101 0.463)	(0.050, 20.21)
RR08	40–60	(1.018, 1.915)	(0.362 0.280 0.358, 1.135 0.653 1.135)	(0.061, 16.50)
RR08	20–40	(1.005, 1.975)	(0.336 0.324 0.340, 0.858 1.008 1.131)	(0.021, 48.07)
RR08	10–20	(1.006, 1.971)	(0.330 0.336 0.334, 0.858 1.073 1.066)	(0.024, 41.90)
RR09	60–80	(1.018, 1.917)	(0.341 0.291 0.368, 1.152 0.630 1.152)	(0.062, 16.01)
RR09	40–60	(1.009, 1.960)	(0.324 0.330 0.346, 1.346 0.863 0.806)	(0.046, 21.65)
RR09	20–40	(0.998, 2.009)	(0.333 0.334 0.333, 1.000 1.000 1.000)	(0.010, 100.0)
RR09	10–20	(0.999, 2.004)	(0.333 0.334 0.333, 1.000 0.999 1.001)	(0.010, 100.0)
RR10	60–80	(0.999, 2.006)	(0.335 0.332 0.333, 1.000 1.000 1.000)	(0.010, 100.0)
RR10	40–60	(1.009, 1.955)	(0.334 0.332 0.334, 1.000 1.000 1.000)	(0.010, 100.0)
RR10	20–40	(1.004, 1.979)	(0.325 0.338 0.337, 0.836 1.079 1.079)	(0.022, 45.89)
RR10	10–20	(1.005, 1.978)	(0.335 0.336 0.329, 1.069 1.072 0.856)	(0.016, 61.68)
RR11	60–80	(1.009, 1.958)	(0.395 0.394 0.211, 0.901 0.901 1.370)	(0.027, 37.49)
RR11	40–60	(0.998, 2.009)	(0.333 0.333 0.334, 1.000 1.000 1.000)	(0.010, 100.0)
RR11	20–40	(1.013, 1.941)	(0.341 0.337 0.322, 1.106 1.103 0.780)	(0.037, 27.23)
RR12	60–80	(1.298, 1.120)	(0.515 0.177 0.308, 0.997 2.696 0.034)	(1.303, 0.769)
RR12	40–60	(1.087, 1.658)	(0.546 0.399 0.055, 0.652 1.606 0.040)	(0.337, 2.970)
RR12	20–40	(1.015, 1.930)	(0.289 0.355 0.356, 0.680 1.130 1.130)	(0.058, 17.29)
RR12	10–20	(1.011, 1.951)	(0.332 0.334 0.334, 0.730 1.113 1.143)	(0.041, 24.18)
RR13	60–80	(1.139, 1.532)	(0.579 0.026 0.395, 1.283 5.427 0.290)	(0.573, 1.747)
RR13	40–60	(1.060, 1.754)	(0.286 0.372 0.342, 0.348 1.300 1.218)	(0.242, 4.128)
RR13	20–40	(1.026, 1.884)	(0.313 0.324 0.363, 0.566 1.102 1284)	(0.106, 9.470)
RR13	10–20	(1.004, 1.983)	(0.334 0.333 0.333, 0.941 0.970 1.089)	(0.012, 80.01)
RR14	60–80	(1.251, 1.259)	(0.355 0.369 0.276, 0.090 2.074 0.738)	(1.114, 0.899)
RR14	40–60	(1.074, 1.700)	(0.133 0.442 0.425, 0.129 0.869 1.410)	(0.291, 3437)
RR14	20–40	(1.051, 1.788)	(0.388 0.353 0.259, 0.672 0.814 1.745)	(0.194, 5.159)
RR15	60–80	(1.274, 1.187)	(0.118 0.464 0.418, 0.010 1.892 0.292)	(1.184, 0.846)
RR15	40–60	(1.026, 1.881)	(0.152 0.419 0.429, 0.390 1.109 1.109)	(0.094, 10.62)
RR15	20–40	(1.024, 1.892)	(0.321 0.348 0.331, 0.993 0.687 1.336)	(0.084, 11.86)
RR16	60–80	(1.072, 1.726)	(0.358 0.362 0.280, 0.624 0.624 1.965)	(0.298, 3.363)
RR16	40–60	(1.031, 1.860)	(0.373 0.254 0.373, 1.222 0.550 1.085)	(0.105, 9.515)
RR16	20–40	(1.003, 1.988)	(0.333 0.333 0.334, 1.004 1.001 0.995)	(0.010, 100.0)
RR17	60–80	(1.043, 1.826)	(0.346 0.231 0.423, 0.715 1.951 0.715)	(0.187, 5.350)
RR17	40–60	(1.031, 1.865)	(0.286 0.427 0.287, 0.855 1.366 0.600)	(0.123, 8.137)
RR17	20–40	(1.016, 1.925)	(0.365 0.272 0.363, 0.851 1.397 0.852)	(0.060, 16.74)

As an approximation to Eq. (6), the mixture proportions $\{\pi_i\}$ and powers $\{\lambda_i\}$, could perhaps be chosen to minimize the mean squared error between $f(x)$ and $f_m(x)$ or, in line with the idea of quantizing the random variable λ, be chosen to minimize the mean squared quantization error. However, as observed in this paper, obtaining the maximum-likelihood mixture parameter estimates through the EM algorithm allows accurate fitting of observed data. This method also has a greater degree of flexibility than the K distribution or the models of Crowther and McDaniel because the mixture parameters [which approximate $f_\lambda(\lambda)$] are estimated directly from the data as opposed to obtaining $f_\lambda(\lambda)$ through the restricted form of a particular model. However, this does imply that if the assumptions made in deriving the K distribution or Crowther's model are correct, then their models may describe the data better than a Rayleigh mixture. When it is not known how accurate these assumptions are, or when estimation of the model parameters is difficult as for Crowther's model, it is prudent to use the Rayleigh-mixture model.

IV. CONCLUSIONS

The statistics of acoustic seafloor reverberation have been investigated for data taken at 80 kHz from 14 shallow-

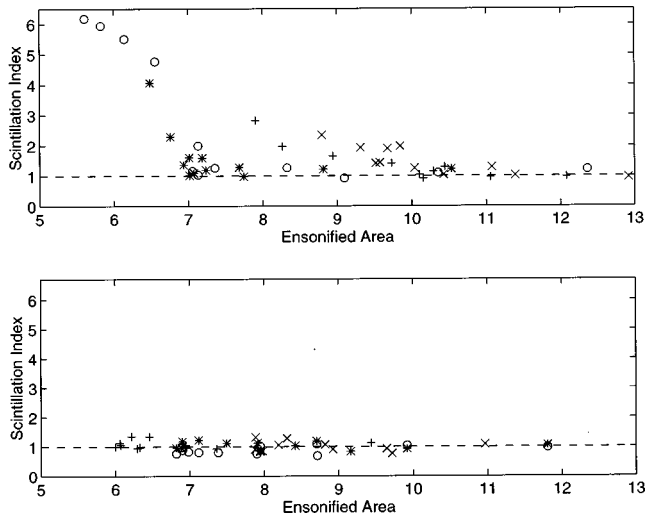


FIG. 8. Scintillation index as a function of ensonified area (m^2); top: stations with non-Rayleigh amplitude distributions, RR16 (+), RR06(x), RR12(o), and RR13(*); bottom: stations with Rayleigh distributions, RR07 (+), RR11(x), RR10(o), and RR09(*).

water sites around Sardinia and Sicily. The bottom characteristics of many of the sites were frequently neither spatially nor temporally homogeneous. Rayleigh PDFs do not accurately represent the statistical properties of the measured reverberation from the study sites, especially at high grazing angles. This result may be attributed to the number of patches of differing scatterer density or strength included in the sonar resolution cell at a given grazing angle. In general, for seafloor environments that are inhomogeneous (patchy), the smaller the resolution cell size the more non-Rayleigh the reverberation. This relationship will depend on the relative sizes of the patches and the ensonified area.

The Rayleigh mixture or K distributions are the closest fit to the observed distributions over all bottom types at high grazing angles and, as these contain the Rayleigh distribution as a submember, they also perform well at low angles. Both these distributions, along with the models of Crowther[7] and McDaniel,[6] can be described by a general integral expression which represents the PDF of the product of two random variables. When the accuracy of the assumptions made in the derivation of the K distribution is not known, the Rayleigh-mixture model may be a better alternative as the parameters are estimated directly from the data as opposed to obtaining them through the restricted form of a model. A three-component Rayleigh-mixture model was used in this analysis, but as two components were usually found to dominate the distribution description, the number of components used in the mixture model could be reduced to two and still give satisfactory results. Additionally, the model could also be expanded to include more components if needed, which would allow it to fit almost any distribution.

The mean scattering levels measured in this study were not found to correlate well with bottom type, which suggests that scattering strength is not an accurate descriptor of sediment type. Any seafloor-classification scheme based on mean scattering strength should include other parameters affecting the scattered level such as the interface roughness spectrum, volume inhomogeneity spectrum, or distributions of discrete scatterers on and in the seafloor.

ACKNOWLEDGMENTS

The authors wish to acknowledge Tuncay Akal for initiating this high-frequency reverberation study and Piero Guerrini for setting up the acoustic and data acquisition systems. The helpful comments and suggestions from three reviewers are also gratefully acknowledged.

[1] N. P. Chotiros, H. Boehme, T. G. Goldsberry, S. P. Pitt, A. L. Garcia, and R. A. Altenburg, ''Acoustic backscattering at low grazing angles from the ocean bottom. II. Statistical characteristics of bottom backscatter at a shallow water site,'' J. Acoust. Soc. Am. **77**, 975–982 (1985).
[2] S. Stanic and E. G. Kennedy, ''Reverberation fluctuations from a smooth seafloor,'' IEEE J. Ocean Eng. **18**, 95–99 (1993).
[3] S. Stanic and E. G. Kennedy, ''Fluctuations of high-frequency shallow-water seafloor reverberation,'' J. Acoust. Soc. Am. **91**, 1967–1973 (1992).
[4] E. Jakeman and P. N. Pusey, ''A model for non-Rayleigh sea echo,'' IEEE Trans. Antennas Propag. **24**, 806–814 (1976).
[5] K. D. Ward, ''Compound representation of high resolution sea clutter,'' Electron. Lett. **17**, 561–563 (1981).
[6] S. T. McDaniel, ''Seafloor reverberation fluctuations,'' J. Acoust. Soc. Am. **88**, 1530–1535 (1990).
[7] P. A. Crowther, ''Fluctuation statistics of sea-bed acoustic backscatter,'' in *Bottom Interacting Ocean Acoustics*, edited by W. A. Kuperman and F. B. Jensen (Plenum, New York, 1980), pp. 609–622.
[8] W. K. Stewart, D. Chu, S. Malik, S. Lerner, and H. Singh, ''Quantitative seafloor characterization using a bathymetric sidescan sonar,'' IEEE J. Ocean Eng. **19**, 599–610 (1994).
[9] R. J. Urick, *Principles of Underwater Sound*, 3rd ed. (McGraw-Hill, New York, 1983).
[10] D. A. Abraham, ''Modeling non-Rayleigh reverberation,'' SACLANT Undersea Research Centre Report SR-266, June 1997.
[11] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, Vol. 1, 2nd ed. (Wiley, New York, 1994).
[12] M. Fisz, *Probability Theory and Mathematical Statistics*, 4th ed. (Macmillan, New York, 1978).

Single sensor source tracking and environmental inversion

Jose Unpingco, William A. Kuperman, and William S. Hodgkiss
Marine Physical Laboratory, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093-0701

Robert Hecht-Nielsen
Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, California 92093-0407

(Received 27 April 1998; revised 14 May 1999; accepted 21 May 1999)

A method is introduced for using a single underwater sensor and knowledge of a shallow water environment to track a moving, multi-frequency, acoustic source. The motion of the source creates a synthetic aperture that is adaptively matched against an acoustic propagation model in a manner similar to matched field processing. However, in this case, many more unknowns must simultaneously be determined. Results are presented for tracking using experimental data subject to unknown source speed, heading, level, and initial position. A new parallel adaptive method manages these simultaneous unknowns. Perturbations in the sound speed based on empirical orthogonal functions are used to extend the approach to operating when there also is uncertainty in this parameter. The performance of the tracking process with hydrophone data for various signal-to-noise cases is considered. © *1999 Acoustical Society of America.*
[S0001-4966(99)01209-6]

PACS numbers: 43.30.Wi, 43.30.Pc, 43.30.Zk, 43.60.Gk [DLB]

INTRODUCTION

This paper considers the tracking of a moving acoustic source in a shallow water ocean environment using the received pressure field at one sensor and historical knowledge of the ocean environment subject to simultaneously unknown source initial position, heading, speed, and level. Additionally, it is shown that in the process of tracking the source, one also can estimate the water column sound speed.

The problem is illustrated in Fig. 1, which shows that as an acoustic source moves in a straight line at a constant speed, its radiated pressure field (shown as concentric circles) is received at the sensor as a function of the surrounding environmental conditions (e.g., bottom geoacoustic characteristics and sound speed structure). The problem is to determine the initial position of the source (x_i, y_i), its heading θ, and its speed using the recorded pressure field at the hydrophone and historical knowledge of the acoustic environment.

The most provocative facet of this problem is the use of a single sensor. While research in tracking problems has focused almost exclusively on sensor arrays, our motivation is to demonstrate, in the limiting case of one sensor, how much hardware can be replaced with computing effort through using prior knowledge of the acoustic environment. Over the past 20 years, a generalization of conventional beamforming methods known as matched field processing (MFP)[1,2] has been developed. MFP incorporates prior environmental knowledge by constructing a mathematical model of propagation based on this knowledge. This model is used to predict the sensor data, which is then compared against actual data under competing hypotheses of source position. The best match between model and data is used as an estimate of source position.

As in MFP, we also employ a propagation model to compare predicted with observed sensor data. However, our point of departure is the lack of a sensor array typically associated with this process.[3] Additionally, we consider only the magnitude of the received pressure field at several frequencies and thus lack the relative element-to-element phase information typically used in MFP. Whereas most MFP discussions consider point sources and do not integrate over the source motion, we consider a moving source that creates a synthetic aperture. Since there can be many source trajectories originating at the same point with different headings and speeds, the combinatorics make for a vast search space of higher dimension than typically considered.[4,5]

Another distinguishing facet of this work is the demonstration of source tracking with real experimental data. The synthetic aperture single sensor tracking problem has been considered in simulation by Collins.[6,7] Subsequent experimental work demonstrated single sensor tracking using directly measured replica fields.[8,9] Our approach here is quite different due to the manner in which the parameter search space is explored.

We also show that the tracking method we develop can be extended to estimating sound speed structure as an integral part of source tracking. This reflects one potential application of this work where a single sensor takes advantage of transiting surface ships to monitor changes in the sound speed structure of a region.

I. PROBLEM DEFINITION

Figure 2 shows the region of interest just west of Point Loma (San Diego, CA). The diamond indicates the receiving sensor position. The contours show the 53-Hz transmission loss (dB). What is important about this figure is the structure of the contours, not their absolute values. As an acoustic source moves through this region relative to the sensor, its

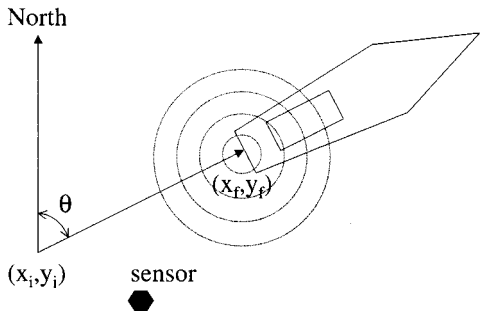


FIG. 1. Single sensor source tracking.

recorded pressure field magnitude at 53 Hz represents a slice across these contours (see Fig. 3). The problem at hand can be stated as finding the spatial track corresponding to a given slice across these contours. Figure 2 shows that along certain tracks, namely those in a radial direction from the sensor, the slice of the map obtained has rich structure. On the other hand, tracks starting away from the sensor and moving east–west cut through relatively flat areas, providing little structural information. This fact alone establishes the ill-posed nature of the problem since, in some cases, the origin of the slice cannot be established uniquely. This also shows how the natural complexity in the acoustic environment supports localization accuracy as a function of different slices across the region.

Figure 3 shows that due to the variation in environmental parameters (e.g., bathymetry), the transmission loss pressure field magnitude (TLPFM) record is a relatively unique signature for each separate source track. This is because, in a shallow water environment, the factors that influence propagation between source and sensor vary spatially along different tracks. Our approach is to exploit this variation by matching modeled pressure field magnitude records against the sensor data from an unknown source track. This is similar to the approach taken in MFP where, instead of processing data from a sensor array, we process data from a single sensor along a synthetic aperture created by the source motion.

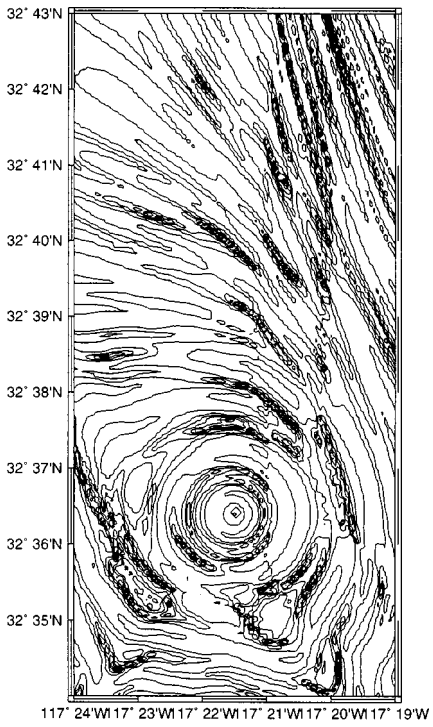


FIG. 2. Transmission loss contours (dB) at 53 Hz west of Point Loma (San Diego, CA). The receiving sensor is indicated by the diamond (◇). Contours are 4 dB apart.

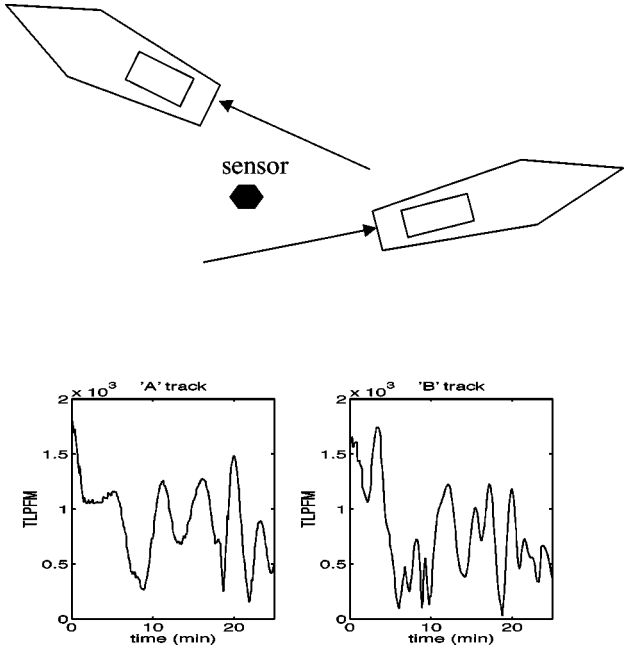


FIG. 3. TLPFM variation with source track.

A. Assumptions

As part of the definition of the problem, we make the following assumptions:

- Source trajectories are distributed in a fixed region.
- Sources move in straight lines for at least 25 min.
- Sources do not change speed or source level for at least 25 min.
- Source speed is bounded between 2 m/s and 4 m/s.
- Sources radiate signature components at multiple discrete frequencies of 53, 69, 85, and 101 Hz.
- Source depth is fixed at 47 m.
- Moving sources are modeled as sequential point sources on a rectangular grid.

Although some of these assumptions are specific to the source tow data which will be analyzed, the framework easily can accommodate the radiated signature components from transiting surface ships.

B. Terminology

The sequence of N sampled points $\{(x_{i,k},y_{i,k})\}_{i=1}^{N}$ of the kth straight trajectory at speed s_k is defined as a ''track vector,''

$$\mathbf{b}_k=[x_{1,k},\ y_{1,k},\ \theta_k,\ s_k],$$

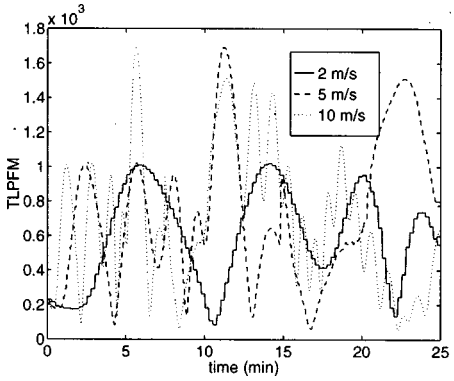


FIG. 4. TLPFM variation with source speed.

where θ_k is the heading of the trajectory. Note that the *x*-direction is along longitude and the *y*-direction is along latitude.

The corresponding sequence of the sampled transmission loss pressure field magnitude at a given frequency (generated by the propagation model)

$$\mathbf{t}_k^f = [|t_f(x_{1,k}, y_{1,k})|,\ |t_f(x_{2,k}, y_{2,k})|, \ldots, |t_f(x_{N,k}, y_{N,k})|]$$

is defined as a ''transmission loss pressure field magnitude'' (TLPFM) vector. Note that here we normalize the pressure field by its value 1 m away from the source and thus all values of $|t(x,y)|$ are bounded by [0,1].

Similarly, the sequence of the sampled received pressure field magnitude at a given frequency (which includes the source level)

$$\mathbf{p}_k^f = [|p_f(x_{1,k}, y_{1,k})|,\ |p_f(x_{2,k}, y_{2,k})|, \ldots, |p_f(x_{N,k}, y_{N,k})|]$$

is defined as a ''pressure field magnitude'' (PFM) vector.

Last, the ''exemplar set'' is the set of pairs of track vectors and corresponding transmission loss pressure field magnitude vectors across a set of discrete frequencies (in this case, 53, 69, 85, and 101 Hz)

$$\mathbf{B} = \{(\mathbf{b}_k,\ \mathbf{t}_k^{53},\ \mathbf{t}_k^{69},\ \mathbf{t}_k^{85},\ \mathbf{t}_k^{101})\}_{k=1}^{M}.$$

C. Source speed uncertainty

Figure 4 shows the effect of source speed variation on the TLPFM during the sampling window of 25 min. This plot shows that there can be wide structural differences in TLPFM vectors that differ only in source speed. Note that the slowest source moving at 2 m/s travels exactly one-fifth the distance of the faster source moving at 10 m/s. For this reason, the TLPFM vector for the 2-m/s source is identical to that of the first $25/5 = 5$ min of the 10-m/s source. Thus variation in source speed corresponds to a compression/expansion of the TLPFM vector.

D. Source level uncertainty

Uncertainty in source level corresponds to an unknown offset between corresponding TLPFM and (noise-free) PFM vectors when expressed in dB (in the absence of modeling errors). Source level uncertainty can be eliminated by subtracting the averaged received level ($\overline{\mathbf{RL}}$) over the length of the data window from the received level (**RL**) and retrieving the magnitude from this new sequence as

$$\mathbf{p}' \equiv 10^{(\mathbf{RL}/20 - \overline{\mathbf{RL}}/20)},$$

where $\mathbf{RL} = 20 \log \mathbf{p}$.

This adjustment is performed for all sensor data (**p**) and modeled transmission loss (**t**) and we make no more explicit reference to it. Note that lack of source level knowledge prohibits estimating the range of the source based on absolute transmission loss.

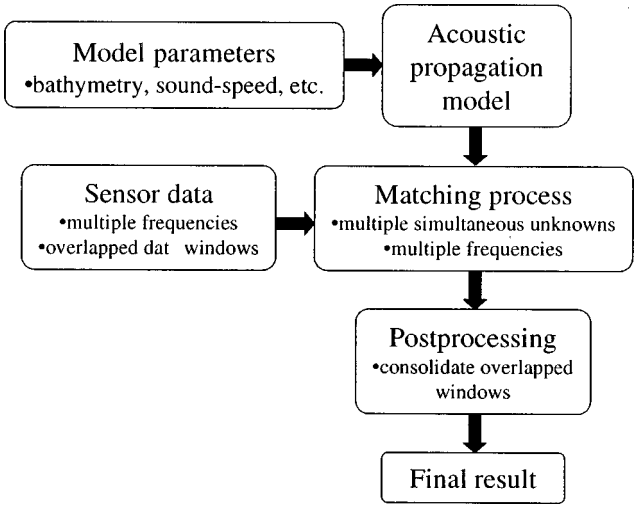


FIG. 5. Block diagram illustrating the solution to the tracking problem.

II. SOLUTION METHOD AND RESULTS

The block diagram in Fig. 5 illustrates the solution to the tracking problem which is discussed a block at a time in this section. Our inputs are the sensor data and historical knowledge of the parameters that influence acoustic propagation (e.g., bathymetry, bottom geoacoustic characteristics, and sound speed). We use prior knowledge of the environment to build a model of propagation in the region. Through an adaptive process, we produce the best match between sensor and modeled data over a track for each overlapped data window. The next step consolidates the individual track estimates from overlapped data windows in a least squares sense to estimate the final coordinates of the overall track.

A. Sensor data

The data come from the SWellEx-3 experiment conducted in the summer of 1994 off the coast of San Diego, California.[10] The event processed included a towed source at a known depth (47 m) emitting a set of discrete frequencies (53, 69, 85, 101 Hz) with the corresponding GPS source track shown in Fig. 6. Specifically, we use data from July 22, 1994 covering the period 21:53–22:26 GMT of Event G:D. The sampling frequency was 1500 Hz and the data were processed using 8192 point FFT's with 50% overlap.

In Fig. 6, the inner rectangle encloses approximately 27 km^2 wherein source tracks are restricted to start, although they can end outside this rectangle. The underlying contours

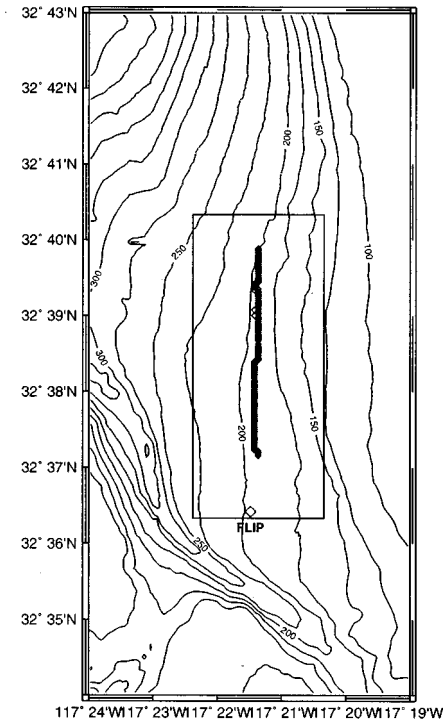


FIG. 6. The GPS source track, receiving sensor ($\diamond$), and underlying bathymetry (m).

indicate the bathymetry which was available on a rectangular 2 arcsec by 2 arcsec (approximately 60 m by 60 m) grid.[11] The geoacoustic parameters also were available.[11] The sound speed profile was measured on July 22, 1994 at 22:25 GMT at 32°39.86′ N and 117°21.31′ W and is shown in Fig. 7. The sensor used for this processing was 82.1 m deep on a vertical array deployed by the R/P FLIP at 32°36.4′ N and 117°21.5′ W and is indicated in Fig. 6 by a diamond ''$\diamond$.''

The source was moving at approximately 5 knots and the observation interval was 32.9 min, which makes the total track length approximately 5 km. The signal levels for the 53-, 69-, 85-, and 101-Hz signals and corresponding noise levels 1 Hz higher in frequency are shown in Fig. 8. The source level was approximately 150 dB *re*: 1 μPa. The SNR for each of these frequencies (computed across the observation interval) was 15, 13, 19, and 17 dB, respectively.

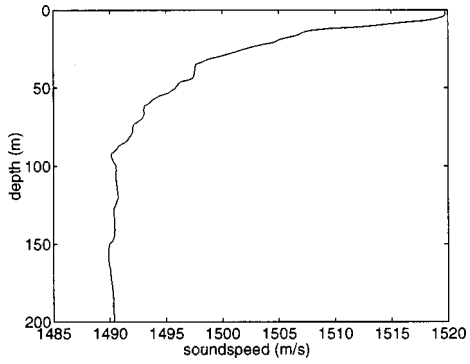


FIG. 7. The sound speed profile measured on July 22, 1994 at 22:25 GMT at 32°39.86′ N and 117°21.31′ W.

B. Acoustic propagation model

KRAKEN[12] was used to compute the waveguide mode functions and horizontal wave numbers on a 30 m by 30 m grid for each frequency (53, 69, 85, 101 Hz). This grid density does not undersample the TLPFM vectors at any of the frequencies of interest. The adiabatic normal mode solution is used for its computational simplicity and physical interpretation, although there is no rigid dependence on this particular solution. Note that the data track is along a region of essentially flat bathymetry so the sensor data is well-modeled by the adiabatic normal mode solution.

The sound speed profile measured nearest in space and time to the source tow experiment is used as the sound speed profile over the entire region (see Fig. 7). In short, the pressure field magnitude at the sensor from every point on the grid is calculated using the same sound speed profile and geoacoustic parameters with only the water depth varying as a function of location. Thus the geographical complexity in the pressure field is strictly through the bathymetry.

Figure 9 shows the data/model mismatch in pressure field magnitude for the 53-, 69-, 85-, and 101-Hz signals along the GPS positions. Since the GPS source path is known, the model is used to compute the pressure field magnitude point-by-point between the sensor and each sequential GPS source location. This means that the line integral in the adiabatic normal mode solution is not taken along colinear paths since the source is not moving exactly in a straight line heading radially away from the sensor. Note that the degree of mismatch is not uniform across frequency.

C. Matching process

Fundamentally, the idea is simple: find a track along which the model matches the data best as efficiently as possible. We construct the following estimate, based on the assumptions in Sec. I A, for the unknown track vector given its pressure field magnitude (PFM) vectors across frequency

$$\hat{\mathbf{b}}=\mathbf{b}_{\min} \text{ where}$$
$$E(\mathbf{b}_{\min},w_1',w_2',w_3')\leqslant E(\mathbf{b}_i,w_1,w_2,w_3) \quad \forall i\in\{1,2,...,N_E\}, \tag{1}$$

where N_E is the total number of exemplars in the region and

$$E(\mathbf{b},w_1,w_2,w_3)=w_1^2\|\mathbf{t}^{53}-\mathbf{p}^{53}\|^2+w_2^2\|\mathbf{t}^{69}-\mathbf{p}^{69}\|^2+w_3^2\|\mathbf{t}^{85}-\mathbf{p}^{85}\|^2+w_4^2\|\mathbf{t}^{101}-\mathbf{p}^{101}\|^2 \tag{2}$$

is the error measure between the data PFM vectors and the computed TLPFM vectors corresponding to $\mathbf{b}$ in the Euclidean norm where the weights sum to unity, i.e., $w_4=1-w_1-w_2-w_3$. $\mathbf{p}^{53}$, $\mathbf{p}^{69}$, $\mathbf{p}^{85}$, $\mathbf{p}^{101}$ correspond to the sensor data at each of the discrete frequencies of interest. Recall that the TLPFM and PFM vectors have been modified as discussed in Sec. I D in order to eliminate the unknown source level. The variable weights permit, per data window, for certain frequencies to be weighted more or less. This is necessary since there is not uniform sensitivity to the environmental param-

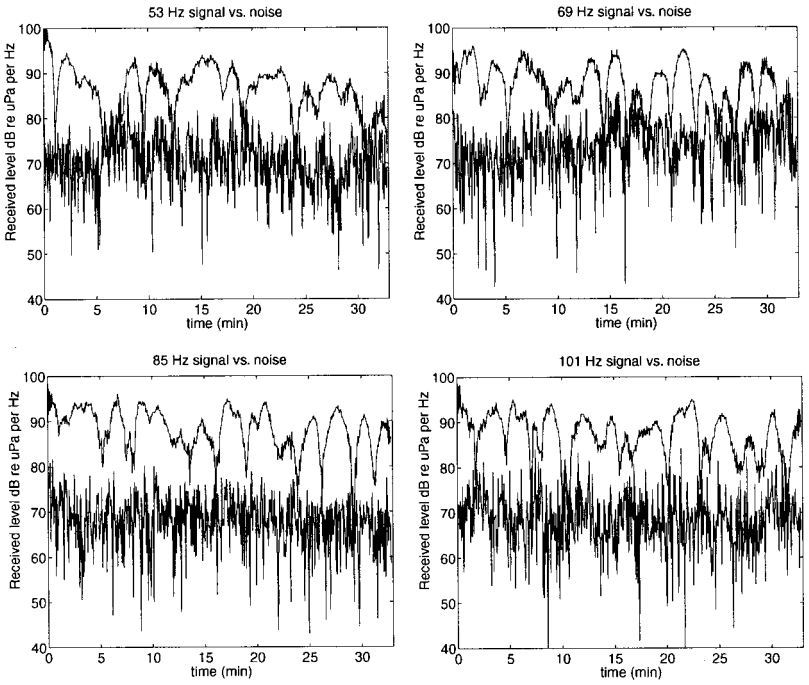


FIG. 8. Experimental source tow data and noise as a function of frequency.

eters across frequency. We will illustrate later the advantage of allowing these weights to be variable.

In principle, for a given a grid density, Eq. (1) is the solution to the problem since it accounts for all of the exemplars. The difficulty is that for a region of reasonable size, N_E becomes very large and the computational cost of checking every exemplar becomes significant. The central issue then becomes how we can obtain the estimate in Eq. (1) without examining all of the exemplars.

Using the definition of the error in Eq. (2), the idea is to use a small subset of exemplars, and, in parallel, reduce each one's error by adjusting its vector components, and then select the exemplar with the minimum error as the estimate of the unknown trajectory. Thus given a set of $M<N_E$ exemplars,

$$\mathbf{B}=\{(\mathbf{b}_i,\ \mathbf{t}_i^{53},\ \mathbf{t}_i^{69},\ \mathbf{t}_i^{85},\ \mathbf{t}_i^{101})\}_{i=1}^{M} \tag{3}$$

adjust each of them in parallel to get

$$\mathbf{B}_{\text{new}}\ \text{ such that, }\quad \text{for all }\ i\ \ E(\mathbf{a}_i^{\text{new}})\leqslant E(\mathbf{a}_i), \tag{4}$$

where $\mathbf{a}=[x,y,\theta,s,w_1,w_2,w_3]^T$ is a notational convenience that concatenates the tracking parameters $\mathbf{b}=[x,y,\theta,s]^T$ and the weights. In this way, the tracking problem reduces to M parallel multi-dimensional nonlinear minimization problems. These problems can be specified more exactly by using the

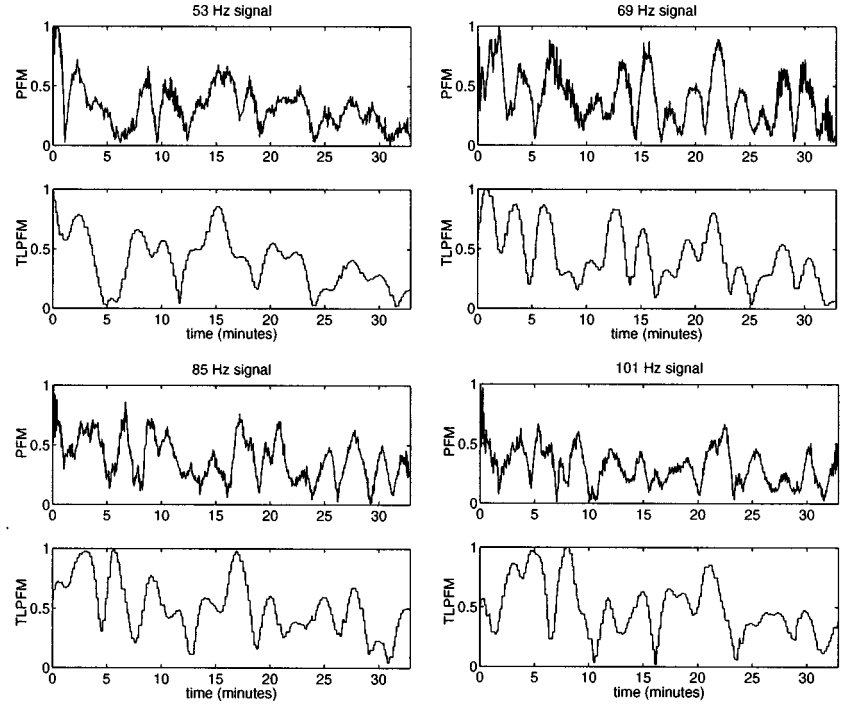


FIG. 9. Data/model mismatch for each frequency (data are in the upper panel and model results are in the lower panel).

assumptions in Sec. I A where the inequality constraints derive from the finite region of interest:

$$\begin{aligned} &\text{minimize} && E(x,y,\theta,s,w_1,w_2,w_3) \\ &\text{subject to:} && X_{lb}\leqslant x\leqslant X_{ub} \\ & && Y_{lb}\leqslant y\leqslant Y_{ub} \\ & && \Theta_{lb}\leqslant \theta\leqslant \Theta_{ub} \\ & && S_{lb}\leqslant s\leqslant S_{ub} \\ & && |w_1|,|w_2|,|w_3|<1 \\ & && w_1+w_2+w_3+w_4=1, \end{aligned} \tag{5}$$

where (X_{lb},X_{ub}), (Y_{lb},Y_{ub}), $(\Theta_{lb},\Theta_{ub})$, and (S_{lb},S_{ub}) are bounds on the x-coordinate, y-coordinate, heading, and speed, respectively. The weights are restricted to sum to unity and are bounded by one.

The problem now stated defines a complicated error surface over a domain described by inequality constraints. That is, the domain is a seven-dimensional box (one dimension for each of seven unknowns, x, y, θ, s, w_1, w_2, w_3). Unfortunately, even though the problem is stated in the classic form, there is no universal best way to solve it in a finite number of steps. In particular, we lack a closed form for the derivatives of Eq. (2) since the function values are based on a simulation. That is, every function evaluation corresponds to simulating the appropriate source motion (described by $\mathbf{b}$), using the propagation model to generate the corresponding predicted sensor data at the four frequencies ($\mathbf{t}^{53}$, $\mathbf{t}^{69}$, $\mathbf{t}^{85}$, $\mathbf{t}^{101}$) and then comparing it in the Euclidean norm to the observed sensor data with the corresponding weighting. This is time consuming and motivates us to avoid computing the function often.

Without derivative information, there is not much theory about how to proceed with the general minimization problem in the absence of global assumptions (even without the additional complication of the inequality constraints). Thus we construct our own customized solution based on unconstrained nonderivative methods and then implement the constraints by a simple parameterization. Specifically, we use a line search technique augmented with a conjugate direction search. Other approaches which have received a great deal of attention recently include simulated annealing[13] and genetic algorithms.[14]

The value of using a line search is that it makes the multi-dimensional problem a one-dimensional problem along a given direction in addition to making the constraints easy to implement. For example, by fixing (y,θ,s,w_1,w_2,w_3), the problem becomes

$$\begin{aligned} &\text{minimize} && E(x) \\ &\text{subject to:} && X_{lb}\leqslant x\leqslant X_{ub}, \end{aligned} \tag{6}$$

where X_{lb} and X_{ub} reflect the finite region of interest. Given this simpler problem, two questions immediately arise. First, how to solve the one-dimensional problem without derivatives; and second, how to pick which variables remain constant with respect to the others. In other words, along which directions should we search along the line. The line search problem is solved using Brent's method[15] and the directions are supplied by Powell's conjugate direction method.[16] Using these two methods in combination requires one additional step to implement the constraints.

1. Constraint implementation

We restate the minimization problem in a more general form to accommodate the arbitrary line search directions of Powell's method. We define $\mathbf{u}=[u_x,u_y,u_\theta,u_s,u_{w1},u_{w2},u_{w3}]$ as a unit vector in an arbitrary direction and α as a scalar variable. The minimization problem becomes

$$\begin{aligned} &\text{minimize} && E(\mathbf{a}+\alpha\mathbf{u}) \\ &\text{subject to:} && \alpha_{lb}\leqslant\alpha\leqslant\alpha_{ub}, \end{aligned} \tag{7}$$

where $(\alpha_{lb},\alpha_{ub})$ are the lower and upper bounds. Although Powell's method does not specify the incorporation of inequality constraints, by stating the problem in terms of a parametric variable, implementation of the constraints simply means using $(\alpha_{lb},\alpha_{ub})$ as the bounds of the line searches.

However, the determination of the upper and lower bounds $(\alpha_{lb},\alpha_{ub})$ are two smaller problems since, for arbitrary directions, the boundary constraints overlap. First, to determine α_{lb}, we solve

$$\begin{aligned} &\alpha_{lb}=\text{minimize} && \alpha \\ &\text{subject to:} && X_{lb}\leqslant x+\alpha u_x\leqslant X_{ub} \\ & && Y_{lb}\leqslant y+\alpha u_y\leqslant Y_{ub} \\ & && \Theta_{lb}\leqslant \theta+\alpha u_\theta\leqslant \Theta_{ub} \\ & && S_{lb}\leqslant s+\alpha u_s\leqslant S_{ub} \\ & && -1\leqslant w_1+\alpha u_{w_1}\leqslant 1 \\ & && -1\leqslant w_2+\alpha u_{w_2}\leqslant 1 \\ & && -1\leqslant w_3+\alpha u_{w_3}\leqslant 1 \\ & && w_1+w_2+w_3+w_4=1, \end{aligned} \tag{8}$$

and second, for α_{ub}, solve

$$\begin{aligned} &\alpha_{ub}=\text{maximize} && \alpha \\ &\text{subject to:} && X_{lb}\leqslant x+\alpha u_x\leqslant X_{ub} \\ & && Y_{lb}\leqslant y+\alpha u_y\leqslant Y_{ub} \\ & && \Theta_{lb}\leqslant \theta+\alpha u_\theta\leqslant \Theta_{ub} \\ & && S_{lb}\leqslant s+\alpha u_s\leqslant S_{ub} \\ & && -1\leqslant w_1+\alpha u_{w_1}\leqslant 1 \\ & && -1\leqslant w_2+\alpha u_{w_2}\leqslant 1 \\ & && -1\leqslant w_3+\alpha u_{w_3}\leqslant 1 \\ & && w_1+w_2+w_3+w_4=1. \end{aligned} \tag{9}$$

The solution is obtained simply by solving for α in each constraint and then picking the smallest positive α for α_{ub} and the largest negative α for α_{lb}.

2. Sectioning

We respond in the standard way to the jagged structure of the error surface by increasing the resolution of the line searches by dividing each of the intervals $(\alpha_{lb},\alpha_{ub})$ into n smaller subintervals wherein the slice of the error function is presumably smoother. The number of costly function evalu-

ations is proportional to n so the rule of thumb for this parameter is to make it as large as the computer resources comfortably allow. The error surface is jagged predominantly because of the structure of the magnitude of the pressure field as seen through the error measure. We will see later (in Fig. 14) that along certain directions the error surface is more jagged than along others.

3. Termination

When the error function is quadratic in the parameters (unlike ours), Powell's method is guaranteed to converge in $n_{\dim}$ cycles,[16] where $n_{\dim}$ is the number of unknowns. Although the error function in some places may be approximately quadratic, this is not the case in general. However, we can establish our own criterion for termination based on the result for quadratic termination and the grid resolution. In practice, the stopping criterion is a negligible change in the error function or (to avoid oscillation) the number of cycles greater than $2n_{\dim}$.

4. Tracking process summary

To review, we construct Eq. (1) to obtain estimates of an unknown track vector. The construction of Eq. (4) requires the solution of M nonlinear nonderivative minimization problems which we accomplish by modifying Powell's conjugate direction method to take advantage of our assumptions in Sec. I A through the use of a parametric variable.

Thus the tracking process is summarized as follows:

(1) Pick an initial set of M exemplars, $\{(\mathbf{b}_k, \mathbf{t}_k^{53}, \mathbf{t}_k^{69}, \mathbf{t}_k^{85}, \mathbf{t}_k^{101})\}_{k=1}^{M}$, and associated weights $w_{1,k}, w_{2,k}, w_{3,k}$ anyway you choose. This means that you now have one point on the error surface $E(x_k, y_k, \theta_k, s_k, w_{1,k}, w_{2,k}, w_{3,k})$ for each exemplar. From experiments, we know that the accuracy of the solution varies asymptotically with M, which means that for large enough M, the incremental benefit from increasing M is negligible. Thus the rule is to make M as large as resources allow. Each of these M exemplars essentially is a probe of the error surface that explores the surface in parallel with each of the others.
(2) Pick a set of mutually conjugate initial search directions, usually the coordinate directions.
(3) Pick a subsectioning value that reflects the granularity of the line search. The greater this value, the slower the processing since it is directly proportional to the number of times the function is evaluated.
(4) In parallel for each exemplar, apply Powell's conjugate direction method augmented with Brent's line search where the constraints are implemented as bounds on the line search determined by an inequality constrained linear optimization problem.
(5) Terminate each search per exemplar when there is a negligible change in the error measure or when the number of iterations exceeds twice the number of unknowns.
(6) Examine the error measures across the newly adjusted exemplars and pick the one that corresponds to the least error measure value. The corresponding track vector is the estimate for the unknown track.

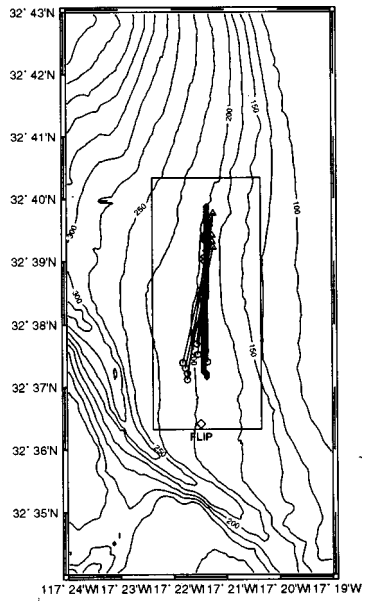


FIG. 10. Tracking results using experimental data.

Note that because the exemplars are adjusted in parallel and do not need to communicate until the end of the process, the structure of the method makes it ideal for implementation on a parallel machine architecture.

D. Results

For these results, the subsection parameter $n=10$ and the number of exemplars $M=26$. Recall that the region wherein tracks are restricted to start is approximately 27 km^2 which means that $(X_{ub}-X_{lb})(Y_{ub}-Y_{lb})\approx 27\,\text{km}^2$. The set of $M=26$ exemplars is evenly distributed northward along the $x=X_{lb}$ axis so that the initial points for the track vectors (x,y) are spaced approximately 500 m apart in latitude. This means that the x-coordinate of the track vectors is the same for all exemplars. The heading component, θ, of the track vectors are spaced π radians apart (i.e., one northward and one southward). The exemplar speed $s=3$ m/s, and the source speed bounds are $[S_{lb}, S_{ub}]=[2,4]$ m/s. The purpose of placing the initial exemplar set along the $x=X_{lb}$ axis is that the initial set of conjugate search directions starts in the x-direction, so that there is built in sampling in this direction through the subsectioning. This is a matter of convenience, however, and a random set can also be used.

Figure 10 shows the results when the processor is evaluated using the experimental 53-, 69-, 85-, and 101-Hz pressure field magnitude data over the source tow path in eight overlapped 25-min windows. Source speed, source level, source initial position, and source heading are unknown *a priori*. Source depth is known and fixed (47 m). On this plot, circles represent the initial points (x, y) of each of eight estimated tracks and ''△'' represents final points. Since the sampling window is 25 min, the length of the segment connecting initial and final points is a function of the estimated source speed. The dark line in the center of the box is the GPS source track. The bathymetry contours underly the plot. The sensor position is indicated by the diamond ''◇.''

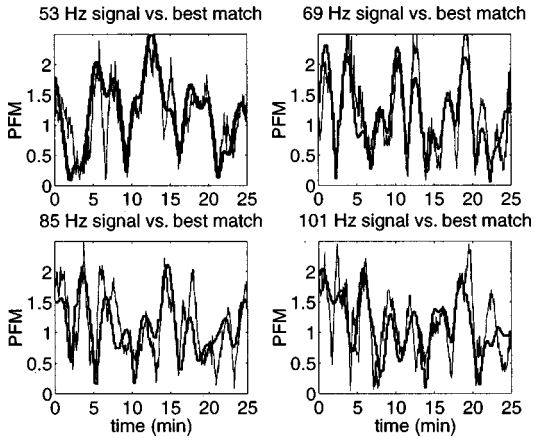


FIG. 11. Data/model mismatch over the estimated track for each frequency.

Figure 11 shows a representative example of the match between the data PFM vector and the model TLPFM vector along the estimated track for the fourth overlapped data window across each of the four frequencies. The dark line represents the modeled pressure field along the estimated track and the other line is the actual sensor data. This result shows that much of the structural information is matched between model and data along the estimated track. Compare this data/model mismatch with that along the correct track in Fig. 9 and note that the estimated track provides a better match than the correct track. We will explore this point shortly. First consider the role of the weights in Eq. (2).

Earlier, we said that allowing the individual frequency mismatches to be weighted differentially in Eq. (2) accommodates the nonuniform data/model mismatch across frequency. Figure 12 shows the weights for each frequency across the eight data windows. This shows that the dominant frequency is 53 Hz, especially for the last several windows. This makes sense since the higher frequencies are more susceptible to data/model mismatch (see Fig. 9). On the other hand, the weights indicate that the other frequencies have predictive value since they are not all zero. The value in allowing the differential weighting is that we do not have to quantify the degree of the mismatch across frequency *a priori*. The searching process does that automatically.

As an example, Fig. 13 shows the weights for each frequency across the eight data windows when the 53-Hz signal

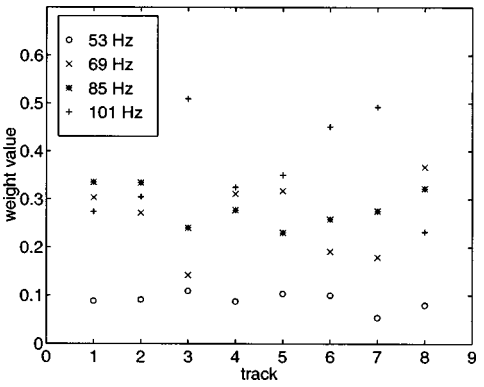


FIG. 13. Adaptive weighting along overlapped data windows when the 53-Hz signal has been replaced by noise.

has been replaced by noise from a nearby frequency (see Fig. 7). This shows that the adaptive weighting process is able to reduce automatically the influence of frequencies with low predictive value. Thus another value of this kind of weighting is that it compensates automatically for poor quality data.

Figure 14 shows, for the first 25-min data window, slices of the error surface along each of the coordinate directions, (x,y,θ,s). We omit the slices along the w_1, w_2, w_3 directions since those are uninteresting parabolas. The abscissa α is the parametric variable that stretches in both directions, originating at the estimated track vector ($\alpha=0$), and terminating on the surface of the seven-dimensional box that represents the bounds on each coordinate. This plot shows the topology of the error surface around the estimated track in the indicated directions. Note that the slices are relatively smooth in the *x*-direction and speed direction. This is because the bathymetry rolls off gently in the *x*-direction and parallel northward tracks have corresponding TLPFM vectors that are not, on a small scale, significantly different in this direction. In the case of the speed slice, we see that, because speed corresponds to a smooth compression/expansion of the TLPFM vector in time, we get a smooth slice. The most jagged slices correspond to the variation in heading (θ) and *y*-direction. The profound variation in heading comes from spinning the track about its initial point, naturally exploring the structure of the error surface in a more radical fashion. Since the experimental track moves nearly radially away from the sensor, the *y*-direction (north–

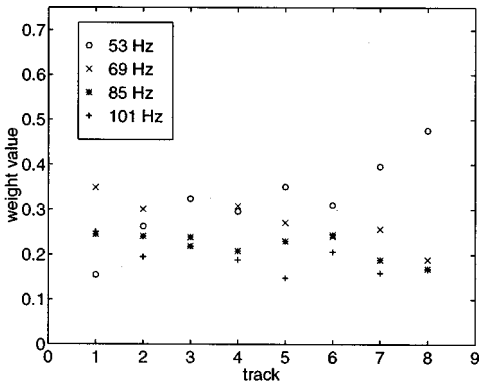


FIG. 12. Adaptive weighting along overlapped data windows.

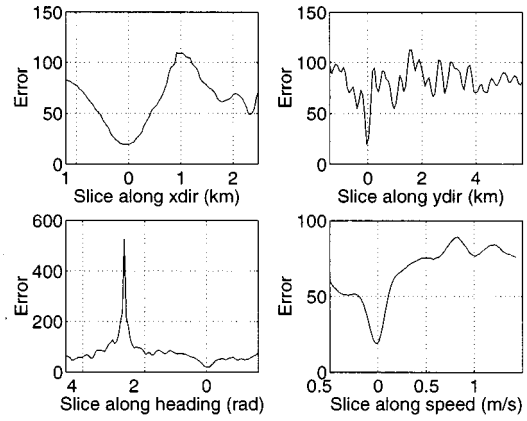


FIG. 14. Error slices about the estimated track.

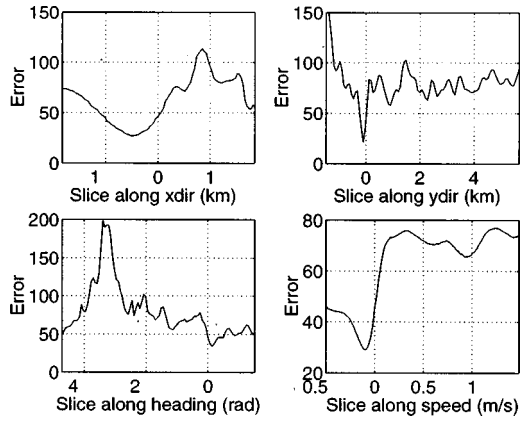


FIG. 15. Error slices about the correct track.

south) is the direction of greatest structural variability in the TLPFM vector. Thus there is high variation in the *y*-direction slice. Finally, note that this kind of detail in the slices is purely for visual effect; the processor itself acts only upon samples of these curves, due to the cost of function evaluation.

Earlier we noted that the estimated track seemed to provide better data/model mismatch than the correct track (compare Fig. 9 with Fig. 11). Figure 15 shows, for the first 25-min data window, slices of the error surface along each of the coordinate directions, (x,y,θ,s) about the correct track. This plot shows that, due to data/model mismatch, the correct answer is slightly displaced from the minimum along all directions, thus making it impossible for a minimum-seeking process to find it.

E. Postprocessing

Because we use overlapped data windows, we know that the track vector estimates for each of those windows should also be overlapped. Furthermore, we know, based on the assumptions of constant source speed, that the lengths of the track vectors (i.e., distance between start and end points) should be the same throughout.

In Fig. 16, the circles and triangles represent the start and end points, respectively of individual track vector estimates. The dashed lines show an example of the kind of track estimates we typically obtain and the solid lines show what these estimates should really look like. In other words, since we know that the data windows are overlapped, it should be the case that the track estimates are also overlapped and co-linear, since, by assumption, the source changes neither heading nor speed. Due to errors in processing, the individual track vector estimates have different speed estimates and thus travel different total distances over a fixed time window, τ_k, whereas they should be constant, τ. Therefore, the objective of post-processing is to consolidate the individual estimates (dashed lines) per data window in a least-squares sense (solid line). The (x_k,y_k), (w_k,z_k) represent the estimated track start and end points, respectively, for the kth track vector estimate. The (a_k,b_k), (c_k,d_k) represent the postprocessed track start and end points, respectively, for the kth track vector, which are functions of the first start point, (ξ_1,φ_1), and the last end point, (ξ_2,φ_2). Thus

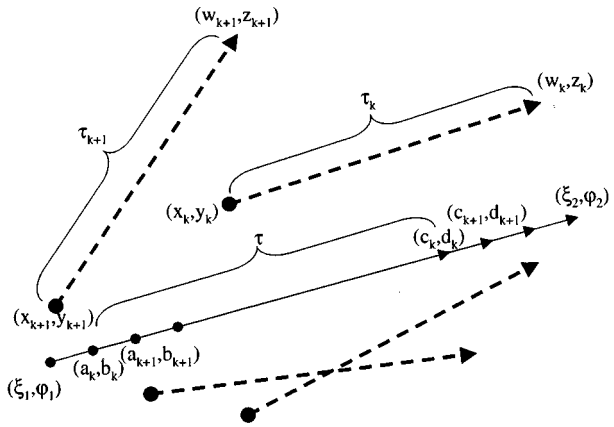


FIG. 16. Post-processing of overlapped data windows.

(a_k,b_k,c_k,d_k) are all functions of $(\xi_1,\varphi_1,\xi_2,\varphi_2)$ since two points describe a line. Consequently, the objective of post-processing is to find $((\xi_1,\varphi_1),(\xi_2,\varphi_2))$ that minimizes the Euclidean distance between each $\{(a_k,b_k),(c_k,d_k)\}$ and $\{(x_k,y_k),(w_k,z_k)\}$, respectively.

To accomplish this postprocessing, we minimize the following error:

$$E(\xi_1,\xi_2,\varphi_1,\varphi_2)=\sum_{k=1}^{N}(x_k-a_k)^2+(y_k-b_k)^2+(w_k-c_k)^2+(z_k-d_k)^2, \tag{10}$$

where

$$a_k=\frac{(1-\nu k+N\nu)\xi_1+\nu(k-1)\xi_2}{N\nu-\nu+1}, \tag{11}$$

$$b_k=\frac{(1-\nu k+N\nu)\varphi_1+\nu(k-1)\varphi_2}{N\nu-\nu+1}, \tag{12}$$

$$c_k=\frac{\nu(N-k)\xi_1+(1-\nu+\nu k)\xi_2}{N\nu-\nu+1}, \tag{13}$$

$$d_k=\frac{\nu(N-k)\varphi_1+(1-\nu+\nu k)\varphi_2}{N\nu-\nu+1}, \tag{14}$$

and $\nu\equiv\mu/\tau$. The simple form of the error makes it solvable analytically. Specifically,

$$\begin{bmatrix}\xi_1\\ \xi_2\end{bmatrix}=\begin{bmatrix}A_{1,1} & A_{1,2}\\ A_{2,1} & A_{2,2}\end{bmatrix}\begin{bmatrix}f_1\\ f_2\end{bmatrix}$$

and

$$\begin{bmatrix}\varphi_1\\ \varphi_2\end{bmatrix}=\begin{bmatrix}A_{1,1} & A_{1,2}\\ A_{2,1} & A_{2,2}\end{bmatrix}\begin{bmatrix}g_1\\ g_2\end{bmatrix},$$

where

$$f_1=\sum_{i=1}^{N}\frac{2\nu(N-i)w_i+2(1-\nu i+N\nu)x_i}{N\nu-\nu+1},$$

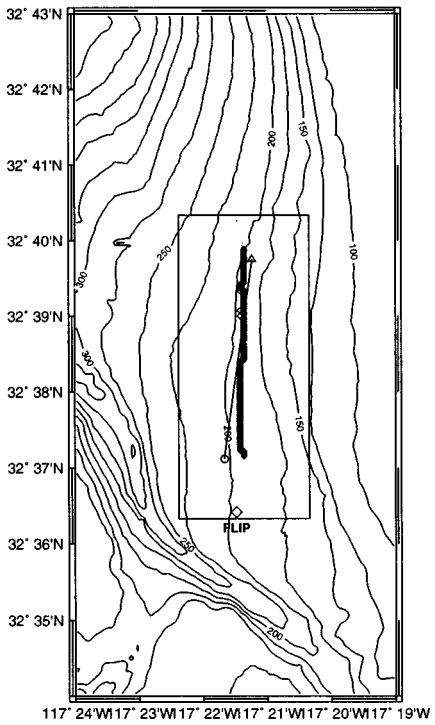


FIG. 17. Post-processed experimental data track estimates.

$$f_2=\sum_{i=1}^{N}\frac{2\nu(i-1)x_i+2(1-\nu+i\nu)w_i}{N\nu-\nu+1},$$

$$g_1=\sum_{i=1}^{N}\frac{2\nu(N-i)z_i+2(1-\nu i+N\nu)y_i)}{N\nu-\nu+1},$$

$$g_2=\sum_{i=1}^{N}\frac{2\nu(i-1)y_i+2(1-\nu+i\nu)z_i}{N\nu-\nu+1},$$

$$A_{1,1}=\frac{\gamma}{2N}(3-3\nu+2\nu^2N^2+\nu^2-3\nu^2N+3\nu N),$$

$$A_{1,2}=\frac{-\gamma\nu}{2N}(N-1)(\nu N-2\nu+3),$$

$$A_{2,1}=\frac{-\gamma\nu}{2N}(N-1)(\nu N-2\nu+3),$$

$$A_{2,2}=\frac{\gamma}{2N}(3-3\nu+2\nu^2N^2+\nu^2-3\nu^2N+3\nu N),$$

$$\gamma\equiv\frac{1}{(N\nu)^2-\nu^2+3}.$$

Figure 17 displays the results of postprocessing the separate track estimates for each of the 25-min data windows shown in Fig. 10. The correct track is the dark vertical (northward) line. Since the sampling window is 25 min, the length of the segment connecting initial and final points is a function of the estimated source speed. Note that the post-processed line is consistent with the north–east (rightward) pull of the estimated tracks. The initial position x-coordinate is 510 m from the correct position. Likewise, the y-coordinate, heading and speed differ by 60 m, 9 degrees, and 0.2 m/s, respectively (the actual source speed was 2.6 m/s).

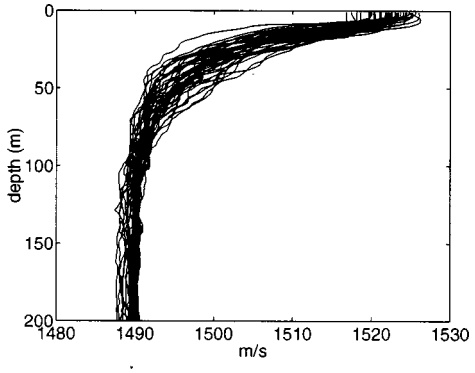


FIG. 18. Sound speed measurements taken during SWellEx-3.

III. SOUND SPEED PERTURBATIONS

In the previous section, we developed a single-sensor tracking method and demonstrated its use with experimental data. Since there is always some inherent uncertainty in the sound speed profile used to build the propagation model, we investigate whether the tracking process can be extended to these cases. We construct empirical orthogonal functions (EOF)[17] using measurements of the sound speed structure over a ten day period of the SWellEx-3 experiment and in the vicinity of the source tow data analyzed (see Fig. 18). We take the eigenvector corresponding to the largest eigenvalue and define a perturbation as

$$\mathbf{c}(\tau)\equiv\tau\mathbf{u}_1+\overline{\mathbf{c}}. \tag{15}$$

We choose the following discrete values for $\tau=\{-120,-60,-30,-20,-10,0,10,20,30,60,120\}$ which allows for a generous range of values considering the distribution ($\sim N(0,\sigma=20)$) of the projections of the measurements shown in Fig. 18 onto the first eigenvector. Figure 19 shows the resulting family of perturbed sound speed profiles using these discrete values of τ.

Recall that, for each frequency, the tracking process compares the sensor data with predicted data generated by a propagation model which incorporates the measured sound speed profile. We extend the tracking process to consider multiple propagation models, each based on a different perturbation of the sound speed. That is, we perform tracking

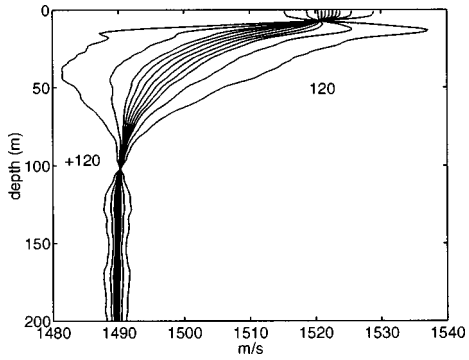


FIG. 19. Family of perturbed sound speed profiles.

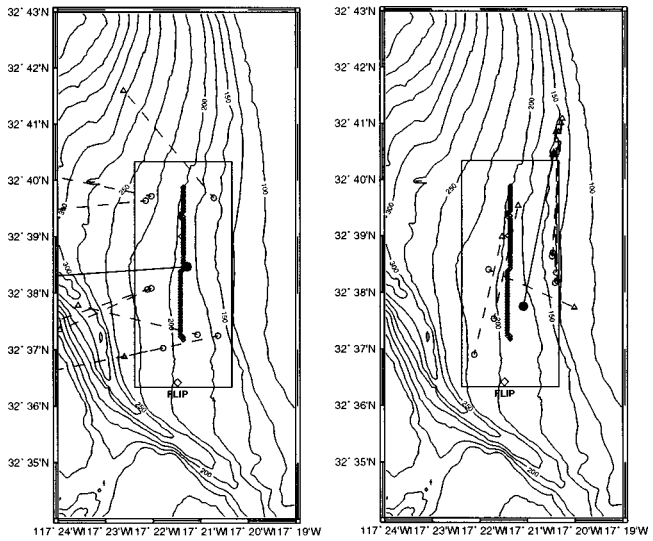


FIG. 20. Experimental tracking results using perturbation $\tau=-60$ on the left and $\tau=-20$ on the right. The dotted lines show each overlapped track vector estimate. The solid line is the post-processed estimate.

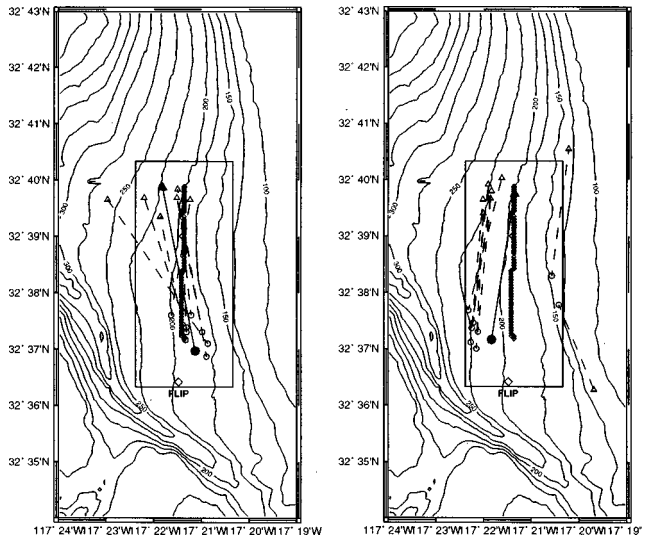


FIG. 22. Experimental tracking results using perturbation $\tau=20$ on the left and for $\tau=60$ on the right. The dotted lines show each overlapped track vector estimate. The solid line is the post-processed estimate.

exactly the same as before except the track vector estimates are conditioned on different sound speed perturbations

$$\hat{\mathbf{b}}_{|\tau=-120},$$

$$\hat{\mathbf{b}}_{|\tau=-60},$$

$$\vdots$$

$$\hat{\mathbf{b}}_{|\tau=120}. \tag{16}$$

Figures 20–22 show the results from tracking using several values of the perturbation parameter τ.

We need a way of choosing between these track vector estimates so we examine the post-processing error Eq. (10) which is a measure of the clustering of the individual track vector estimates (see Fig. 23). The minimum occurs for $\tau=10$ which coincidentally is the perturbation closest to Fig. 8, in the sense of the EOF expansion. This result shows that it is possible to estimate the correct sound speed through the perturbation by tracking. The minimum post-processing error is a meaningful criterion since a value of zero corresponds to the case of perfectly colinear individual track vector estimates.

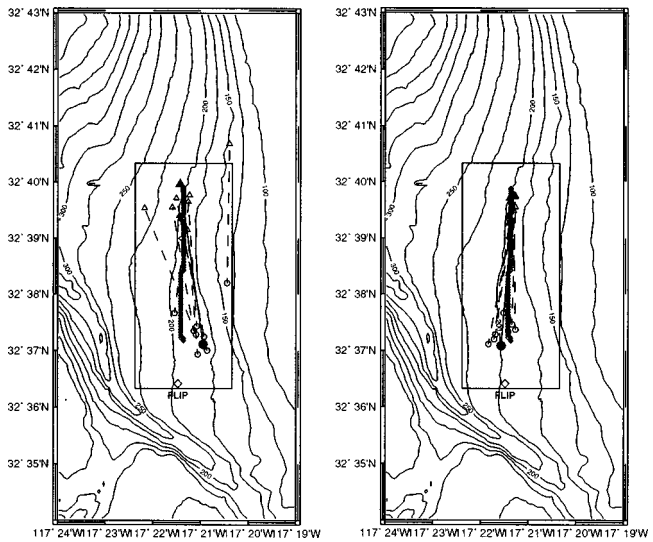


FIG. 21. Experimental tracking results using perturbation $\tau=0$ on the left and $\tau=10$ on the right. The dotted lines show each overlapped track vector estimate. The solid line is the post-processed estimate.

IV. TRACKING IN NOISE

In this section, we will investigate the performance of the tracking process with respect to different levels of signal versus noise. We begin by taking measurements of the ambient noise during the source tow experiment at frequencies near to the frequencies of interest. We use these noise data as input to the tracking process. Figure 24 shows the experimental noise data measured 1 Hz away from each of the frequencies of interest (54, 70, 86, 102 Hz). Since the source does not emit energy in these bins, they represent a good sample of ambient noise around the frequencies of interest.

Using the experimental noise as input to the processor, we obtain the track estimates on the left side of Fig. 25 for

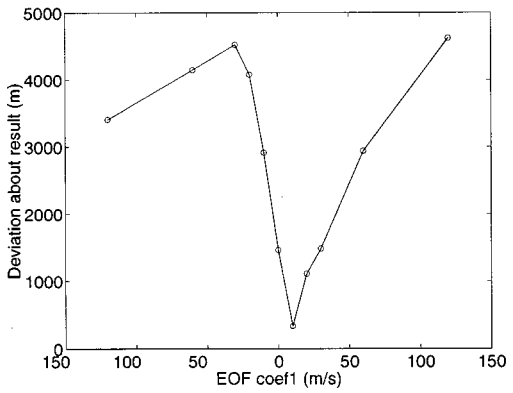


FIG. 23. Post-processing error as a function of sound speed structure perturbation parameter τ.

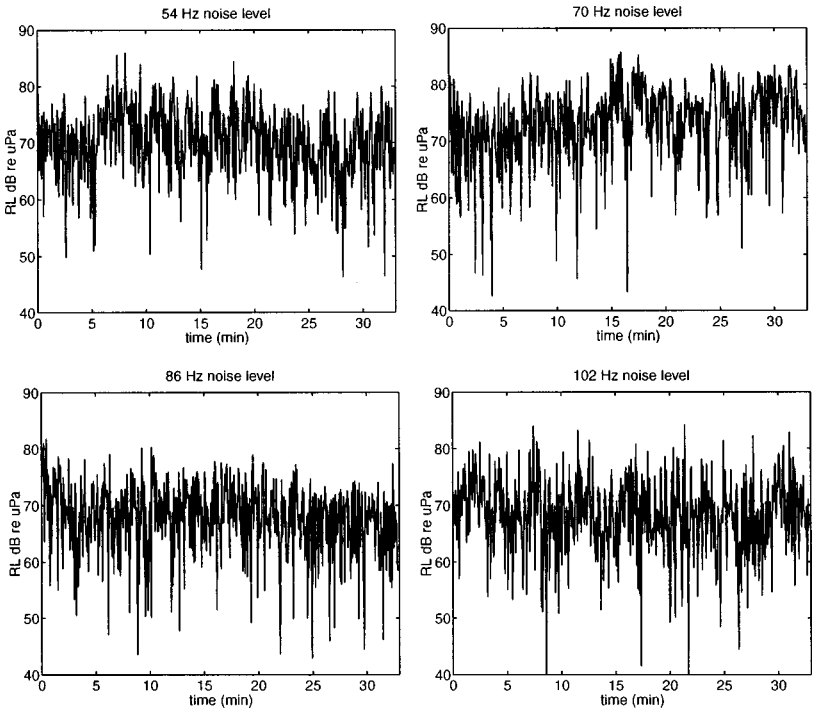


FIG. 24. Experimental noise as a function of frequency.

the overlapped data windows. The right side of Fig. 25 shows the result of post-processing these individual estimates to obtain the final estimated track.

As a representative example of the TLPFM vectors the processor finds to match the experimental noise, consider Fig. 26. Note that there are flat TLPFM vectors that fit the noise in a least-squares sense and correspond to the track vectors shown in Fig. 25. This result indicates that the tracking process does not produce a random splatter of track estimates when the input is noise; rather, it produces track estimates consistent with minimizing the error measure.

A. Adding noise to experimental data

To see how the process incrementally degrades with noise, we contaminate the experimental data for each frequency with complex additive Gaussian noise. Figure 27, on the left, shows the tracking results (assuming the known sound speed profile) using the experimental data with white noise added to a signal-to-noise ratio of 5 dB as input to the processor. Figure 27, on the right, shows the tracking results using the experimental data added with white noise to a signal-to-noise ratio of 10 dB as input to the processor. In this case, the tracking process at least favors a northerly track.

These results indicate that the tracking process behaves in a consistent way when noise is added to the input signals. That is, as more signal is available, the process is able to better estimate the underlying track. However, these results also show that the process is so internally consistent that it can create track vector estimates even when there is no signal present in the input. Thus the tracking process requires an additional component to quantify reliability of track estimates. This is the topic of the next section.

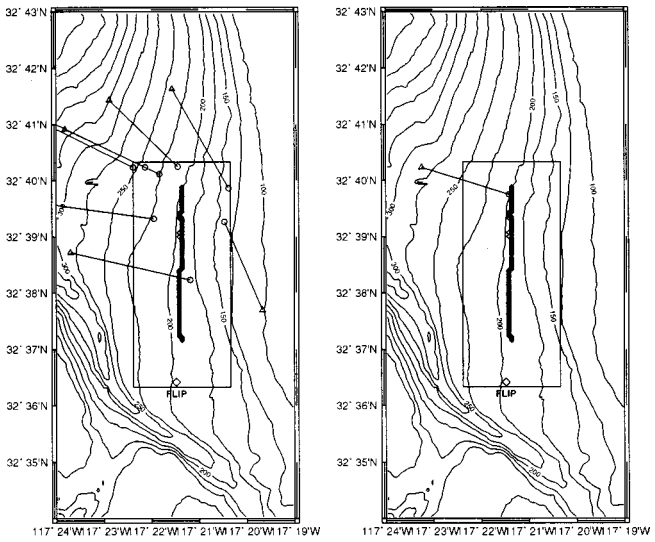


FIG. 25. Experimental noise tracking results.

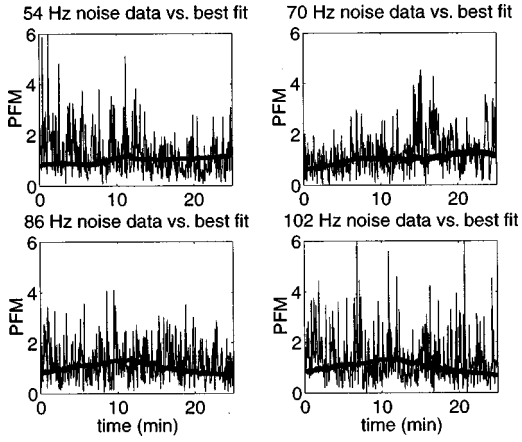


FIG. 26. Model (heavy line) fitting experimental noise (light line).

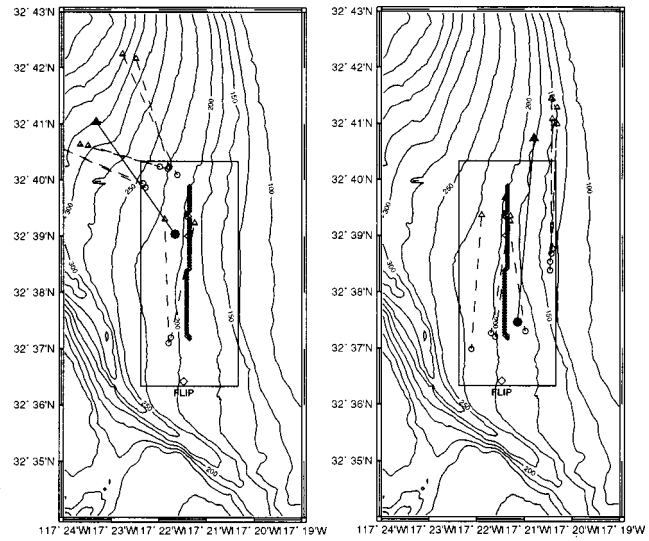


FIG. 27. On the left, SNR=5 dB tracking results; on the right, SNR=10 dB tracking results.

B. Data admissibility

The approach to quantifying the reliability of the track vector estimates is to develop a classifier to decide to what degree the input is representative of signal. We do this by defining a signal subspace as the span of the 20 eigenvectors corresponding to the 20 largest eigenvalues of an EOF decomposition of signal TLPFM vectors. To generate this EOF decomposition, we select randomly 300 TLPFM vectors at each frequency assuming that each of the tracking coordinates (x,y,θ,s) is distributed uniformly. We place these vectors in the columns of a matrix and find a basis for them using the same kind of EOF decomposition as for the sound speed perturbation. That is, using the eigenvectors corresponding to the 20 largest eigenvalues.

$$\hat{\mathbf{p}} \equiv \sum_{i=1}^{20} \alpha_i \mathbf{u}_i + \overline{\mathbf{t}}, \tag{17}$$

where $\alpha_i = (\mathbf{p} - \overline{\mathbf{t}})^T \mathbf{u}_i$ and $\overline{\mathbf{t}}$ is the row-wise average vector over the 300 TLPFM vectors. Recall that the TLPFM and PFM vectors have been modified as discussed in Sec. I D in order to eliminate the unknown source level.

We added white noise to the experimental data for the tracking results shown in Fig. 27. We now approximate those PFM (i.e., $\mathbf{p}_{s+n}$) vectors using the EOF decomposition according to Eq. (17) and consider the following average ratio:

$$\overline{d} = \frac{1}{8} \sum_{j=1}^{8} \frac{\|\hat{\mathbf{p}}_{s+n,j}\|}{\|\mathbf{p}_{s+n,j}\|},$$

where $j \in \{1,2,...,8\}$ for each of eight overlapped windows. This is a measure of how much of the noise corrupted PFM vectors lie in the signal subspace defined by the EOF expansion.

In Table I, the entries show how much of the noise corrupted data projects onto the signal subspace. Recall from Sec. II A, that the experimental data itself varies in SNR between 13 dB (69 Hz) and 19 dB (85 Hz). The value of

TABLE I. Average data projection onto the signal subspace ($\overline{d}$).

Frequency (Hz)	53	69	85	101
Exp data	0.97	0.95	0.99	0.97
Exp data+noise, SNR=10 dB	0.96	0.94	0.96	0.94
Exp data+noise, SNR=5 dB	0.93	0.92	0.94	0.92
Exp noise	0.86	0.88	0.85	0.81

Table I is that it provides a way to define a threshold through which either data could be rejected for processing or the confidence in the track vector estimate could be quantified. For example, it is evident from Table I that $\overline{d}=0.95$ would provide an appropriate threshold for the track investigated.

V. CONCLUSIONS

There are two primary conclusions from this work. First, it is possible with real data to localize the trajectory of a moving acoustic source using a single sensor in a complicated propagation environment. Second, it is possible to estimate environmental parameters (e.g., sound speed structure) as an integral part of the tracking process. Potentially, the evolution of sound speed structure could be estimated over long time scales as part of tracking different sources (e.g., surface ships transiting through a region).

There are several shortcomings in the research presented. First, there is only one example using real data, and so it is difficult to make generalizations on that basis. Although an attempt was made to find other suitable tracks in the data set, nothing of sufficient length along a straight course was found for which the adiabatic normal mode solution provided good data/model matching.

The next immediate research direction is to consider tracking a surface ship instead of a towed source. Other directions for further work include using this process in different/larger environments, with longer data windows, with different source depths/frequencies using other propagation models. These directions may require some modest change to the details of the tracking process, although the thrust of the implementation should likely be unaltered. Beyond investigating sensitivity to the sound speed, investigating the estimation of other environmental parameters (e.g., seafloor geoacoustic parameters) either separately or in tandem is another issue. Last, whereas we consider straight tracks, nothing in the construction of the process prohibits the investigation of other types of tracks, as long as they can be parameterized by relatively few terms (e.g., splines).

ACKNOWLEDGMENTS

The San Diego Supercomputer Center provided much of the computational resources used in this study. J. Murray (MPL) assisted with both the experimental data and the initial modeling using KRAKEN. C. Guest (ECE/UCSD) and G. D'Spain (MPL) provided many insightful comments throughout this work. Fellowship support for J. Unpingco initially was provided by the Ford Foundation and subsequently by the Office of Naval Research, Code 321 US.

[1] H. Bucker, ''Use of calculated sound fields and matched-field detection to locate sound sources in shallow water,'' J. Acoust. Soc. Am. **59**, 368–373 (1976).
[2] A. B. Baggeroer, W. A. Kuperman, and P. N. Mikhalevsky, ''An overview of matched field methods in ocean acoustics,'' IEEE J. Ocean. Eng. **18**, 401–424 (1993).
[3] L. Frazer and P. Pecholcs, ''Single-hydrophone localization,'' J. Acoust. Soc. Am. **88**, 995–1002 (1990).
[4] H. Bucker, ''Matched-field tracking in shallow water,'' J. Acoust. Soc. Am. **96**, 3809–3811 (1994).
[5] J. Fawcett and B. Maranda, ''A hybrid target motion analysis/matched-field processing localization method,'' J. Acoust. Soc. Am. **94**, 1363–1371 (1993).
[6] M. D. Collins, L. E. Tinker, J. S. Perkins, N. C. Makris, and W. A. Kuperman, ''Single-frequency matched-field processing with a single hydrophone,'' J. Acoust. Soc. Am. **90**, 2366 (1991).
[7] M. D. Collins, L. T. Fialkowski, W. A. Kuperman, and J. S. Perkins, ''Environmental source tracking,'' J. Acoust. Soc. Am. **94**, 3335–3341 (1993).
[8] W. A. Kuperman, M. D. Collins, J. S. Perkins, L. T. Fialkowski, T. L. Krout, L. Hall, R. Marrett, L. J. Kelly, and A. Larsson, ''Environmental source tracking using measured replica fields,'' J. Acoust. Soc. Am. **94**, 1844 (1993).
[9] J. S. Perkins, L. T. Fialkowski, M. D. Collins, W. A. Kuperman, and T. L. Krout, ''TESPEX: Initial processing results,'' in Proc. of the 22nd Annual International Meeting of the Technical Cooperative Program, Subgroup G, Technical Panel Nine, pp. 1–17 (1993).
[10] N. O. Booth, P. A. Baxley, J. A. Rice, P. W. Schey, W. S. Hodgkiss, G. L. D'Spain, and J. J. Murray, ''Source localization with broadband matched field processing in shallow water,'' IEEE J. Ocean. Eng. **21**, 402–412 (1996).
[11] R. Bachman, P. Schey, N. Booth, and F. Ryan, ''Geoacoustic databases for matched-field processing: Preliminary results in shallow water off San Diego, California,'' J. Acoust. Soc. Am. **99**, 2077–2085 (1996).
[12] M. Porter, ''The KRAKEN Normal Mode Program,'' Technical report, SACLANT Undersea Research Center (1990).
[13] M. Collins and W. Kuperman, ''Focalization: Environmental focusing and source localization,'' J. Acoust. Soc. Am. **90**, 1410–1422 (1991).
[14] P. Gerstoft, ''Inversion of seismoacoustic data using genetic algorithms and *a posteriori* probability distributions,'' J. Acoust. Soc. Am. **95**, 770–782 (1994).
[15] R. P. Brent, *Algorithms for Minimization Without Derivatives* (Prentice-Hall, Englewood Cliffs, NJ, 1972).
[16] M. Powell, ''A view of unconstrained minimisation algorithms that do not require derivatives,'' ACM Trans. Math. Softw. **1**, 97–107 (1975).
[17] L. LeBlanc and F. Middleton, ''An underwater acoustic sound velocity data model,'' J. Acoust. Soc. Am. **67**, 2055–2062 (1980).

Acoustic phase conjugation by nonlinear piezoelectricity. I. Principle and basic experiments

Masahiro Ohno, Ken Yamamoto, Akira Kokubo, Keiji Sakai, and Kenshiro Takagi
Institute of Industrial Science, University of Tokyo, 7-22-1 Roppongi, Minato-ku, Tokyo 106, Japan

(Received 4 February 1998; revised 20 April 1999; accepted 1 June 1999)

The principle and the basic experimental results of the acoustic phase conjugation by nonlinear piezoelectricity are described. Acoustic phase conjugate waves at frequency ω are generated as a result of the interaction between incident acoustic waves at ω and pump electric fields at 2ω. An original explanation based on the modulation of sound velocity by the electric field together with the concept of time grating is given. Coupling equations for the PZT ceramics as a nonlinear material are derived. Experimental results of the sound velocity modulation and the phase conjugate generation in nonlinear piezoelectric PZT ceramics are shown. Amplitude reflectivity of the phase conjugation was 23% for the acoustic incidence at 10 MHz. © *1999 Acoustical Society of America.* [S0001-4966(99)02709-5]

PACS numbers: 43.35.Sx [HEB]

INTRODUCTION

The wave equation in a lossless medium is unchanged if the sign of time is inverted. Therefore, if a wave $f(r,t)$ is a solution of a given wave equation, its time-reversed wave $f(r,-t)$ is also the solution. This idea gives the basis of the concept of ''phase conjugation.'' A ''phase conjugate wave'' is defined as a wave which has the same spatial distribution as the incident wave but travels in the opposite direction. This phenomenon was first observed by Zel'dovich in a stimulated scattering of light.[1] After this discovery in 1972, phase conjugation has been extensively studied in optics.[2,3] It is not only a very interesting physical phenomenon itself, but also possesses the possibility of wide range of application. In particular, its application to imaging technique, based on the aberration-free nature, is very attractive and many papers have been reported on this subject.

Studies in acoustics started in 1980's and several kinds of methods to generate phase conjugate waves have been proposed. It was found that some already known physical phenomena contain the process of phase conjugation. Since then, a considerable number of studies on this matter have been reported. The range of study extends from the basic study to the application to practical ultrasonic instruments. It seems that both the quality and the quantity of the study of acoustic phase conjugation are growing in recent years.

In this and in accompanying papers, we describe the results of our study of acoustic phase conjugation in nonlinear piezoelectric media. The phase conjugate generation via nonlinear piezoelectricity has been studied since the 1980's, and several kinds of materials have been used. In our study, we pay special attention to PZT ceramics as a nonlinear piezoelectric medium. In Secs. I A and I B, we describe the principle of generating acoustic phase conjugate waves via nonlinear piezoelectricity. We give an original explanation for this phenomenon based on the modulation of sound velocity by the electric field together with the concept of time grating. In Sec. I C, we derive the coupling equations for the case of PZT ceramics as a nonlinear material. Experimental results on the sound velocity modulation and the generation of phase conjugate waves are given in Sec. II.

The methods for acoustic phase conjugation reported so far can be categorized into four groups: purely acoustic method, nonlinear piezoelectric method, nonlinear magnetoacoustic method, and purely electrical method. The first three methods involve some physical interaction between acoustic waves themselves or between acoustic waves and electromagnetic fields, whereas the fourth method makes use of signal processing with electrical circuits.

To the knowledge of the authors, most of the studies on the earliest stage are categorized into the purely acoustic method. Bunkin *et al.* suggested and experimentally demonstrated the acoustic phase conjugate generation via nonlinear effects in liquid containing bubbles,[4,5] liquid surface,[6] and thermal waves.[7] Sato *et al.* reported acoustic phase conjugate waves were generated via four wave mixing (FWM) in liquid suspending small particles.[8] Watanabe *et al.*[9] utilized a thin layer containing polymer microparticles flickering at a frequency twice the acoustic frequency. These purely acoustic methods operate at relatively low acoustic frequencies (kHz—MHz). This is because the nonlinear interaction occurs in liquid, in which the acoustic absorption is substantially large at higher frequencies.

The second category (nonlinear piezoelectric method), which is the main interest of this paper, needs some explanation on its history. As is explained later, this method is based on the parametric interaction between acoustic waves at a frequency ω and an electric field at a frequency 2ω. Svaasand[10] reported on this effect for the first time within the knowledge of the authors. Thompson and Quate[11] studied this phenomenon extensively, with special interest in the generation of an electric field at 2ω as a result of the mixing of two acoustic waves at ω. In their work, and in the following studies in this field, the main interest was in the ''convolving effect'' of two acoustic waves. The phase conjugate effect involved in this phenomenon seems to have been paid little attention at that time.

In the 1980's, after the prosperous research of optical phase conjugation, this nonlinear piezoelectric interaction attracted the attention of researchers with special interest in the wavefront of acoustic waves. It was pointed out that two acoustic fields in this interaction were in the relation of phase conjugation, and this phenomenon was nothing but a method of generating acoustic phase conjugate waves with the help of electrical field.[12,13] Papers on phase conjugation using nonlinear piezoelectric materials such as $LiNbO_3$,[14] CdS, BK7 glass,[15] and PZT ceramics[16,17] have been presented. The detail of this interaction will be described in subsequent sections of this paper.

The third category, nonlinear magneto-acoustic method, can be described as a magnetic counterpart of above method. It was reported that an acoustic phase conjugate wave at a frequency ω was generated as a result of the parametric interaction between the incident acoustic wave at a frequency ω and a magnetic field at 2ω in nonlinear magneto-acoustic media such as α-Fe_2O_3 and hematites. An outstanding feature of this method is that the conversion ratio from the incident wave to the phase conjugate wave is extremely high. The intensity of the phase conjugate wave is often larger than that of the incident wave, and sometimes larger by several 10 dB. Preobrazhensky *et al.* have reported on this method extensively.[18–20]

It should be noted here that, in the nonlinear piezoelectric method and in the nonlinear magneto-acoustic method, operating acoustic frequency can be relatively high (MHz—GHz). This is related to the fact that these interactions occur in solid media, in which acoustic absorption is much smaller than in liquid.

The last category is a purely electrical methods.[21–23] In these methods, no physical interaction was utilized, but instead, electrical circuits and signal processing techniques do all the processing. Usually an array of ultrasonic transmitter/receiver is used. Memorizing the delay time of the transmission of each array and then synthesizing the time-reversed waveform are the key techniques. Although the fidelity of phase conjugate waves is not very high in this method, it has some advantages over other methods: The conversion ratio can be set arbitrarily, and the real-time or nonreal-time operations are selectable. Fink *et al.*[23] extended the function of phase conjugation to nonmonochromatic waves and referred to it as time-reversal mirrors. In these purely electrical methods, operating frequency is relatively low (kHz–MHz).

I. THEORY

A. Principle of acoustic phase conjugation via nonlinear piezoelectricity

In this section, we give an outline of the principle for generating acoustic phase conjugate waves via nonlinear piezoelectricity. The basic explanation on the definition of phase conjugation and its characteristics are available, for example, in Refs. 2 and 3. Here we suppose that we have a medium in which the multiplication of an acoustic field and an electric field occurs. In other words, the medium brings about a parametric interaction between acoustic and electric fields. Suppose an acoustic wave given by

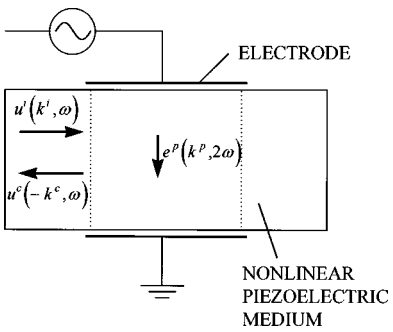


FIG. 1. Basic structure of an acoustic phase conjugator with a nonlinear piezoelectric medium.

$$u^i(r,t)=\tfrac{1}{2}U^i(r)\exp\{i(k^i\cdot r-\omega t)\}+\text{c.c.} \tag{1}$$

is incident on this medium. Here, k^i denotes the wave vector and $U^i(r)$ expresses the amplitude which is assumed to vary slowly compared to the wavelength. The complex conjugate is expressed by c.c. We suppose an electric field given by

$$e^p(r,t)=\tfrac{1}{2}E^p\exp\{i(k^p\cdot r-2\omega t)\}+\text{c.c.} \tag{2}$$

is applied on this medium simultaneously. Here the wave vector is denoted as k^p and the frequency is twice that of the incident acoustic wave. The amplitude E^p is assumed to be spatially uniform. We call this field the pump electric field.

Multiplication of Eqs. (1) and (2) consists of four terms. The cross-term is expressed as

$$u^c(r,t)=\tfrac{1}{4}U^i(r)^*E^p\exp[i\{(k^p-k^i)\cdot r-(2\omega-\omega)t\}]+\text{c.c.} \tag{3}$$

This equation represents a wave of which frequency is ω, the same as that of the incident acoustic wave, and of which amplitude is proportional to $U^i(r)^*$. The wave vector of Eq. (3) is k^p-k^i. Since the electromagnetic velocity c is about 10^5 times larger than any acoustic velocity ν, the pump wave number $|k^p|=2\omega/c$ is negligibly smaller than the acoustic wave number $|k^i|=\omega/\nu$. Therefore, Eq. (3) can be approximated as

$$u^c(r,t)=\tfrac{1}{4}U^i(r)^*E^p\exp\{i(-k^i\cdot r-\omega t)\}+\text{c.c.} \tag{4}$$

The field expressed by Eq. (4) has the same frequency as the incident acoustic wave and the spatial part $U^i(r)^*E^p\times\exp(-ik^i\cdot r)$ is proportional to the complex conjugate of that of the incident wave. Therefore, this field is the phase conjugate wave of the incident acoustic wave.

Figure 1 shows the abovementioned mechanism. The most notable feature of this method is that it does not need any kind of fine adjustments of geometry. In other words, the whole process is automatic. In the phase conjugation by FWM, two pump fields must be aligned so as to cancel their total momentum. In the acoustic phase conjugation mentioned above, on the other hand, the momentum of the pump field is zero with the accuracy of 10^{-5}. In this sense, this method is an approximate method, but its precision is very high. IN FWM, two counterpropagating pump waves must be plane waves or phase conjugate each other. When the wavefronts of the pump waves are deformed by diffraction or some other causes, the deformations are superposed on the phase conjugate wave, and this decreases the fidelity of the

phase conjugate waves to a degree not negligible in some cases. On the other hand, phase conjugate waves generated by the present method have the fidelity which allows only 10^{-5} error, as far as the pump field E^p is spatially uniform. Thus a very high fidelity in generating phase conjugate waves can be achieved.

B. Time grating

The description on the acoustic phase conjugate generation in the preceding section is a phenomenological one. In this section, we give another explanation on this phenomenon that is more understandable from a physical point of view. Here, we invoke a concept of ''time grating.''

The well-known piezoelectric equations of *d*-form are as follows:[24]

$$S = d_{(1)}E + s^E_{(1)}T, \tag{5a}$$

$$D = \epsilon^T_{(1)}E + d_{(1)}T. \tag{5b}$$

Here, S is the strain, D the electric displacement, E the electric field, T the stress, $d_{(1)}$ the linear piezoelectric constant, $s^E_{(1)}$ the linear elastic compliance, and $\epsilon^T_{(1)}$ the linear dielectric constant. Next, we consider how these equations should be modified in nonlinear piezoelectric media. Following Ref. 24, the material equations in such media are described as

$$S = d_{(1)}E + s^E_{(1)}T + \tfrac{1}{2}s^E_{(2)}TT + {}^1d_{(2)}ET + \tfrac{1}{2}{}^2d_{(2)}EE, \tag{6a}$$

$$D = \epsilon^T_{(1)}E + d_{(1)}T + \tfrac{1}{2}{}^1d_{(2)}TT + {}^2d_{(2)}ET + \tfrac{1}{2}\epsilon^T_{(2)}EE. \tag{6b}$$

Here $s^E_{(2)}$ is the second-order nonlinear elastic compliance, $\epsilon^T_{(2)}$ the second-order nonlinear dielectric constant, and ${}^1d_{(2)}$ and ${}^2d_{(2)}$ are the two different second-order nonlinear piezoelectric constants. In above equations, six second-order terms are added to the linear piezoelectric equations. These higher-order terms give rise to various kinds of physical phenomena, including second harmonic generation, frequency summation and subtraction, and so on. However, our interest is limited to the case in which an acoustic field at a frequency ω and an electric field at 2ω are applied to the material. Therefore we take only three fields into consideration: The strain field $S(\omega)$ generated by the incident acoustic wave, the electric field $E(\omega)$ accompanying the incident acoustic field via linear piezoelectric effect, and the pump electric field $E(2\omega)$. With these assumptions, Eqs. (6a) and (6b) are rewritten as

$$S(\omega) = [d_{(1)} + {}^2d_{(2)}E(2\omega)]E(\omega) + [s^E_{(1)} + {}^1d_{(2)}E(2\omega)]T(\omega), \tag{7a}$$

$$D(\omega) = [\epsilon^T_{(1)} + \epsilon^T_{(2)}E(2\omega)]E(\omega) + [d_{(1)} + {}^2d_{(2)}E(2\omega)]T(\omega). \tag{7b}$$

Comparing these equations with the linear equations (5a) and (5b), we notice that they have the same apparent form, but the compliance, the dielectric constant, and the piezoelectric constant are changed to the terms in which the second-order constants multiplied by $E(2\omega)$ are added. The sound velocity in this nonlinear piezoelectric media will thus be modulated at frequency 2ω,

$$\nu(t) = \nu + \Delta\nu \sin(2\omega t), \tag{8}$$

where ν is the sound velocity without the pump electric field, and $\Delta\nu$ is the modulation amplitude in the velocity caused by the application of the pump electric field. In the following, we consider what happens to an acoustic wave that propagates in a medium of which sound velocity is temporally modulated at a frequency twice the acoustic frequency. The wave number of the acoustic field will be temporally modulated as

$$k(t) = k + \Delta k \sin(2\omega t), \tag{9}$$

where $k = \omega/\nu$ and $\Delta k = -k(\Delta\nu/\nu)$. We assume that a plane acoustic wave at ω is incident into this medium. Due to the wave vector modulation, it will be modulated as

$$\begin{aligned}
&\exp\{i(k(t)x - \omega t)\} \\
&= \exp[i\{(k + \Delta k \sin(2\omega t))x - \omega t\}] \\
&= \exp\{i(kx - \omega t)\} \sum_{n=-\infty}^{\infty} J_n(\Delta kx)\exp(in2\omega t) \\
&= J_0(\Delta kx)\exp\{i(kx - \omega t)\} + J_1(\Delta kx)\exp\{i(kx + \omega t)\} \\
&\quad + \text{other terms.}
\end{aligned} \tag{10}$$

In above equation, the term $J_0(\Delta kx)\exp\{i(kx - \omega t)\}$ expresses a wave which has the same frequency and the same propagating direction as the incident wave. This wave is the transmitted part of the incident wave. The second term $J_1(\Delta kx)\exp\{i(kx + \omega t)\}$ indicates the existence of a wave of which frequency is the same as that of the incident wave but the propagation direction is inverted. This can be regarded as the reflected part of the incident wave. Here we should note that two important facts are involved in this process. First, the reflection occurs in a manner to invert the propagating direction completely. Second, this inversion occurs *independent of the propagating direction* of the incident wave. This is understood from the fact that we have made no specific assumption in deriving Eq. (10) with respect to the relation between the propagating direction and the material coordination. From these facts, we can state that the incident plane acoustic wave with any propagating direction will be converted into the completely direction-inverted plane wave. Next problem is what happens to the incident acoustic wave which has an arbitrary spatial distribution. Any propagating waves can be expressed by the superposition of plane waves, i.e., spatial frequency Fourier components. In the nonlinear piezoelectric process, each of these Fourier components is direction-inverted. Therefore, an incident acoustic wave of any spatial distribution will be converted into the correctly backward-propagating wave with the same spatial distribution and the same frequency, independent of the material coordination. This wave satisfies the definition of the phase conjugate wave.

Hereafter we try to explain this phenomenon by the concept of ''grating'' in comparison with FWM. In FWM, three waves are incident on a nonlinear medium: the incident wave, the first pump wave, and the second pump wave. Two pump waves are counterpropagating. The FWM process is often interpreted as follows. First, the incident wave and the

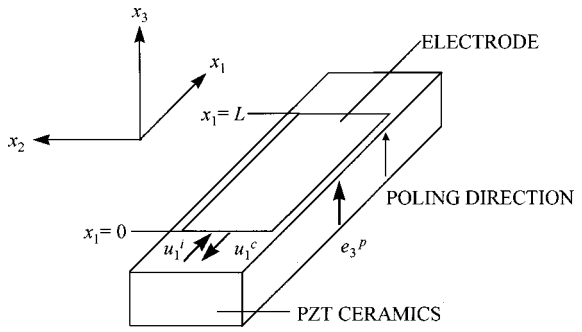


FIG. 2. Configuration of the phase conjugator for the analysis of coupling equations.

first pump wave interfere in the nonlinear medium and create a spatial interference pattern (spatial grating). Next, the second pump wave is incident on this grating and is diffracted. This diffracted wave becomes the phase conjugate wave of the incident wave. When the spatial form of the incident wave changes, the pattern of the spatial grating changes accordingly so as to make the diffracted wave become the precise phase conjugate wave. In this process, the incident wave always ''experiences'' the spatial velocity modulation in the nonlinear medium in which the grating is created. On the other hand, in the acoustic phase conjugation via nonlinear piezoelectricity, a grating is formed in time domain. In this process, the pump electric field is spatially uniform. But it is temporally oscillating at a frequency at 2ω and is modulating the sound velocity of the medium. When the incident acoustic wave propagates in this medium, it also ''experiences'' the velocity modulation. This brings about the same result to the incident wave as what happened to the wave in the spatially modulated medium. Even if the spatial form of the incident wave changes, this ''time grating'' acts exactly in the same way because of its spatial uniformity. Therefore, this ''time grating'' is equivalent to the ''spatial grating'' in FWM. It should be noted that this time grating method is possible only in acoustic phase conjugation, because the ''simultaneous 2ω modulation'' is based on the fact that the velocity of the acoustic wave is negligibly smaller than the velocity of the pump wave.

C. Coupling equations

In this section we derive and solve coupling equations for the amplitudes of the incident wave and the phase conjugate wave. The coupling equations for this phenomenon are extensively studied by Nelson[25] for arbitrary polarizations and wave vector directions. In this paper, we do not give a general expression of the coupling equations, but limit our analysis to the case for the experimental configuration described in Sec. II. Our analysis is further limited to the case that the nonlinear medium is a uniaxially poled PZT ceramics. Analyses for more general cases are available in Ref. 25.

Figure 2 shows the configuration to be analyzed here. A block of PZT cerarnics is placed in a coordinate (x_1,x_2,x_3). We assume that both the incident acoustic wave and the phase conjugate wave are of longitudinal mode propagating along the x_1 direction. The pump electric field is applied in the x_3 direction with the electrodes which extend from $x_1=0$ to L. The poling direction of the PZT ceramics is parallel to the x_3 direction. The nonlinear material equations are written as follows in *d*-form of piezoelectric expression:[26,27]

$$S_1 = d_{31}E_3 + s_{11}^E T_1 + \tfrac{1}{2}s_{111}^E T_1^2 + {}^1d_{311}E_3T_1 + \tfrac{1}{2}{}^2d_{(33)1}E_3^2, \quad (11a)$$

$$D_3 = \epsilon_{33}^T E_3 + d_{31}T_1 + \tfrac{1}{2}{}^1d_{311}T_1^2 + {}^2d_{(33)1}E_3T_1 + \tfrac{1}{2}\epsilon_{3(33)}^T E_3^2. \quad (11b)$$

Here S_1 denotes the strain, D_3 the electrical displacement, and T_1 the stress. Each coefficient has the same meaning as described in Sec. I B. The suffix of the variables and coefficients is expressed by the conventional abbreviating form. The displacement, stress, and the electrical displacement must satisfy the following equations (the equation of motion and the Gaussian equation):

$$\rho \frac{\partial^2 u_1}{\partial t^2} = \frac{\partial T_{11}}{\partial x_1}, \quad (12)$$

$$\frac{\partial D_3}{\partial x_1} = 0, \quad (13)$$

where ρ is the density of the nonlinear medium. We assume three fields in this medium: the incident acoustic wave, the phase conjugate wave, and the pump electric field, respectively, written by

$$u_1^i(r,t) = \tfrac{1}{2}U^i(r)\exp\{i(k^i\cdot r - \omega t)\} + \text{c.c.}, \quad (14)$$

$$u_1^c(r,t) = \tfrac{1}{2}U^c(r)\exp\{i(k^c\cdot r - \omega t)\} + \text{c.c.}, \quad (15)$$

$$e_3^p(r,t) = \tfrac{1}{2}E^p \exp\{i(k^p\cdot r - 2\omega t)\} + \text{c.c.} \quad (16)$$

Here subscripts denote the spatial coordinates and superscripts define the difference of incident, conjugate, and pump waves. We note that the propagating electric field accompanying the incident acoustic wave and the phase conjugate wave are not taken into consideration, because these fields do not contribute to the phase conjugate generation when an uniaxially poled PZT is placed in the configuration shown in Fig. 2.

Substituting Eqs. (14)–(16) into Eqs. (11a)–(13) and comparing the coefficients of the corresponding Fourier components, we have

$$\frac{\partial U^{i*}}{\partial x_1} + \frac{1}{\nu}\frac{\partial U^{i*}}{\partial t} = i\kappa U^c \exp(-i\delta k x_1), \quad (17a)$$

$$\frac{\partial U^c}{\partial x_1} + \frac{1}{\nu}\frac{\partial U^c}{\partial t} = i\kappa^* U^{i*} \exp(-i\delta k x_1), \quad (17b)$$

Here, the second-order derivatives have been neglected (the approximation of slowly varying amplitude), ν is the sound velocity, and κ is the coupling constant given by

$$\kappa = \frac{-\omega\rho\tilde{d}_{311}\nu E^p}{4}, \quad (18)$$

where the effective second-order piezoelectric constant $\tilde{d}_{311}$ is defined by

$$\tilde{d}_{311} = d_{311} - 2s_{11}^E d_{31}, \quad (19)$$

and δk is the phase mismatch:

$$\delta k = \frac{2\Delta\omega}{v}. \tag{20}$$

When all the applied fields are continuous and $\delta k = 0$, the solution of the coupling equations (17a) and (17b) is given by

$$U^{i*}(x_1) = \frac{i|\kappa|}{\kappa^*}\frac{\sin(|\kappa|x_1)}{\cos(|\kappa|L)}U^c(L) + \frac{\cos[|\kappa|(x_1-L)]}{\cos(|\kappa|L)}U^{i*}(0), \tag{21a}$$

$$U^c(x_1) = \frac{\cos(|\kappa|x_1)}{\cos(|\kappa|L)}U^c(L) + \frac{i|\kappa|}{\kappa^*}\frac{\sin[|\kappa|(x_1-L)]}{\cos(|\kappa|L)}U^{i*}(0). \tag{21b}$$

When the phase conjugate wave is null at the end of the interaction region ($x_1 = L$),

$$U^{i*}(L) = U^{i*}(0)\sec(|\kappa|L), \tag{22a}$$

$$U^c(0) = -i|(\kappa/|\kappa|)\tan(|\kappa|L)|U^{i*}(0). \tag{22b}$$

Thus we can define the amplitude reflectivity of phase conjugation r by the ratio of the amplitude of the phase conjugate wave to that of the incident wave at the entrance of the interaction region, $x_1 = 0$. That is,

$$r = \frac{U^c(0)}{U^{i*}(0)} = -i\frac{\kappa}{|\kappa|}\tan(|\kappa|L). \tag{23}$$

Here we note that $r \to \infty$ at $|\kappa|L = \pi/4$, which means the resonance of the system. Practically, r is limited to a finite value because of the absorption and higher order nonlinear effects. In magneto-acoustic phase conjugate generation, this was experimentally observed.[20]

Although the solution of the CW case is very comprehensive, we have to analyze the non-CW case, because the experiments of acoustic phase conjugate generation is usually done in a pulse (tone-burst) mode. When the incident acoustic wave is a rectangle-shaped tone-burst wave with a duration τ, the phase conjugate reflectivity is given by

$$r = -i\frac{\kappa}{|\kappa|}\tan(|\kappa|L) \quad : 2L/v \leqslant \tau, \tag{24a}$$

$$r = \frac{\tau}{2L/v}\tan(|\kappa|L) \quad : \tau < 2L/v. \tag{24b}$$

More general analysis of the non-CW operation is available in Ref. 2.

II. EXPERIMENTS

A. Sound velocity variation

In this section, we describe our experiments on the generation of acoustic phase conjugate waves with several kinds of PZT ceramics. Prior to the generation of phase conjugate waves, we made the measurement of the sound velocity variation in PZT ceramics under electric fields. If the dependence of the sound velocity on the electric field is known, the phase conjugate reflectivity can be estimated using Eqs. (24a) and (24b).

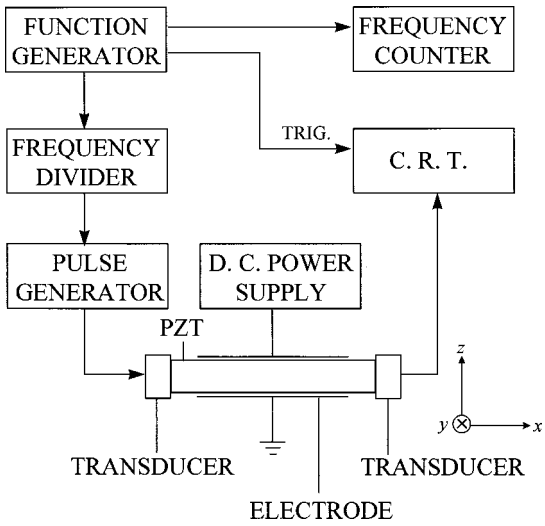


FIG. 3. Experimental configuration for the measurement of the sound velocity variations of PZT ceramics by dc electric fields (static method).

We adopted two different methods to measure the change in sound velocity. The first is the static method, in which the pulse-echo overlapping method was used under the application of dc electric fields. This method is appropriate to roughly estimate the velocity change over a wide range of the applied electric field (from -400 kV m^{-1} to 400 kV m^{-1}). The second is the dynamic method.[27] Each PZT ceramic sample was shaped into a thin plate which vibrates in its fundamental mode. An ac voltage was applied to this plate. Because of the nonlinearity in piezoelectricity, the vibration of the plate is deformed from a sinusoidal curve. The dependence of the sound velocity on electric fields can be estimated by analyzing the spectrum of this nonsinusoidal vibration. This dynamic method is more appropriate to connect the property of the samples to the phase conjugate generation, because the measurement is done under ac electric fields. However, the range of applicable electric fields was smaller (from -6 kV m^{-1} to 6 kV m^{-1}) than the static method. Therefore, we used the data from the static method to roughly estimate the change in sound velocity over a wide range of applied electric fields, and used the data from the dynamic method to precisely determine the sound velocity change.

Figure 3 shows the experimental configuration for velocity measurement in the static method. The size of PZT ceramics was $30(x)\times 10(y)\times 5(z)$ mm. A pair of electrodes of which the length was 20 mm were formed on (001) and $(00\bar{1})$ faces of the sample. The voltage was applied to these electrodes. A pair of ultrasonic transducers to transmit/receive the acoustic waves were attached on both (100) and $(\bar{1}00)$ faces of the sample. Pulse-echo overlapping method was used to measure the sound velocity. The acoustic frequency was 10 MHz. Figure 4 shows the configuration of the dynamic method, in which a spectrum-analyzing circuit was formed to measure the nonsinusoidal vibration of the thin-plate sample.

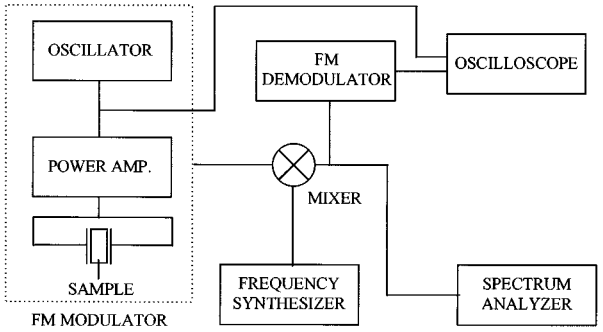


FIG. 4. Experimental configuration for the measurement of the sound velocity variations of PZT ceramics by ac electric fields (dynamic method).

We have measured the sound velocity variation for eight kinds of PZT ceramics (C2, C21, C22, C23, C202, C203, C213, C3). All the samples were produced by Fuji Ceramics Co. and the names (C2, C21....) denote the commercial codes. These samples are different in the composition and in the manufacturing process. They were initially poled in the z direction. The typical experimental results obtained in three samples, C2, C22, and C3, are shown in Fig. 5, in which the abscissa is the applied electric field and the ordinate is the relative sound velocity variation which is defined as $\Delta\nu$ divided by the sound velocity without electric fields. The plus sign means the applied electric field is in the direstion of the initial poling direction. In each figure, the inset is the result of the dynamic method, whereas the background is the result of the static method.

From these results we see the common feature of the sound velocity variation of PZT ceramics. The sound velocity decreases in every sample, when the applied electric field is negative (against the initial poling direction). This can be understood that the external field works to soften the material. The values of the relative sound velocity variation are large in this condition. On the other hand, when the electric field is positive, the variation of the sound velocity differs among samples. The magnitude of the sound velocity variation is generally smaller than in the case of negative electric field. In C2, C21, and C23, the sound velocity variations are positive. In C22, C202, and C203, it seems that the values of sound velocity saturate and decrease at larger electric fields. In C213 and C3, the sound velocity variations are very small in the positive electric field region. All of these features are believed to be related to the difference in electromechanical features of the PZT ceramics, specifically the value of initial poling, although the physical explanation at microscopic level is not available at present.

We give characteristic values of these PZT ceramics samples in Table I. The density ρ, the linear piezoelectric constant d_{31}, and the linear compliance s_{11}^{E} are from catalogue values. The sound velocity ν and the absorption coefficient α were estimated by the experimental values. The nonlinear piezoelectric constants d_{311}, $\tilde{d}_{311}$ were determined from the data of the dynamic measurement method. The phase conjugate coupling constant κ was calculated by Eq. (18). The phase conjugate reflectivity r was estimated by

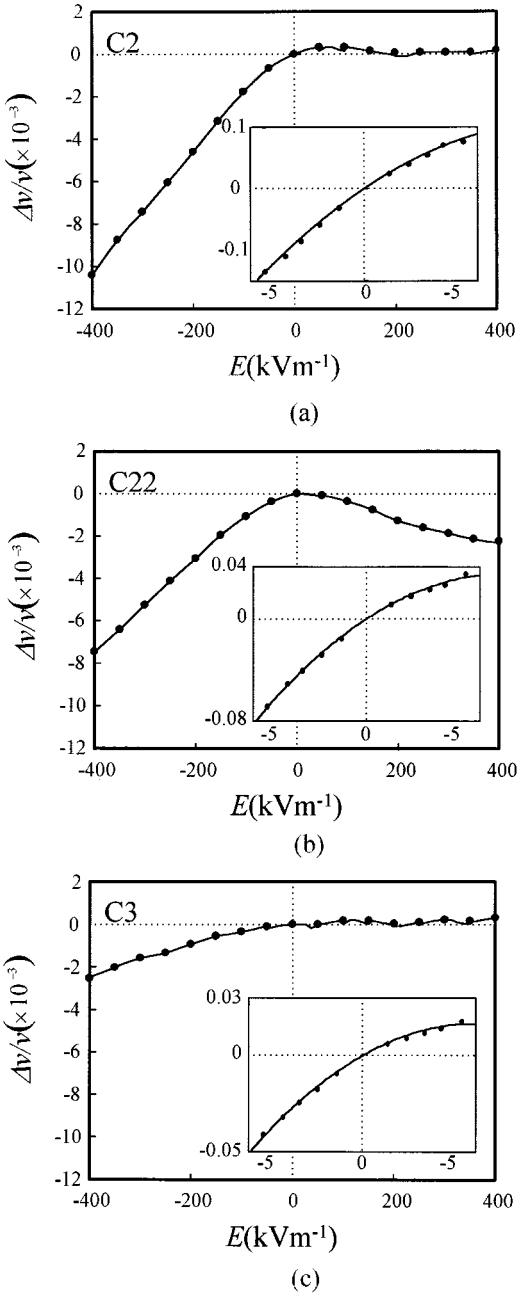


FIG. 5. The sound velocity variations of the three typical PZT ceramics observed under dc and ac (insets) electric fields.

these values with the interaction length $L=20\,\mathrm{mm}$ and the absorption coefficients of samples.

B. Generation of phase conjugate waves

We have generated acoustic phase conjugate waves with PZT ceramics of which characteristics have been studied in the preceding section. Figure 6 shows the schematic block diagram of the experimental setup. A PZT transducer was attached to the end surface of the sample and the experiment was done with the pulse-echo technique. Output of the CW signal source at 10 MHz was gated to a square-shaped pulse with 5 μs duration and drove the transducer to inject the incident acoustic waves into the sample. At the same time, frequency of the driving pulse was doubled to 20 MHz for

TABLE I. Physical properties, linear and nonlinear piezoelectric constants, and the amplitude reflectivity of phase conjugation for eight different materials of PZT ceramics.

	C2	C21	C22	C23	C202	C203	C213	C3	
ρ	7.5	7.85	7.5	7.6	7.85	7.7	7.8	7.5	$\times 10$ kg/m^3
ν	4181	4315	4045	4247	4189	4125	4214	4450	m/s
d_{31}	−120	−130	−140	−100	−130	−145	−135	−57	$\times 10^{-12}$ m/V
d_{311}	−7.21	−5.98	−2.76	−4.94	−3.74	−3.39	−3.57	−1.15	$\times 10^{-19}$ m^3/VN
S_{11}^E	13.3	13.7	14.3	12.0	12.2	13.4	12.2	10.5	$\times 10^{-12}$ m^2/N
$\tilde{d}_{311}$	−7.18	−5.95	−2.72	−4.92	−3.71	−3.35	−3.54	−1.11	$\times 10^{-19}$ m^3/VN
κ	35.4	31.6	12.9	24.9	19.1	16.7	18.3	5.8	m^{-1}
α	0.015	0.012	0.013	0.015	0.013	0.012	0.010	0.013	neper/mm
r	17.5	16.9	6.1	12.1	9.4	8.2	9.6	3.0	%

the pump electric field. The 20 MHz pulse which was amplified and applied to the electrodes gave the pump field of 60 kV m^{-1} peak-to-zero. In order to shift the operation point of the phase conjugate generation, a dc electric field E^b was superposed as a bias electric field. The results of the velocity measurement in Fig. 5 show that the velocity variation depends on the operation point and therefore the higher phase conjugate reflectivity can be obtained by adjusting the bias electric field.

Figure 7 shows the acoustic signals observed by the oscilloscope. Sample C2 was used in this experiment. Figure 7(a) is the signal without the pump electric field. A series of echoes from the other side of the sample is observed. When the pump electric field with the duration 10 μs was applied 2 μs after the transmission of the incident acoustic wave, the received signal changed as shown in Fig. 7(b). An additional pulse indicated by an arrow was observed. This signal is regarded as the phase conjugate wave for the following reasons: (1) the frequency was 10 MHz; (2) it appeared before the first echo; (3) it appeared only when both the incident wave and the pump electric field were on; (4) its amplitude was proportional to the amplitude of the pump electric field; and (5) the amplitude was varied by changing the bias electric field. Figure 7(c) shows the signal when the bias electric field $E^b = -200$ kV m^{-1} was applied. The sign of the bias

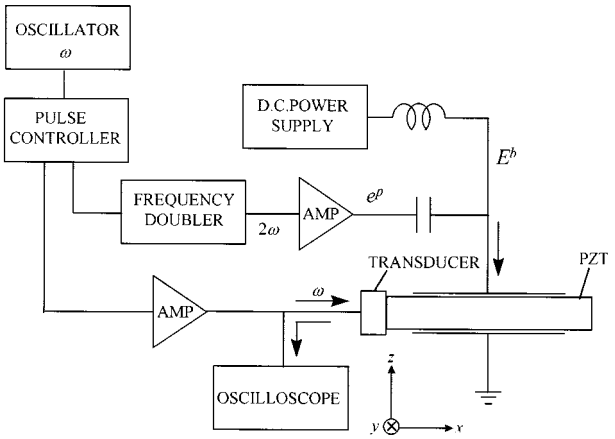


FIG. 6. Experimental configuration for the generation of phase conjugate waves in PZT ceramics.

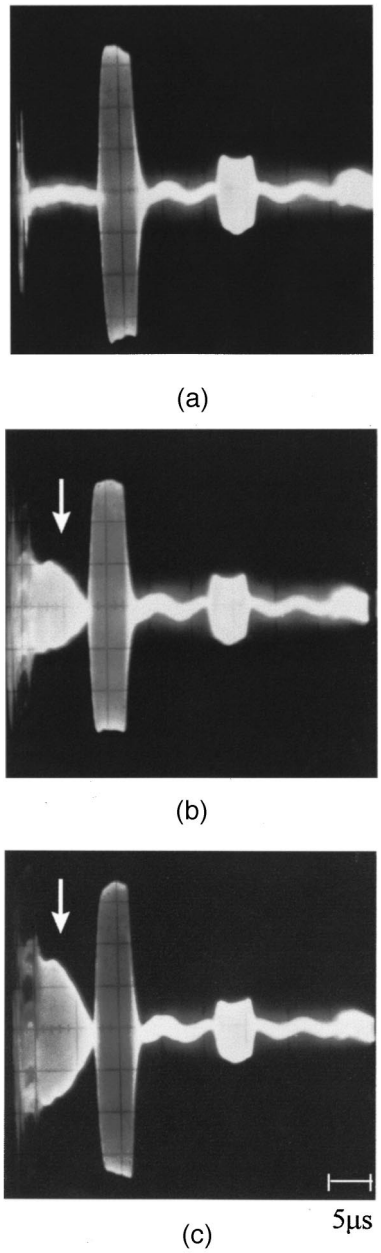


FIG. 7. Output signals in the experiment of Fig. 6 (a) without pump electric field, (b) with a pump electric field of 60 kV m^{-1}, and (c) with a pump electric field of 60 kV m^{-1} together with a bias electric field of −200 kV m^{-1}. Signals indicated by arrows are the phase conjugate waves.

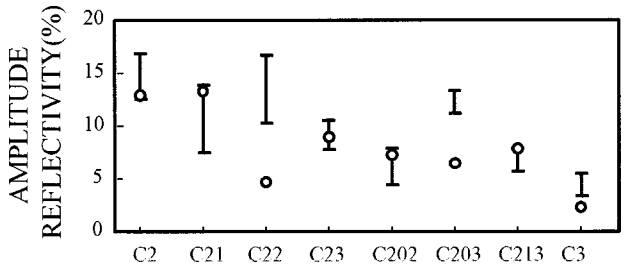


FIG. 8. Amplitude reflectivity of the phase conjugation for eight different PZT ceramics. Solid bars and circles represent the experimental results and the calculated values, respectively.

electric field E^b is taken as positive when it is in the direction of the poling direction of PZT.

In Fig. 8, we show the experimental and theoretical values of amplitude reflectivity of phase conjugation for eight different PZT ceramics. Experimental values were evaluated by the ratio of the amplitude of the phase conjugate signal to the first echo from the other side of PZT. The envelope of the phase conjugate wave becomes a trapezoid for the incident wave of a rectangle envelope for a lossless nonlinear medium.[14] For a lossy nonlinear medium such as PZT, the trapezoid shape gets a small change (decay) due to the sound absorption. In our evaluation, the maximum values of the phase conjugate waves were used together with the calibration by the absorption effect. Theoretical values were calculated using the $\tilde{d}_{311}$ values obtained in the previous section. Both values are in good agreement except for two cases, C22 and C203. In these cases, the pump electric field actually applied might have been larger than our estimation for some kind of resonance effect. The agreement between the experimental and theoretical values in Fig. 8 supports the theoretical treatment on this phenomenon described in Sec. I.

Figure 9 shows the experimental values of the amplitude reflectivity of phase conjugation with and without the bias electric field. As seen in Fig. 5, the gradient of the sound velocity variation with regard to the electric field is large when the electric field is negative in all samples. Therefore we tried to enhance the phase conjugate reflectivity by applying negative bias fields. The solid bars in Fig. 9 are the reflectivity when the bias electric field $E^b = -200\,\mathrm{kV\,m^{-1}}$ was applied, whereas the dotted bars are the values without bias fields. The values of phase conjugate reflectivity were obviously increased by the bias field. The difference in the enhancement effect among samples are due to the difference of electric field-sound velocity curves shown in Fig. 5.

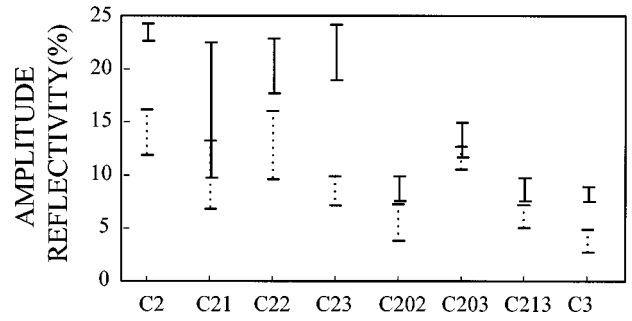


FIG. 9. Experimental values of the amplitude reflectivity of the phase conjugation for eight different PZT ceramics. Solid bars and dotted bars represent the results with the bias fields of $-200\,\mathrm{kV\,m^{-1}}$ and without, respectively.

III. CONCLUSIONS

In this paper, the principle of the acoustic phase conjugation by nonlinear piezoelectricity and the experimental results of the generation of phase conjugate waves at 10 MHz have been described. Acoustic phase conjugate waves at a frequency ω can be generated in a nonlinear piezoelectric medium via parametric interaction between the incident acoustic wave at ω and externally applied pump electric field at 2ω. It has been shown that the physical essence of this interaction is the modulation of the sound velocity by the pump electric field. A concept of time grating has been proposed to explain this interaction comprehensively. The basis which ensures the principle of this interaction is the great difference between the sound velocity and the electromagnetic velocity. Coupling equations for a particular configuration with PZT piezoelectric ceramics have been derived based on the nonlinear material equation in *d*-form. It has been shown that the coupling constant is proportional to the effective second-order piezoelectric constant $\tilde{d}_{311}$ and the amplitude of the pump electric field E^p. The reflectivity of the phase conjugation was described using the coupling constant.

In the experiments, the measurement of sound velocity variation by electric fields and the generation of phase conjugate waves have been achieved. Sound velocity variation has been measured for eight different kinds of PZT ceramics. The common feature of the results was that the sound velocity was greatly reduced when the electric field was applied in the opposite direction to the initial poling direction. The magnitude of the sound velocity variation and some other features were different among samples. The explanation of this difference based on the microscopic physical properties remains for the future work. Generation of phase conjugate waves at 10 MHz has been achieved in every PZT samples. Experimental values of the amplitude reflectivity of phase conjugation were in good agreement with the calculated values with effective second-order piezoelectric constants $\tilde{d}_{311}$. The reflectivity was different among samples, and the maximum value reached to 23%. Our future work is to increase the reflectivity. This will be achieved by searching the materials which have the larger piezoelectric nonlinearity together with the microscopic study of the physical origin of this property.

The confirmation of the fidelity of phase conjugate waves by visualizing the ultrasonic fields and the application of the phase conjugation to scanning ultrasonic imaging systems will be presented in the accompanying paper.

[1] B. Y. Zel'dovich, V. I. Popovichev, V. V. Ragul'skii, and F. S. Faizullov, ''Connection between the wave fronts of the reflected and exciting light in stimulated Mandel'shtam-Brillouin scattering,'' Sov. Phys. JETP Lett. **15**, 109–113 (1972).

[2] R. A. Fisher, Ed., *Optical Phase Conjugation* (Academic, New York, 1983).

[3] B. Ya. Zel'dovich, N. F. Pilipetsky, and V. V. Shkunov, *Principles of Phase Conjugation* (Springer Verlag, Berlin, 1985).
[4] D. V. Vlasov, E. A. Zabolotskaya, and Yu. A. Kravtsov, ''Acoustic phase conjugation in water containing bubbles,'' Sov. Phys. Acoust. **29**, 69–70 (1983).
[5] E. A. Zabolotskaya, ''Phase conjugation of sound beams in connection with four-phonon interaction in a liquid containing gas bubbles,'' Sov. Phys. Acoust. **30**, 462–463 (1984).
[6] F. V. Bunkin, D. V. Vlasov, and Yu. A. Kravtsov, ''Phase conjugation and self-focusing of sound by a nonlinear interction with a liquid surface,'' Sov. Tech. Phys. Lett. **7**, 138–140 (1981).
[7] F. V. Bunkin, D. V. Vlasov, E. A. Zabolotskaya, and Yu. A. Kravtsov, ''Phase conjugation of sound beams in four-phonon interaction with temperature waves,'' Sov. Phys. Acoust. **28**, 440–441 (1983).
[8] T. Sato, H. Kataoka, T. Nakayama, and Y. Yamakoshi, ''Ultrasonic phase conjugator using micro particle suspended cell and its application,'' *Acoustical Imaging*, Vol. 17 (Plenum, New York, 1989), pp. 361–370.
[9] T. Yanase, Y. Watanabe, and Y. Urabe, ''Acoustic phase conjugation mirror using microcapsules,'' J. Acoust. Soc. Jpn. **51**, 221–226 (1995) (in Japanese).
[10] L. O. Svaasand, ''Interaction between elastic surface waves in piezoelectric materials,'' Appl. Phys. Lett. **15**, 300–302 (1969).
[11] R. B. Thompson and C. F. Quate, ''Nonlinear interaction of microwave electric fields and sound in $LiNbO_3$,'' J. Appl. Phys. **42**, 907–919 (1971).
[12] A. P. Brysev, F. V. Bunkin, D. V. Vlasov, and Yu. E. Kazarov, Sov. Phys. Tech. Lett. **8**, 237 (1982).
[13] V. I. Reshetzky, ''Phase conjugate reflection and amplification of a bulk acoustic wave in piezoelectric cryatals,'' J. Phys. C **17**, 5887–5891 (1984).
[14] M. Ohno, ''Generation of acoustic phase conjugate waves using nonlinear electroacoustic interaction in $LiNbO_3$,'' Appl. Phys. Lett. **54**, 1979–1980 (1989).
[15] N. S. Shiren, R. L. Melcher, and T. G. Kazyaka, ''Multiple-quantum phase conjugation in microwave acoustics,'' IEEE J. Quantum Electron **QE-22**, 1457–1460 (1986).
[16] M. Ohno and K. Takagi, ''Acoustic phase conjugation in highly nonlinear PZT piezoelectric ceramics,'' Appl. Phys. Lett. **64**, 1620–1622 (1993).
[17] M. Ohno and K. Takagi, ''Enhancement of the acoustic phase conjugate reflectivity in nonlinear piezoelectric ceramics by applying static electric or static stress fields,'' Appl. Phys. Lett. **69**, 3483–3485 (1996).
[18] A. P. Brysev, F. V. Bunkin, D. V. Vlasov, L. M. Krutianskii, V. L. Preobrazhenskii, and A. D. Stakhovskii, ''Parametric phase conjugation of an ultrasonic wave in a ferrite,'' Sov. Phys. Acoust. **34**, 642–643 (1989).
[19] A. P. Brysev, F. V. Bunkin, D. V. Vlasov, A. D. Stakhovsky, V. N. Streltsov, L. M. Krutiansky, V. L. Preobrazhensky, Yu. V. Pyl'nov, and E. A. Ekonomov, ''Some recent findings on acoustic phase conjugation in parametric media,'' Opt. Acoust. Rev. **1**, 107–120 (1989).
[20] L. M. Krutiansky, V. L. Preobrazhensky, Yu. V. Pyl'nov, A. P. Brysev, F. V. Bunkin, and A. D. Stakhovsky, ''Observation of ultrasonic waves in liquid under overthreshold parametric phase conjugation in ferrite,'' Phys. Lett. A **164**, 196–200 (1992).
[21] M. Nikoonahad and T. L. Pusateri, ''Ultrasonic phase conjugation,'' J. Appl. Phys. **66**, 4512–4513 (1989).
[22] D. R. Jackson and D. R. Dowling, ''Phase conjugation in underwater acoustics,'' J. Acoust. Soc. Am. **89**, 171–181 (1991).
[23] M. Fink, ''Time reversal of ultrasonic fields—Part I: Basic principles,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 555–566 (1992).
[24] B. A. Auld, *Acoustic Fields and Waves in Solids*, Vols. I and II (Krieger, Florida, 1990).
[25] D. F. Nelson, ''Three-field electroacoustic parametric interactions in piezoelectric cryatals,'' J. Acoust. Soc. Am. **64**, 891–895 (1978).
[26] Y. Cho and J. Wakita, The transactions of IEICE, **J76-A**, 108–117 (1993) (in Japanese).
[27] Y. Cho and F. Matsuno, Rev. Sci. Instrum. **64**, 1244 (1993).

Acoustic phase conjugation by nonlinear piezoelectricity. II. Visualization and application to imaging systems

Ken Yamamoto, Masahiro Ohno, Akira Kokubo, Keiji Sakai, and Kenshiro Takagi
Institute of Industrial Science, University of Tokyo, 7-22-1 Roppongi, Minato-ku, Tokyo 106, Japan

(Received 4 February 1998; revised 25 January 1999; accepted 1 June 1999)

Phase conjugate waves of ultrasound were generated in PZT ceramics through nonlinear piezoelectric interaction between an incident ultrasonic field at ω and an electric field at 2ω. The amplitude reflectivity of the phase conjugator was 23% at 10 MHz. The behavior of the incident waves and the phase conjugate waves were visualized by stroboscopic schlieren technique. Time-reversal property and the automatic correction of wavefront distortion in the phase conjugate process were confirmed. A scanning ultrasonic imaging system with a PZT phase conjugator has been built. This system was used to visualize test samples composed of solid plates and phase disturbers. Images by phase conjugate reflection yielded clear figures of the solid plates in spite of the existence of phase disturbers made of agarose gel with rough surfaces, whereas conventional images showed serious distortion. © *1999 Acoustical Society of America.*
[S0001-4966(99)02809-X]

PACS numbers: 43.35.Sx [HEB]

INTRODUCTION

It is almost a quarter century since a term ''phase conjugation'' was given to some phenomena of the retrogressive behaviors in wave propagation, though essentially the same concept had been found in the stimulated Brillouin scattering[1] and some kinds of holography. This concept has attracted much attention since it was shown that the time-reversal property in this phenomenon could be applied to improve the characteristics of imaging systems. The wavefront distortion suffered in the propagation through an inhomogeneous medium can be removed automatically in phase undoing in the path of phase conjugate waves.

The study first began in optics and various experiments have been made. From a viewpoint of imaging application, the most successful would be the four-wave mixing through the strong photorefractive effect in a single crystal of $BaTiO_3$. In typical experiments, wavefront distortion was given, for example, by a lump of glass with an irregular shape to a laser beam conveying an image, then it was removed in the phase conjugate path to reconstruct the initial image. The investigation of acoustic phase conjugation started a few years later, and various methods of phase conjugation have been reported, including the liquid surface holography,[2–4] holograms written in polystyrene latex,[5] and nonlinear elastic response of water with many micro bubbles.[6,7] Some solid state devices with piezoelectric[8] or magneto-elastic[9] materials have also been made. Nevertheless, the acoustic phase conjugation seems to be on its first stage and the application to imaging systems has not been fully tried with one exception.[10]

In the accompanying paper, we have described the principle and the basic operation of an acoustic phase conjugator made with PZT ceramics. The nonlinear piezoelectricity of the material causes three-wave mixing, in which one is the electric field and others are the incident and the phase conjugate waves. The experiments have shown the phase conjugate amplitude reflectivity of the order of 10^{-1} and the real-time response. These properties are suitable for the application to scanning ultrasonic imaging systems.

The purpose of this paper is twofold: First, to visually confirm the feasibility of the time-reversal property and the automatic correction of phase distortion in the phase conjugate process via nonlinear piezoelectricity. We utilized a stroboscopic schlieren technique[11] for this purpose. Secondly, to introduce a scanning ultrasonic imaging system incorporating a phase conjugator, which owns a novel characteristic compared to the conventional ultrasonic imaging systems.

I. VISUALIZATION OF THE PHASE CONJUGATE PROPAGATION

In this section, we give the experimental results that have visually confirmed the time-reversal and the aberration-correcting processes in phase conjugation by nonlinear piezoelectric interaction. Figure 1 shows the experimental setup. The phase conjugator is a rectangular block of PZT ceramics (C2) $30\times10\times5$ mm in size with a pair of electrodes deposited on side faces. It has been initially poled along the direction from one electrode to the other. This PZT block was fixed at the bottom of a water tank in a manner that only the top surface was in touch with water. An anti-reflection layer made of SiO_2 with the thickness of 150 μm was deposited on the aperture surface in order to increase the transmissivity of the ultrasound. Although the acoustic impedance of SiO_2 ($13\times10^6\,\mathrm{kg\,s^{-1}\,m^2}$) was not the ideal value ($7\times10^6\,\mathrm{kg\,s^{-1}\,m^2}$) for an anti-reflection layer between water and PZT, this layer has enhanced the transmissivity up to 25%. Ultrasonic tone-burst waves at a center frequency of 10 MHz and a duration of 10 μs are radiated from a transducer into water and are incident on the top surface of PZT. Pump electric fields at 20 MHz, twice the ultrasonic frequency, are applied to the electrodes while the ultrasonic waves travel inside the PZT. The amplitude of the pump electric field was 60 $\mathrm{kV\,m^{-1}}$ peak to peak. As a result of the parametric inter-

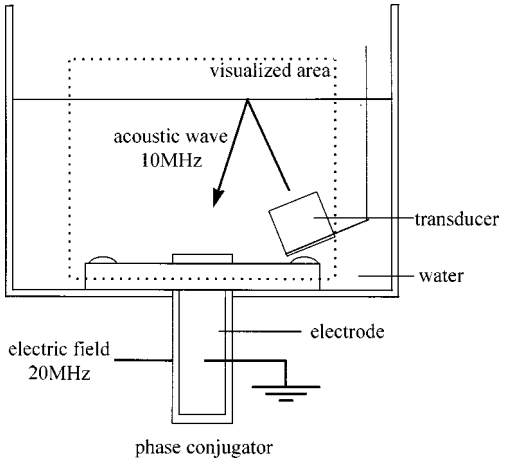


FIG. 1. Experimental setup to visualize the acoustic phase conjugate wave in water by schlieren technique. The visualized area is shown by the dotted square.

action via nonlinear piezoelectricity, phase conjugate waves at 10 MHz are generated in the region between the electrodes. A dc bias field of 200 kV m^{-1}, of which polarity is inverse of the PZT's initial poling, was applied together with the pump electric field. This field works to shift the operation point to the range where the piezoelectricity has stronger nonlinearity.[12] Under these conditions, the amplitude reflectivity of the phase conjugator was 23%.

The system described above was set in a schlieren optics, as shown in Fig. 2, to visualize the behavior of ultrasonic waves in the area indicated by the dotted square in Fig. 1. The light source was a xenon flashlamp of which radiating duration was 180 ns. The water tank was put where the light was expanded and collimated by a system of lenses. The ultrasonic pulses were radiated into water at a repetition rate of 150–200 Hz. The ultrasonic fields diffract the light into several orders through Raman–Nath diffraction. These lights were focused to give several spots in accordance with the order of diffraction. Only the + first-order light was detected through a slit placed at the focal plane. The xenon lamp was triggered with a certain delay time after the radiation of ultrasonic waves. If the delay time is fixed, a still image of the ultrasonic field is obtained. The ultrasonic wave propagates in water for 270 μm during the radiation of light (180 ns). Since this is larger than the ultrasonic wavelength in water at 10 MHz, 150 μm, what is obtained in our system is the averaged amplitudes of ultrasound, not each wavefront. If the delay time is increased continuously, moving pictures of the ultrasound can be obtained.

First, we visualized the simplest case in which a pulse of a plane wave with 10 μs duration is incident on the PZT at

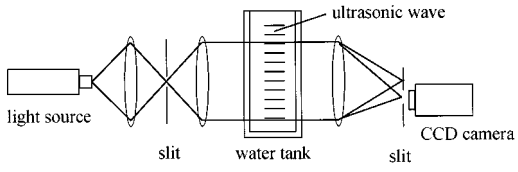


FIG. 2. Experimental setup for the schlieren visualization of acoustic fields.

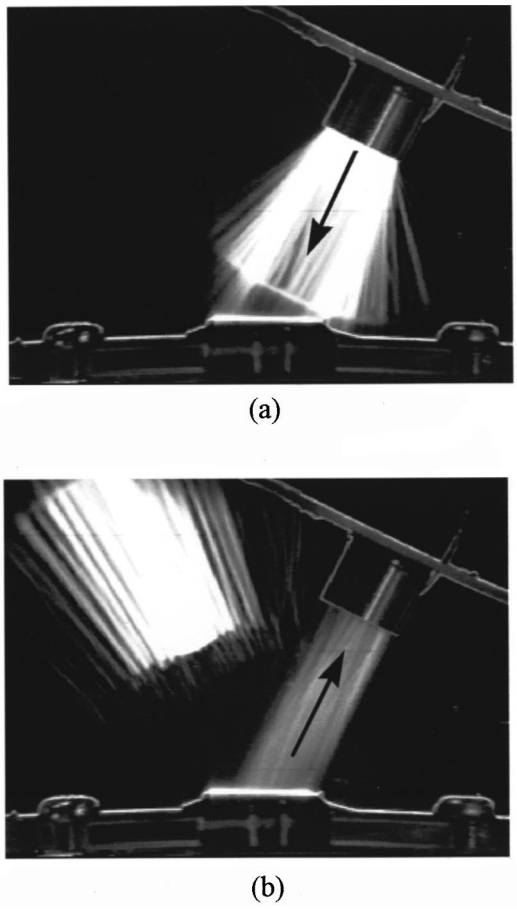


FIG. 3. Schlieren images of (a) the incident wave traveling toward the phase conjugator, and (b) the phase conjugate wave coming back to the transducer. The bright portion in the left is the pulse reflected at water/PZT interface.

an incident angle of 30°. Figure 3 shows the result. Figure 3(a) is the schlieren image of the incident wave. Figure 3(b) shows the schlieren image taken 30 μs after Fig. 3(a). Besides the specularly reflected wave which travels toward the upper left, one can observe a phase conjugate wave which emerges from the top surface of PZT and travels toward the transducer. The path of the phase conjugate wave seems to be the same as that of the incident wave. However, its amplitude is smaller and its duration are longer compared to those of the incident wave. The former is the result of the small phase conjugate reflectivity (less than unity), and the latter is caused by the fact that the phase conjugate wave is generated in every point in the active area of PZT, which has a finite volume.

In the second experiment, a pulse of a plane wave was once emitted toward the water surface, then its reflection was taken into the phase conjugator. Figure 4 shows the experimental results. Figure 4(a) is the plane wave emitted from the transducer. Figure 4(b) shows the wave reflected by the water surface. In this experiment, the water surface was perturbed by a small glass rod. This caused moving deformation on the water surface and the ultrasonic wave reflected by it was deformed substantially. This can be observed as a nonuniform amplitude distribution in Fig. 4(b). Figure 4(c) shows the schlieren image taken 40 μs after image (b), in which a phase conjugate wave is just arriving at the initial

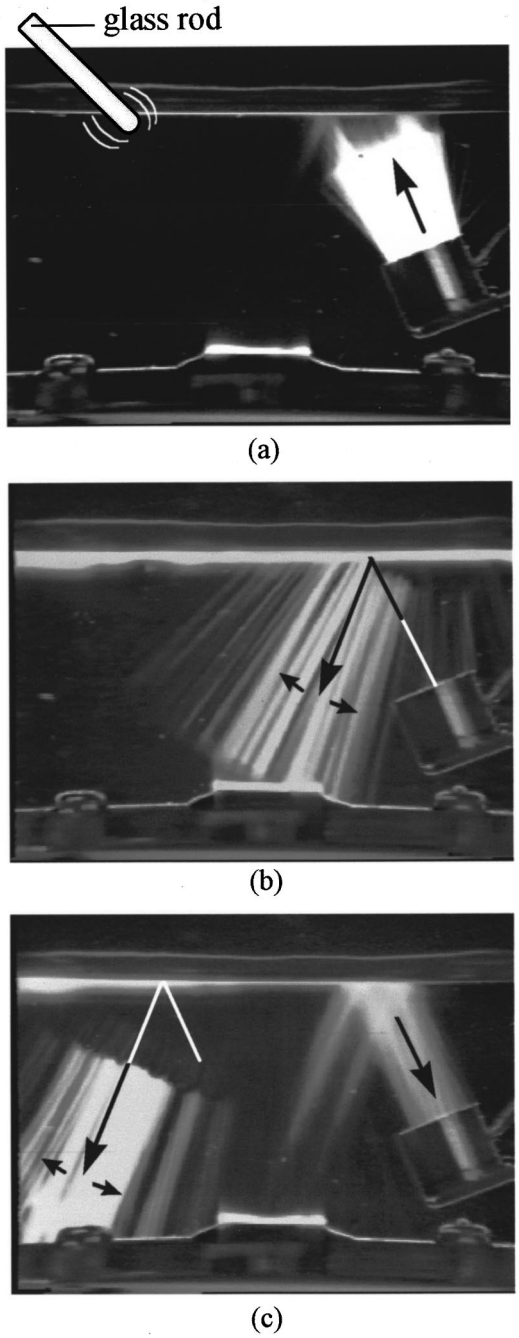


FIG. 4. Images of (a) the incident wave going toward the water surface, (b) the wave once reflected at water surface disturbed by a grass rod, and (c) the phase conjugate wave coming back to the transducer. This bright portion in the left is the pulse twice reflected at the water surface.

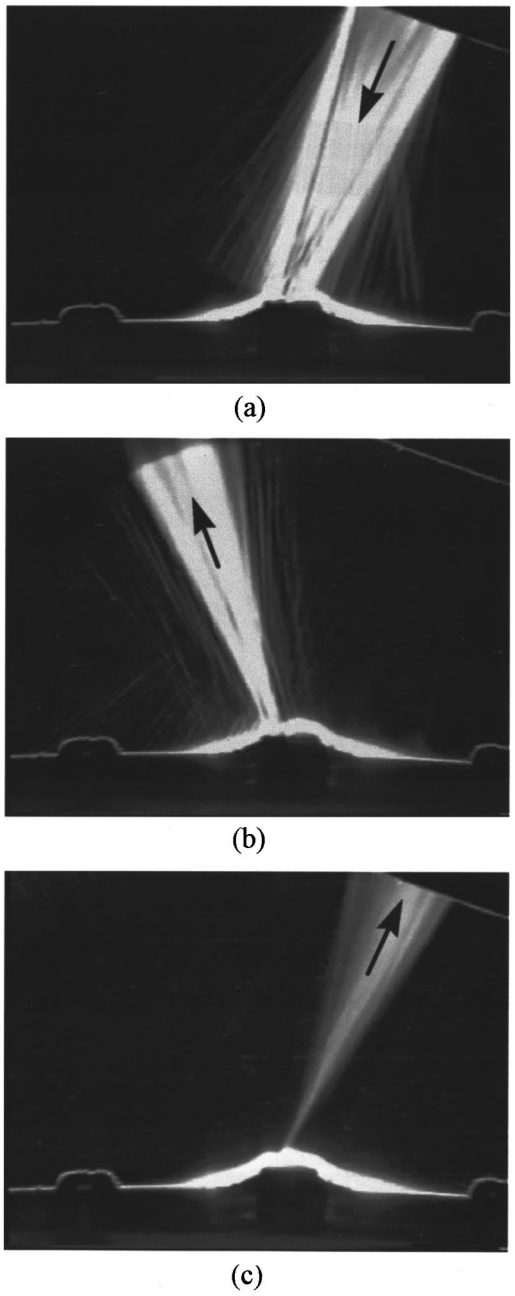


FIG. 5. Images of (a) the incident wave focusing on the phase conjugator, (b) the specularly reflected wave from the top surface, and (c) the phase conjugate wave which has the same amplitude distribution as (a), although the intensity is much different.

transducer. In this picture, the amplitude distribution of the phase conjugate wave is not perturbed as in (b). It seems to restore its initial smooth distribution. This is the result of wavefront recovery in a phase-conjugate roundtrip with a reflector which moves very slowly compared to the acoustic propagation. The speed of the surface motion given by the glass rod is in the order of 10 ms at the fastest, while the roundtrip time of each ultrasonic pulse is shorter than 50 μs. This means the water surface is frozen and the phase conjugate wave is reflected by the surface of the same shape.

In the third experiment, we have tested the case of focused incidence. A PZT transducer with a cylindrical concave curvature was used to emit line-focused ultrasonic beams. Its focal length was 30 mm and its diameter was 10 mm. Line focusing has been used because it provides larger diffraction efficiency in schlieren visualization than point focusing. Figure 5(a) shows the incident pulse arriving at the phase conjugator. The beam is focused on the top surface of the phase conjugator. Figure 5(b) is the image taken 20 μs after (a) in which the specular reflection of the incident wave is observed. Figure 5(c) shows the phase conjugate wave which appeared 20 μs after (b). It is observed that the phase conjugate wave propagates to the transducer as a diverging wave with its focus corresponding to that of the incident wave. Although the amplitude level is much lower, the phase

conjugate wave seems to have the same amplitude distribution as that of the incident wave. The reason the focus looks broader in the incident wave than in the phase conjugate wave is that the intensity is much higher in the former. The existence of dark areas in the image of the incident wave is attributed to the nonlinear relation between the ultrasonic intensity and the image brightness in schlieren visualization. Considering these facts, one can conclude that the fidelity of the time-reversed wave for the case of focused incidence is fairly good in the phase conjugation of our system. This result gives the fundamental basis of the ultrasonic imaging system incorporating a phase conjugator, which is discussed in Sec. II.

Lastly, we note here that we have recorded slow motion pictures of these three processes, which have shown the whole travel from the start of the incident pulse to the return of the phase conjugate wave, providing us with the visual confirmation of the time-reversal property.

II. APPLICATION TO SCANNING IMAGING SYSTEMS

The final aim of our study is to give a basis for the application of phase conjugators to ultrasonic imaging systems. Ultrasonic imaging is widely used to visualize internal structures of optically opaque samples. Conventional ultrasonic imaging systems provide reliable sub-surface images for samples with flat surfaces. However, if the sample has a nonflat surface, the image is degraded because the ultrasonic probe beam is refracted differently from point to point. This gives rise to a problem that the surface image is superposed on the internal image. Similar degradation is observed in the samples of which acoustic velocity is inhomogeneous throughout the visualizing volume. All of these defects are attributed to the distortion in the wavefront of ultrasonic probe beams. The ability of automatic wavefront correction in phase conjugation can be applied to improve these ultrasonic images. In order to demonstrate this, we have built a scanning ultrasonic imaging system shown in Fig. 6. This figure shows the configuration of the ''transmission type.'' Its basic operation is the same as that of the conventional scanning imaging system. Ultrasonic tone-burst beams with 10 MHz carrier frequency and 5 μs duration are emitted from a focusing transducer with 30 mm focal length and 6 mm diameter. This incident beam is radiated to the sample, and its transmitted part is taken into a phase conjugator, which is fixed at the bottom of the water tank. The sample is scanned in a raster manner. All the temporary operations, i.e., the emission of the incident beams, the application of the pump electric fields for the phase conjugator, and the detection of the signals, are in sequence with the positioning signals from the *XY*-scanner. This transmission configuration can easily be changed to ''reflection-type'' configuration. Figure 7 illustrates the arrangement of the transducer, sample, and the phase conjugator for these two different configurations.

In this imaging system, the phase conjugator works as follows. If the surface of the sample is not flat, the ultrasonic wave suffers the distortion in its wavefront. When this distorted wave is incident on the phase conjugator, its phase conjugate wave is generated in real time. This phase conjugate wave has the same distortion in wavefront as the incident wave, but travels in the completely opposite direction. After passing through the sample in its backward path, the distortion in the wavefront of the phase conjugate wave is automatically removed. Thus the transducer detects ultrasonic waves which always have the same wavefront, independent of the shape of the sample surface. In this sense, this system is insensitive to the surface roughness of the sample. However, the attenuation of the ultrasonic wave is not restored even in this process. If the incident wave is blocked or attenuated by some possible structures inside the sample, such as flaws, delaminations, or air bubbles, phase conjugate waves will be weakened accordingly. Therefore, the resultant image will show the attenuating structure of the sample without the influence of the surface structure. By using a phase conjugator, internal structures can be distinguished from surface structures and can be observed clearly.

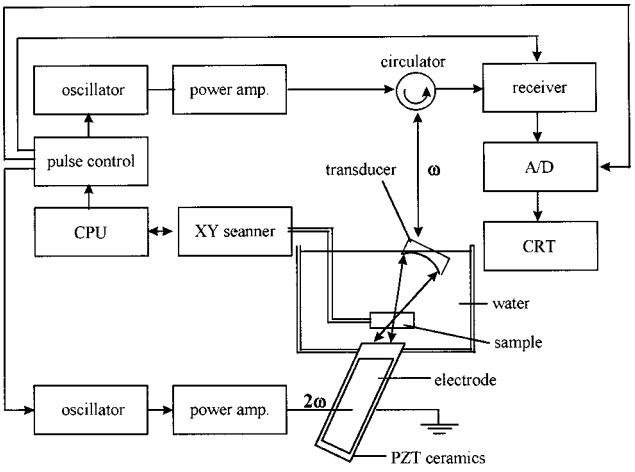


FIG. 6. Schematic diagram of the scanning imaging system with a phase conjugator.

Figure 8 shows the phase conjugator installed in this imaging system. The material is PZT ceramics (C2) described in the former section. This conjugator has a ''funnel-like'' shape. This is the result of designing to realize both large aperture area and large pump electric field. The aperture from which the ultrasonic wave enters has a size of 12×7.5 mm. Anti-reflection layers of polyethylene tereph-

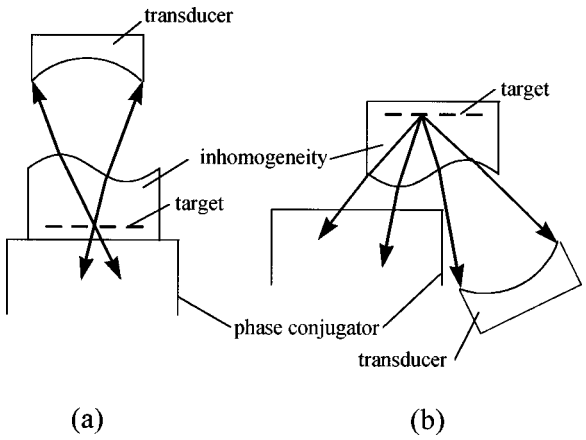


FIG. 7. Two different configurations of imaging, transmission type (a) and reflection type (b).

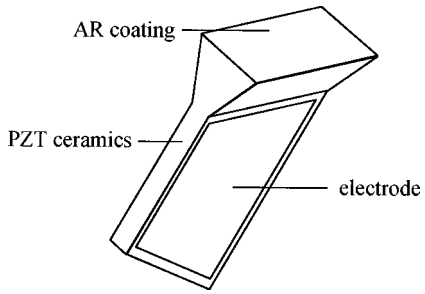


FIG. 8. The shape of the phase conjugator used for imaging.

thalate and fused quartz (SiO_2) are deposited on this surface. The body of the conjugator, on which the pump electric field is applied, has a thickness of 2.5 mm, which is much thinner than the aperture part. This makes a large pump electric field possible for the limited amplitude of the 2ω signal and thus leads to the higher phase conjugate reflectivity. This phase conjugator was designed for the ultrasonic incident angle of approximately 30° so that the specular reflections from the top surface of the phase conjugator and/or the surfaces of the samples could be avoided. The top surface of the conjugator was cut slantingly at 30° so that the incident ultrasonic waves could be guided into the interaction region with the least numbers of reflection.

In Figs. 9–12 are shown the structure of the samples and the experimental results of ultrasonic imaging. The samples were made of agarose gel and targets embedded inside. The agarose gel was made from NaCl solution so as to be given acoustic velocity larger than that of water. Its value was 1650 ms^{-1} for a 15% solution of NaCl and 1700 ms^{-1} for 20%. The difference in acoustic velocity causes refraction at the water/agarose interface. The top surface of agarose was made to have undulation so that the refraction angle differs

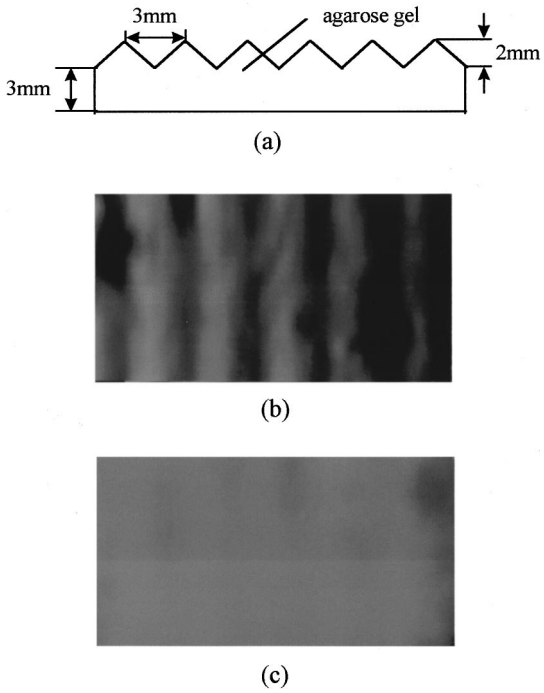


FIG. 9. (a) Cross section of the sample. (b) Conventional image. (c) Phase conjugate image.

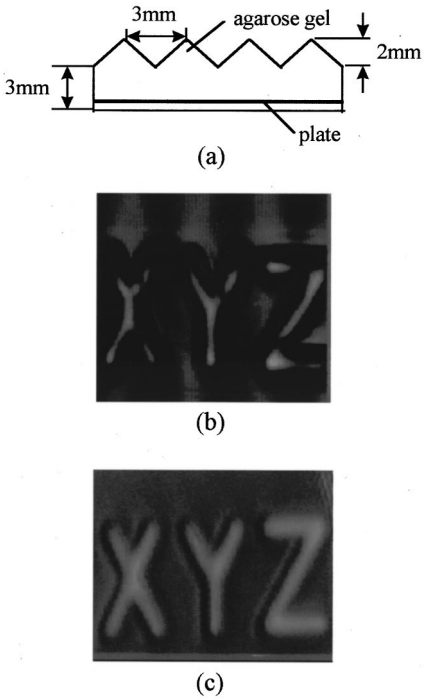


FIG. 10. (a) Cross section of the sample. (b) Conventional image. (c) Phase conjugate image.

from point to point. This causes the distortion in wavefronts of the incident ultrasonic waves. We have taken ultrasonic images of these samples in two different manners. One is the conventional method, in which the simply transmitted or re-

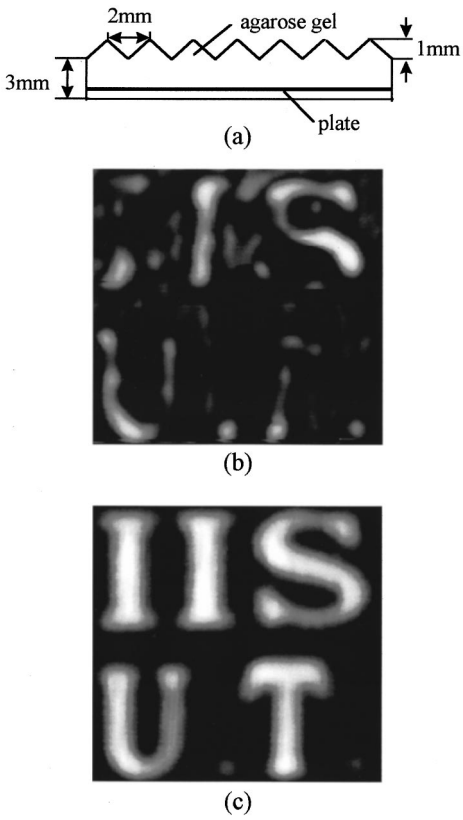


FIG. 11. (a) Cross section of the sample. (b) Conventional image. (c) Phase conjugate image.

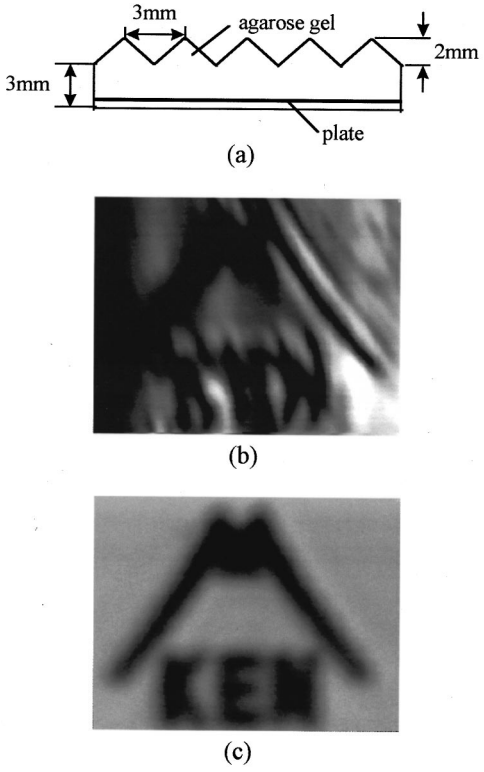


FIG. 12. (a) Cross section of the sample. (b) Conventional image. (c) Phase conjugate image.

flected ultrasonic waves were detected. The other is to detect the phase conjugate waves in a round-trip configuration.

Figure 9 shows the result of the transmission imaging of the first sample. The agarose had a corrugated surface with notches, as illustrated in (a), with the acoustic velocity of 1650 ms^{-1}. This sample had no target inside. The conventional image and the phase conjugate image are shown in (b) and (c), respectively. The dark and bright fringes are observed in (a). This is due to the refraction of the ultrasonic waves at the undulated surface. A substantial part of the incident ultrasonic wave was deflected and was not received by the transducer. This pattern almost disappeared in Fig. 9(c). This experimental result ensures that the phase conjugate image is insensitive to the surface roughness of the sample.

Figure 10 shows the experimental results for a sample which is composed of an agarose disturber and a target embedded inside. The shape and the acoustic velocity of the agarose were the same as those in Fig. 9. The target was made of a plastic plate of 1 mm thick with holes of letters ''*XYZ*.'' Ultrasonic images were taken in the transmission configuration. In the conventional image (b), a stripe pattern due to the surface shape of agarose are superposed on the letter *XYZ*, while they are absent in the phase conjugate image (c). In this experiment, it was confirmed that the inner target was successfully visualized without the influence of the surface roughness in the phase conjugate image.

Figure 11 shows the results for another sample in the transmission configuration. The acoustic velocity of agarose was 1700 ms^{-1} for this case. A pattern of a pyramid array was formed on its surface. Each pyramid was 2 mm in length and 1 mm in height. The target embedded was a 10-μm-thick plate of stainless steel with holes of letters ''I I S U. T.'' The conventional image (b) is seriously destroyed because of the surface roughness of agarose and letters are hardly recognizable. On the other hand, the letters are clearly visualized in the phase conjugate image (c).

Figure 12 shows results for the reflection configuration. The disturber was agarose gel which had the same acoustic velocity and the surface shape as those used in Fig. 9. The target inside was a metal plate with holes of logo ''Mountain-KEN.'' Comparing the conventional image (b) and the phase conjugate image (c), one can see that the influence of the wavefront distortion is inevitable in the former, whereas it is compensated in the latter.

III. CONCLUDING REMARKS

Ultrasonic imaging is commonly used in industry for detecting inner defects such as flaws, cracks, and foreign substances in a nondestructive manner. The scanning acoustic microscope, for example, emits ultrasonic beams onto the sample and receives the reflected signals through a liquid couplant. If the surface of the sample is not flat, image distortion or overlapping of surface images take place. In this paper, we have demonstrated that a phase conjugator could effectively be used to solve these problems. The automatic correction of wavefront distortion in a round-trip configuration with a phase conjugator has worked to eliminate the image distortion and to visualize the inner structures more clearly. The phase conjugator of the present study has the amplitude reflectivity of 50%, which is sufficient for the purpose of imaging. The optimization of the anti-reflection coating for the phase conjugator will improve its total function.

In this paper, we have also investigated the property of phase conjugate waves by visualizing the ultrasonic waves by schlieren technique. From the still or the moving pictures of the incident and the phase conjugate waves, it was confirmed that the time-reversal of the incident wave and the automatic correction of the wavefront distortion really worked in the phase conjugation process. This gives the fundamental basis for the operation of the ultrasonic imaging system with a phase conjugator discussed above.

[1] B. Y. Zel'dovich, V. I. Popovichev, V. V. Ragul'skii, and F. S. Faizullov, ''Connection between the wave fronts of the reflected and exciting light in stimulated Mandel'shtam-Brillouin scattering,'' Sov. Phys. JETP Lett. **15**, 109–113 (1972).

[2] F. V. Bunkin, D. V. Vlasov, and Yu. A. Kravstov, ''Phase conjugation and self-focusing of sound by a nonlinear interaction with a liquid surface,'' Sov. Tech. Phys. Lett. **7**, 138–140 (1981).

[3] N. P. Andeeva, F. V. Bunkin, D. V. Vlasov, and K. Karshiev, ''Experimental observation of acoustic phase conjugation at a liquid surface,'' Sov. Tech. Phys. Lett. **8**, 45–46 (1982).

[4] F. V. Bunkin, D. V. Vlasov, K. Karshiev, and D. Stakhovskii, ''Experimental observation of the suppression of a wave field by means of a phase-conjugating (wavefront-reversing) mirror (PCM),'' Sov. Phys. Acoust. **31**, 80–81 (1985).

[5] T. Sato, H. Katatoka, and Y. Yamakoshi, J. Acoust. Soc. Jpn. **44**, 122–129 (1988) (in Japanese).

[6] D. V. Vlasov, E. A. Zabolotskaya, and Yu. A. Kravtsov, ''Acoustic phase conjugation in water containing bubbles,'' Sov. Phys. Acoust. **29**, 69–70 (1983).
[7] E. A. Zabolotskaya, ''Phase conjugation of sound beams in connection with four-phonon interaction in a liquid containing gas bubbles,'' Sov. Phys. Acoust. **30**, 462–463 (1984).
[8] V. I. Reshetzky, ''Phase conjugate reflection and amplification of a bulk acoustic wave in piezoelectric crystals,'' J. Phys. C **17**, 5887–5891 (1984).
[9] A. P. Brysev, F. V. Bunkin, D. V. Vlasov, L. M. Krutyanskii, V. L. Preobrazhenskii, and A. D. Stakhovskii, ''Parametric phase conjugation of an ultrasonic wave in a ferrite,'' Sov. Phys. Acoust. **34**, 642–643 (1988).
[10] M. Ohno, ''An acoustic imaging system using phase conjugate waves,'' Jpn. J. Appl. Phys., Suppl. **29**, 299–301 (1989).
[11] K. Yamamoto, A. Kokubo, M. Ohno, K. Sakai, and K. Takagi, ''Nonlinaer piezoelectricity of PZT ceramics and acoustic phase conjugate waves,'' Jpn. J. Appl. Phys. **35**, 3210–3213 (1996).
[12] M. Ohno and K. Takagi, ''Enhancement of the acoustic phase conjugate reflectivity in nonlinear piezoelectric ceramics by applying static electric or static stress fields,'' Appl. Phys. Lett. **69**, 3483–3485 (1996).

Lamb wave characterization of the effects of long-term thermal-mechanical aging on composite stiffness

Michael D. Seale[a] and Eric I. Madaras
NASA Langley Research Center, Mail Stop 231, Hampton, Virginia 23681

(Received 4 June 1998; revised 4 February 1999; accepted 1 June 1999)

Lamb waves offer a promising method of evaluating damage in composite materials. The Lamb wave velocity is directly related to the material parameters, so an effective tool exists to monitor damage in composites by measuring the velocity of these waves. The Lamb Wave Imager™ (LWI) uses a pulse/receive technique that excites an antisymmetric Lamb mode and measures the time-of-flight over a wide frequency range. Given the material density and plate thickness, the bending and out-of-plane shear stiffnesses are calculated from a reconstruction of the dispersion curve. In this study, the time-of-flight as well as the elastic stiffnesses D_{11}, D_{22}, A_{44}, and A_{55} for composite samples which have undergone combined thermal and mechanical aging are obtained. The samples examined include a baseline specimen with 0 cycles, specimens which have been aged 2350 and 3530 cycles at high strain levels, and one specimen aged 3530 cycles at low strain levels.
 [S0001-4966(99)02309-7]

PACS numbers: 43.35.Zc, 43.40.Le, 43.20.Mv [HEB]

INTRODUCTION

Future advanced aerospace composite materials will be required to respond well to combined thermal and mechanical loading. These materials will be used for the next generation of aerospace structures, such as the proposed high-speed civil transport (HSCT). The flight profile of this aircraft will include speeds of up to Mach 2.4, altitudes of 60 000 ft, and exterior temperatures up to 200 °C.[1] The materials used in the design of such a structure will need to perform well under these strenuous conditions for thousands of flight hours. Therefore, it is of interest to investigate the feasibility of nondestructively monitoring thermal-mechanical aging in composites.

Among the various techniques available, ultrasonic Lamb waves offer a convenient method of evaluating these composite materials. Studies have been conducted which show a reduction in Lamb wave velocity due to a loss of stiffness caused by matrix cracking.[2–5] Seale *et al.*[2] showed a correlation between Lamb wave velocity and stiffness measured with strain gages as well as a correlation between crack density and velocity for fatigued composite samples. Dayal and Kinra[3] found the wave speed as well as attenuation to be sensitive to cracking for a Lamb wave propagating in the plane of the plate. Tang and Henneke[4] noted that, since composites are commonly designed to carry in-plane loads, Lamb wave measurements are more useful because they provide information about the in-plane properties of a plate. Similarly, Dayal *et al.*[5] noted that Lamb waves provide an effective method to detect damage due to transverse matrix cracking because the Lamb wave interaction with cracks is much stronger in the plane of the plate.

Since the Lamb wave velocity depends on the elastic stiffness of a material, it provides an excellent method to nondestructively obtain information about the integrity of a material. Karim *et al.*[6] and Mal *et al.*[7] have used inversion techniques to determine the material parameters of composites from experimental Lamb wave data. Recently, Shih *et al.*[8] used Lamb wave velocity measurements to calculate laminate stiffness constants in fatigued composites. Lamb wave techniques have also been used to study composite defects such as delaminations,[9–12] porosity,[12,13] and fiber misalignment.[13]

Under the normal use environment for many composite aerospace structures, thermal degradation as well as damage due to mechanical fatigue may occur. While studies have been conducted using Lamb waves to examine fatigue damage,[2–5,8] few studies have been conducted which monitor either thermal degradation or thermal-mechanical aging in polymer matrix composites using ultrasonic nondestructive evaluation techniques. However, ultrasonic Lamb waves have been shown by Seale *et al.*[2] as well as Bar-Cohen *et al.*[14] to be an effective method for characterizing thermal damage in composites.

A new scanning system has been developed that determines the velocity of the lowest order antisymmetric Lamb mode over a wide frequency range. The elastic bending and out-of-plane shear stiffnesses of the material are computed from a reconstruction of the velocity dispersion curve which best fits the data. For a laminated composite, the D_{11}, D_{22}, A_{44}, and A_{55} stiffness matrix components can be determined. The measurement technique has been shown to be sensitive to changes in bending stiffness caused by mechanical degradation such as impact damage, delaminations, and debonds.[15] The scanner can also be used to monitor stiffness variations created during manufacturing which may be due to changes in fiber volume content, porosity content, or due to process control variables.

In this measurement, the average stiffness of the material on a line connecting the two sensors, which is the direction of Lamb wave propagation, is determined. The velocity at each frequency is calculated from the known transducer

[a]National Research Council Associate.

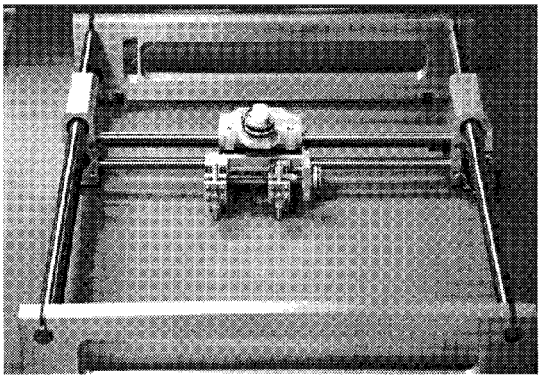


FIG. 1. Photo of the Lamb wave scanning system.

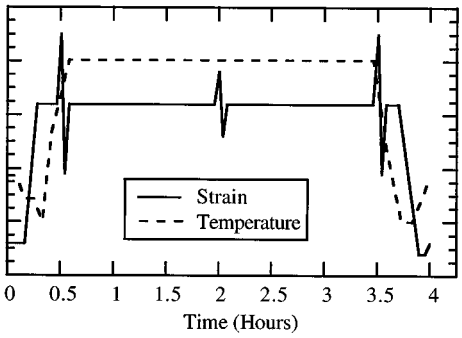


FIG. 2. Typical supersonic flight profile.

separation and the measured time-of-flight. The same peak in the waveforms received at various distances is used to measure the time difference between the signals. The velocity measurements are accurate and repeatable to within 1%, resulting in reconstructed stiffness values repeatable to within 4%.[15] A mechanical scanner, shown in Fig. 1, is used to move the sensors over the surface to map the time of flight, velocity, or stiffnesses of the entire specimen. Access to only one side of the material is required and, because the sensors are dry coupled to the surface of the plate, neither immersing the sample nor couplants are required.

This study will explore the use of the Lamb wave technique to monitor thermal-mechanical aging in the composite specimens. The Lamb wave scanning system was used to measure time-of-flight and velocity on thermal-mechanically aged composite samples. For each specimen, the stiffness was obtained from an inversion of the velocity dispersion data. The resulting stiffness reduction for samples which have been aged 3530 cycles at high and low strain levels and one specimen which has been aged 2350 cycles at low strain levels are discussed.

I. SAMPLES AND AGING PROFILE

The composite material studied was a graphite fiber reinforced amorphous thermoplastic polyimide with a stacking sequence of $[45/0/-45/90]_{2S}$. The specimens were 90 cm by 30 cm and had a nominal thickness of 0.22 cm. The samples were subjected to thermal-mechanical aging in either 98-kN (22-kip) or 222-kN (50-kip) capacity load frames equipped with environmental chambers which had a usable temperature range of $-54\,^{\circ}\mathrm{C}$ to $+344\,^{\circ}\mathrm{C}$. The combined thermal-mechanical aging to which these materials were subjected was representative of the conditions which supersonic airplanes will encounter. Flights which cover takeoff, climbing to high altitudes, flying at extreme speeds, and descending to land subject the aircraft to a wide range of temperatures and load variations.

A typical temperature-strain profile for a supersonic aircraft flight is similar to that shown in Fig. 2. The temperature profile shown was adapted from information contained in Noor and Venneri[16] and Sutton.[17] The strain profile shown was adapted from information obtained from Harpur.[18] The units of the temperature and strain have been omitted due to the fact that different areas of the aircraft are subject to differing levels of both strain and temperature. Obviously, the leading edges will be subjected to a higher temperature excursion than other portions of the plane. The skin temperature at various locations on a Mach 2.4 aircraft as well as the skin temperature as a function of Mach number and altitude can be found in Noor and Venneri.[16] The temperature profiles in Ref. 16 show skin temperatures for a Mach 2.4 flight at 60 000 ft to be in the neighborhood of 150 °C.

In this study, each aging cycle had a duration of 255 min. The temperature extremes for all samples were chosen to be $-18\,^{\circ}\mathrm{C}$ to $+177\,^{\circ}\mathrm{C}$ with a sustained temperature of $+177\,^{\circ}\mathrm{C}$ for 180 min. The load levels for the low-strain sample ranged from 0 to 2000 microstrain with a load at or above 1040 microstrain for 180 min. The load levels for the high-strain samples ranged from 0 to 3000 microstrain with a load at or above 1560 microstrain for 180 min.

II. THEORY

Lamb waves arise from a coupling between the shear-vertical (SV) and compressional (P) waves reflected at the stress-free boundaries at the top and bottom of a thin plate. There are two classes of Lamb mode solutions, symmetric and antisymmetric, which are defined in terms of the displacement of the plate with respect to the mid-plane. As the frequency of the wave increases, higher order symmetric and antisymmetric modes begin to propagate. At low frequencies (below 1 MHz for thin plates), only the lowest order symmetric mode, S_0, and lowest order antisymmetric mode, A_0, propagate. In this region, the S_0 mode is almost nondispersive and the A_0 mode is highly dispersive. The velocity of each Lamb mode is directly related to the properties of the material. Therefore, an effective tool exists to calculate the stiffness of a composite by measuring the velocity of these waves. This study will investigate solutions for the A_0 mode.

For a composite lamina with the *x*-axis defined as in the fiber direction, the *y*-axis transverse to the fibers, and the *z*-axis being out of the plane of the plate, the stress-strain relationship for a lamina under a condition of plane stress is given by

$$\begin{bmatrix}\sigma_1\\ \sigma_2\\ \tau_4\\ \tau_5\\ \tau_6\end{bmatrix}=\begin{bmatrix}Q_{11} & Q_{12} & 0 & 0 & 0\\ Q_{12} & Q_{22} & 0 & 0 & 0\\ 0 & 0 & Q_{44} & 0 & 0\\ 0 & 0 & 0 & Q_{55} & 0\\ 0 & 0 & 0 & 0 & Q_{66}\end{bmatrix}\begin{bmatrix}\epsilon_1\\ \epsilon_2\\ \gamma_4\\ \gamma_5\\ \gamma_6\end{bmatrix}, \tag{1}$$

where σ and τ represent the normal and shear stresses, respectively, and ϵ and γ represent the normal and shear strains, respectively. The Q_{ij} are the reduced stiffness components. Relationships between the reduced stiffnesses and the elastic stiffness constants, c_{ij}, can be found in Jang[20] and equations relating the engineering constants (E, G, and ν) to the reduced stiffnesses can be found in Daniel and Ishai.[21] The bending stiffnesses, D_{11} and D_{22}, and the out-of-plane stiffnesses, A_{44} and A_{55}, are obtained by integrating the Q_{ij} through the thickness of the plate. These stiffness values are defined as[19]

$$D_{ii}=\int_{-h/2}^{h/2}(Q'_{ii})_k z^2\,dz, \quad i=1,2,6 \tag{2}$$

and

$$A_{jj}=k_j^2\int_{-h/2}^{h/2}(Q'_{jj})_k\,dz, \quad j=4,5, \tag{3}$$

where the subscript k represents each layer in the laminate, k_j is a shear correction factor, and h is the total thickness of the plate. The Q'_{ij} are the transformed stiffness coefficients which take into account the orientation of each ply with respect to the wave propagation direction.

For a symmetric quasi-isotropic plate, the dispersion relation for the A_0 or flexural plate mode propagating in the 0° direction is given by[19]

$$\begin{aligned}&(D_{11}k^2+A_{55}-I\omega^2)(D_{66}k^2+A_{44}-I\omega^2)(A_{55}k^2-\rho^*\omega^2)\\ &\quad-(D_{16}k^2)^2(A_{55}k^2-\rho^*\omega^2)\\ &\quad-(A_{55}k)^2(D_{66}k^2+A_{44}-I\omega^2)=0.\end{aligned} \tag{4}$$

For waves propagating in the 90° direction of a symmetric quasi-isotropic plate, the relation is[19]

$$\begin{aligned}&(D_{22}k^2+A_{44}-I\omega^2)(D_{66}k^2+A_{55}-I\omega^2)(A_{44}k^2-\rho^*\omega^2)\\ &\quad-(D_{16}k^2)^2(A_{44}k^2-\rho^*\omega^2)\\ &\quad-(A_{44}k)^2(D_{66}k^2+A_{55}-I\omega^2)=0.\end{aligned} \tag{5}$$

In the equations, ω is the angular frequency and k is the wave number. I and ρ^* are defined as[22]

$$I=\int_{-h/2}^{h/2}\rho\,dz \tag{6}$$

and

$$\rho^*=\int_{-h/2}^{h/2}\rho z^2\,dz, \tag{7}$$

where ρ is the density and h is the plate thickness. Although multiple roots exist for the dispersion relations given by Eqs. (4) and (5), only those roots which satisfy the condition that the velocity approaches zero as the frequency approaches zero are the ones corresponding to the A_0 mode.[22] If the density, thickness, and stiffnesses are known, the dispersion curve can be obtained by choosing a wide range of wave numbers, k, and solving Eqs. (4) and (5) for the angular frequencies, ω, which correspond to each wave number. Once ω and k are known, the phase velocity, ν, is given by

$$\nu=\frac{\omega}{k} \tag{8}$$

and the frequency, f, is obtained from the angular frequency by using the relation

$$f=\frac{\omega}{2\pi}. \tag{9}$$

The dispersion curve is then generated by plotting the velocity as a function of frequency. For a further description of plate theory and how the laminate stiffnesses relate to the dispersion curve, the reader is referred to Tang *et al.*[19]

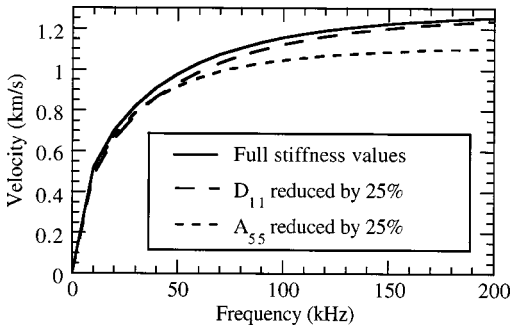


FIG. 3. Flexural dispersion curves using full values of all five stiffness parameters, with D_{11} reduced by 25%, and with A_{55} reduced by 25%. The curves with the D_{16}, D_{66}, and A_{44} stiffness reduced by 25% could not be discerned from the curve with the full stiffness parameters.

For propagation in the 0° direction, the effects on the flexural dispersion curve of reducing each of the stiffness constants in Eq. (4) by 25% are shown in Fig. 3. Decreasing the values of D_{16}, D_{66}, and A_{44} did not significantly alter the dispersion curve and the shift in the curve could not be discerned from the curve generated using the full stiffness parameters. The effects of altering D_{16}, D_{66}, and A_{44} will be shown in the subsequent figure. As can be seen in Fig. 3, decreasing D_{11} by 25% changes the curve only slightly and a larger shift in the curve in seen when A_{55} is reduced by 25%.

Figure 4 shows a plot of the percent reduction in velocity as a function of frequency for 25% reductions in each of the five stiffness values. The figure clearly shows that changing D_{11} has a greater effect on the velocity at lower frequencies and changes in A_{55} alter the dispersion curve at higher frequencies. The constants D_{16}, D_{66}, and A_{44} did not alter the dispersion curve by more than 0.5%. Since the only constants affecting the curve are D_{11} and A_{55}, the curve fitting to the experimental data will only use these two parameters. The remaining stiffness values will be fixed at their theoretical values obtained from Eqs. (2) and (3). The dispersion curve obtained from Eq. (5) for propagation of the flexural in the 90° direction is the same as shown in Fig. 3 except the

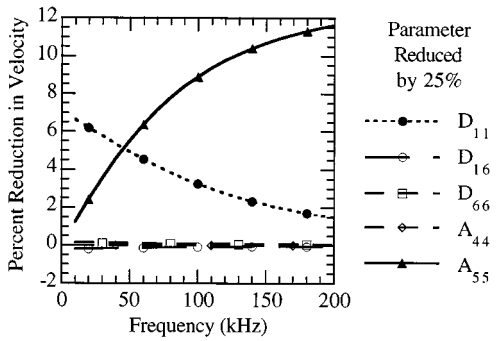


FIG. 4. Plot of the percent reduction in velocity as a function of frequency for a 25% decrease in each of the stiffness parameters.

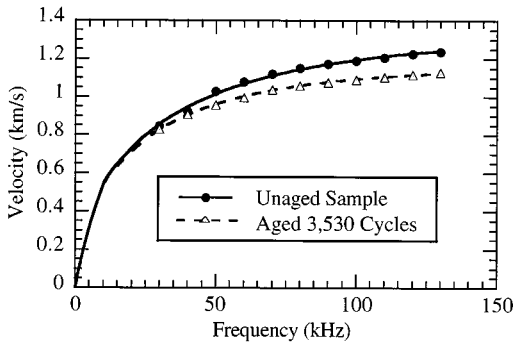


FIG. 5. Experimental dispersion curves for an unaged sample and a sample aged 3530 cycles at high-strain levels. Also shown are the reconstructed dispersion curves for each sample.

constants controlling the behavior are D_{22} and A_{44}. Since the values of D_{16}, D_{66}, and A_{55} do not affect the dispersion curve for Lamb waves propagating in the 90° direction, these values will be held constant and only the values of D_{22} and A_{44} will be used to fit the experimental data.

III. LAMB WAVE MEASUREMENTS

The elastic bending and out-of-plane shear stiffnesses of the material were computed using experimental velocity measurements of the flexural plate mode over a wide range of frequencies. For propagation in the 0° direction, the D_{11} and A_{55} stiffnesses in Eq. (4) were determined by adjusting their values in order to produce the best fit to the experimentally obtained values of ω and k. The stiffnesses D_{22} and A_{44} in Eq. (5) were computed in a similar manner from data obtained for propagation in the 90° direction. Further details of the velocity and stiffness measurements using this technique can be found in Huang *et al.*[15]

The Lamb wave scanning system was used to measure time of flight, velocity, and stiffness in the 0° and 90° directions on one unaged and three aged specimens. For the measurements, a four-cycle Gaussian-enveloped sine wave was used to generate the signal and the received signal was captured at sampling rate of 25 MHz. The sensor separation was varied from 2.75 cm to 4.75 cm in increments of 0.5 cm. The small step size was used in order to assure that the same peak in the waveform was followed over the total propagation distance. The frequency was swept from 30 kHz to 130 kHz in 10-kHz steps and the velocity at each frequency was obtained from the known transducer separation and measured time of flight. The thickness was as given earlier and the density, estimated from common values for composites, was taken to be 1560 kg/m^3. From the data taken in the 0° direction, the dispersion curve was reconstructed by altering the values of A_{55} and D_{11} in Eq. (4) in order to produce the best nonlinear least squares fit to the experimental velocity measurements. From the reconstruction, values for the out-of-plane stiffness, A_{55}, and bending stiffness, D_{11}, were obtained. Similarly, from data taken in the 90° direction, the stiffnesses A_{44} and D_{22} were obtained by altering these values in Eq. (5) in order to produce the best fit to the experimental velocity measurements.

Shown in Fig. 5 are the experimental velocity measurements for an unaged sample and a sample with 3530 cycles of aging at high-strain levels. Also shown are the reconstructed dispersion curves. As can be seen from the figure, the dispersion curve for the aged sample is clearly shifted from that of the unaged sample. The dispersion curves shown in Fig. 5 were reconstructed using values for the out-of-plane stiffness, A_{55}, and the bending stiffness, D_{11}, which best fit the experimental data. For the curves shown, the value of A_{55} decreased by over 22% and the value of D_{11} increased by 1.3% for the aged sample as compared to the unaged sample.

It is expected, and previous strain gage measurements show,[2] that matrix cracking due to fatigue damage in composites leads to a decrease in elastic moduli. Since the samples were mechanically loaded as well as thermally cycled, a likely damage mechanism causing a stiffness reduction is matrix cracking. The stiffness A_{55} is controlled heavily by the matrix since the out-of-plane shear carrying capabilities of the composite are matrix dominated. Therefore, the decrease in the stiffness A_{55} was as expected. However, the increase in D_{11} would not be anticipated. The reason for the increase in the value is most likely due to the insensitivity of the dispersion curve to changes in D_{11} over the measurement frequency range (30 kHz–130 kHz). In this region, the parameter dominating the behavior of the curve will be A_{55} (see Fig. 4). Only a few data points exist at the very low frequencies (below 50 kHz) where D_{11} controls the behavior of the curve. Due to this lack of data, the constant D_{11} will be inaccurate. To address this problem, new sensors have been developed which have better low frequency response. With improved data collection capabilities at low frequencies, future data gathered using the scanner will provide more accurate values of D_{11}. For this study, the improved sensors were not available. Therefore, the more accurate constant measured for the aged samples was the out-of-plane stiffness, A_{55}.

Velocity measurements were made in 45 different regions across the entire sample to obtain a mapping of the out-of-plane stiffness, A_{55}, for each of four different samples with various degrees of aging. The stiffness was measured in increments of 5 cm along the length of the specimen and 6 cm across the width of the specimen for a total scan area of

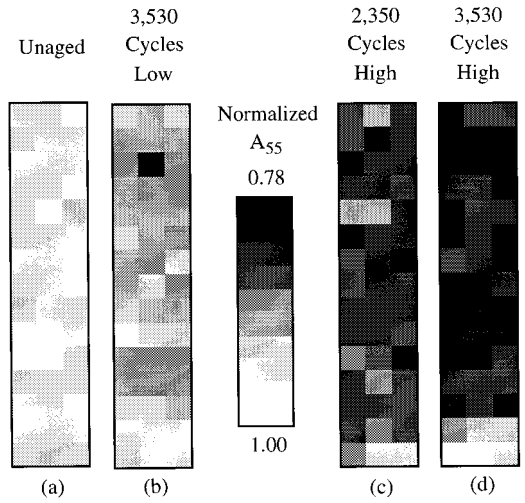


FIG. 6. A_{55} stiffness mapping for four composite samples: (a) unaged; (b) aged 3530 cycles at low-strain levels; (c) aged 2350 cycles at high-strain levels; and (d) aged 3530 cycles at high-strain levels.

70 cm by 12 cm. The results are shown in Fig. 6. The loss of stiffness with increasing aging is shown very clearly in the figure. The stiffnesses have been normalized to the highest value for A_{55} obtained for the unaged sample. In the figure, the higher stiffness values in the regions at the bottom of the aged specimens were due to that portion of the sample being out of the ovens and, therefore, not subjected to the same aging process as the rest of the sample. The average normalized stiffness and standard deviation for the 45 measurement regions across each sample, excluding the measurements in the areas which were outside of the oven, are shown in Table I. Also shown in the table are the values of D_{11} measured for each sample.

From the table, it can be seen that the average value of A_{55} for the samples subjected to high strain levels changed by over 10% for the 2350-cycle sample and by 14% for the 3530-cycle sample. The sample which was aged at low strain levels for 3530 cycles showed a change of 5% for A_{55}. The standard deviation for the 45 measurements across each sample was also observed to be only around 4% as well. In contrast, the values of D_{11} were highly variable and had errors on the order of 15% for all of the samples. The large standard deviations associated with the average values suggest that the measured values were not reliable. As mentioned earlier, this inaccuracy was probably due to this lack of data at low frequencies.

Stiffness results were also obtained for values of D_{22} and A_{44} from measurements in the 90° direction. Due to the mechanical constraints of the scanner, only 18 stiffness values in the middle portion of the sample could be acquired. The total scan area was 25 cm along the length of the specimen and 12 cm across the width of the specimen. The results for the 90° scans are shown in Fig. 7. As can be seen from the figure, the value of A_{44} decreased with increased aging. However, the values did not decrease as much as the values of A_{55}. This is expected since the load was applied in the 0° direction during the aging process. Loading in this direction would tend to produce cracks in the 90° plies as well as the $\pm 45°$ plies. This type of damage would be indicated by a decrease in A_{55}. Since the loading occurred in the 0° direction, matrix cracking in the 0° plies probably did not occur or will not be as pronounced as in the 90° plies. Therefore, the measurement of A_{44} will only be sensitive to the cracks in the $\pm 45°$ plies. This would lead to a less significant decrease in stiffness for A_{44}, which is consistent with observation.

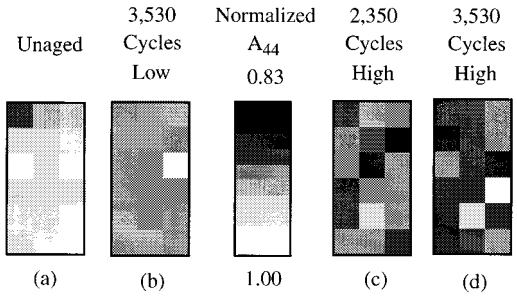


FIG. 7. A_{44} stiffness mapping for four composite samples: (a) unaged; (b) aged 3530 cycles at low-strain levels; (c) aged 2350 cycles at high-strain levels; and (d) aged 3530 cycles at high-strain levels.

The average stiffness and standard deviation for each sample for the 90° measurements are shown in Table II. Also shown in the table are the values of D_{22} for each sample. From the table, it can be seen that the average value of A_{44} for all of the samples changed by roughly 5%. As was observed with the values of D_{11}, the values of D_{22} were highly variable and had large errors.

In addition to the stiffness measurements, time-of-flight scans in the 0° and 90° directions were conducted on the samples. A four-cycle Gaussian sine wave at a fixed frequency of 150 kHz was used to generate the antisymmetric Lamb wave. The scan area was 75.0 cm by 24.0 cm in the 0° direction and 25 cm by 24 cm in the 90° direction. The step size for all scans was 1.0 cm and the transducers were held fixed at a separation distance of 2.75 cm. The results for the 0° and 90° scans are shown in Figs. 8 and 9, respectively. In the figures, the values have been normalized to the lowest measured time-of-flight (fastest velocity) for all samples. As can be seen in the figures and consistent with the previous stiffness measurements, the time-of-flight is clearly shifted to longer times for the aged samples.

The degraded stiffness in the aged specimens leads to a slower Lamb wave velocity which is reflected by the longer

TABLE I. Normalized stiffness values in the 0° direction for aged samples.

Normalized stiffness	0-cycle baseline	3530 cycles Low strain	2350 cycles High strain	3530 cycles High strain
A_{55}	0.95±0.02	0.90±0.02	0.84±0.02	0.81±0.02
D_{11}	0.69±0.05	0.66±0.06	0.74±0.10	0.66±0.12

TABLE II. Normalized stiffness values in the 90° direction for aged samples.

Normalized stiffness	0-cycle baseline	3530 cycles Low strain	2350 cycles High strain	3530 cycles High strain
A_{44}	0.94±0.03	0.90±0.01	0.88±0.02	0.88±0.04
D_{22}	0.89±0.07	0.78±0.03	0.81±0.08	0.76±0.08

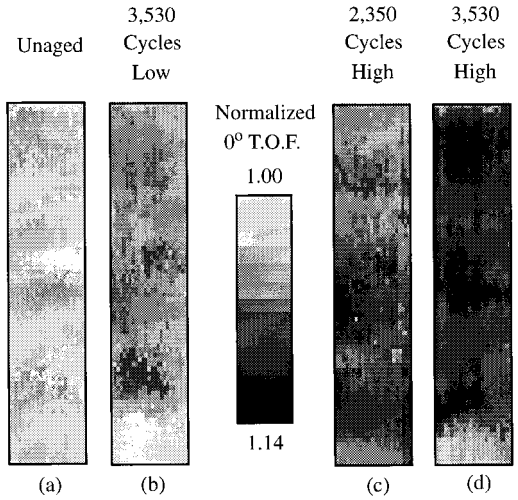


FIG. 8. Time-of-flight (TOF) scans in the 0° direction for four composite samples: (a) unaged; (b) aged 3530 cycles at low-strain levels; (c) aged 2350 cycles at high-strain levels; and (d) aged 3530 cycles at high-strain levels.

time-of-flight. The average time-of-flight and standard deviation for the 0° and 90° measurements are shown in Table III. As can be seen in the data, which is also consistent with the stiffness measurements, the shift in time-of-flight in the 90° directions is less than the shift in 0° direction. Again, this difference was most likely due to the loading in the 0° direction causing more matrix cracks to develop in the 90° plies.

IV. DISCUSSION

The Lamb Wave Imager™ is a unique new tool for nondestructively measuring the elastic properties of a composite material. It has been shown by this study to be a very effective method in providing a quantitative measure of thermal-mechanical aging in composite materials. The out-of-plane stiffness measurements in both the 0° and 90° directions correlated well with aging levels. Time-of-flight scans showed a significant increase for the aged samples as compared to an unaged sample. This corresponds to a slower velocity, which was in agreement with the reduced stiffness measured in the aged samples. The values for the bending stiffness of the samples were also obtained, but the large standard deviation associated with the measurement of this parameter masked any true significance.

The Lamb wave scanning system provides a fast and accurate method of nondestructively obtaining quantitative information about materials. The system is portable and can be used to make *in situ* measurements of stiffness. Since the Lamb wave technique used here provides information about the out-of-plane stiffness, it complements standard ultrasonic measurements of velocity and attenuation. Additionally, the

TABLE III. Normalized time-of-flight values in the 0° and 90° directions for aged samples.

Normalized time-of-flight	0-cycle baseline	3530 cycles Low strain	2350 cycles High strain	3530 cycles High strain
0°Direction	1.03±0.01	1.06±0.02	1.07±0.02	1.09±0.02
90°Direction	1.05±0.01	1.06±0.01	1.07±0.01	1.09±0.01

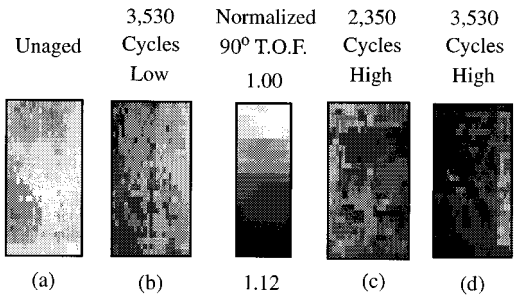


FIG. 9. Time-of-flight (TOF) scans in the 90° direction for four composite samples: (a) unaged; (b) aged 3530 cycles at low-strain levels; (c) aged 2350 cycles at high-strain levels; and (d) aged 3530 cycles at high-strain levels.

Lamb wave technique offers several advantages over conventional through-the-thickness measurements. Access to only one side of the material is required, no immersion or couplants are needed because the sensors are dry coupled to the surface of the plate, and the measurement is not limited to the thinness of the plate. Also, since actual values of the stiffness of a composite can be obtained, the Lamb wave velocity measurements provide useful data about the integrity and load-carrying capabilities of the material. This type quantitative information is important when making decisions concerning the safety of composite structures.

In addition to measuring the effects of thermal-mechanical aging, the Lamb wave technique is also sensitive to stiffness changes caused by impact damage, delaminations, fatigue, and debonds. Therefore, it is reasonable to expect that this system will detect manufacturing anomalies such as inconsistent thickness, fiber misalignment, porosity, and low fiber volume fraction. This instrument provides new capabilities to inexpensively and accurately map out material properties, which is particularly useful for determining variations in material properties across a given specimen due to damage, monitoring quality and process control, and verifying the design of composites.

ACKNOWLEDGMENTS

This work was performed while M.D.S. held a National Research Council NASA-LaRC Research Associateship. The authors would also like to thank Steve Ziola, Wei Huang, and John Dorighi of Digital Wave Corporation for their technical support involving the scanning system.

[1] E. Baer, A. Hiltner, and R. J. Morgan, ''Biological and synthetic hierarchical composites,'' Phys. Today **45**, 60–67 (1992).
[2] M. D. Seale, B. T. Smith, and W. H. Prosser, ''Lamb wave assessment of fatigue and thermal damage in composites,'' J. Acoust. Soc. Am. **103**, 2416–2424 (1998).
[3] V. Dayal and V. K. Kinra, ''Ultrasonic NDE of composites for transverse cracking,'' in *Optical Methods in Composites; Proceedings of the SEM Fall Conference on Experimental Mechanics* (A88-13876 03-24) (Society for Experimental Mechanics, Inc., Bethel, CT, 1986), pp. 17–22.
[4] B. Tang and E. G. Henneke II, ''Lamb-wave monitoring of axial stiffness reduction of laminated composite plates,'' Mater. Eval. **47**, 928–934 (1989).
[5] V. Dayal, V. Iyer, and V. K. Kinra, ''Ultrasonic evaluation of microcracks in composites,'' in *Advances in Fracture Research; Proceedings of the Seventh International Conference on Fracture* (ICF7), Vol. 5 (A90-41276 18-39) (Pergamon, New York, 1989), pp. 3291–3300.
[6] M. R. Karim, A. K. Mal, and Y. Bar-Cohen, ''Inversion of leaky Lamb

wave data by simplex algorithm,'' J. Acoust. Soc. Am. **88**, 482–491 (1990).
[7] A. K. Mal, M. R. Gorman, and W. H. Prosser, ''Material characterization of composite laminates using low-frequency plate wave dispersion data,'' in *Review of Progress in Quantitative Nondestructive Evaluation*, Vol. 11, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1992), pp. 1451–1458.
[8] J.-H. Shih, A. K. Mal, and M. Vemuri, ''Plate wave characterization of stiffness degradation in composites during fatigue,'' Res. Nondestruct. Eval. **10**, 147–162 (1998).
[9] K. J. Sun, ''Application of guided acoustic waves to delamination detection,'' in *Review of Progress in Quantitative Nondestructive Evaluation*, Vol. 11, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1992), pp. 1213–1219.
[10] N. Guo and P. Cawley, ''The interaction of Lamb waves with delaminations in composite laminates,'' J. Acoust. Soc. Am. **94**, 2240–2246 (1993).
[11] N. Guo and P. Cawley, ''Lamb waves for the NDE of composite laminates,'' in *Review of Progress in Quantitative Nondestructive Evaluation*, Vol. 11, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1992), pp. 1443–1450.
[12] Y. Bar-Cohen and D. E. Chimenti, ''NDE of defects in composites using leaky Lamb waves,'' in *Symposium on Nondestructive Evaluation*, 15th Proceedings (A86-47129 22-38) (Nondestructive Testing Information Analysis Center, San Antonio, TX, 1986), pp. 202–208.
[13] K. Balasubramaniam and J. L. Rose, ''Guided plate wave potential for damage analysis of composite materials,'' in *Review of Progress in Quantitative Nondestructive Evaluation*, Vol. 9, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1990), pp. 1505–1512.
[14] Y. Bar-Cohen, S-S. Lih, A. Mal, and Z. Chang, ''Rapid characterization of the degradation of composites using plate wave dispersion data,'' in *Review of Progress in Quantitative Nondestructive Evaluation*, Vol. 17, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1998), pp. 1171–1176.
[15] W. Huang, S. M. Ziola, J. F. Dorighi, and M. R. Gorman, ''Stiffness measurement and defect detection in laminated composites by dry-coupled plate waves,'' in *Process Control and Sensors for Manufacturing*, edited by R. H. Bossi and D. M. Pepper (SPIE, Bellingham, WA, 1998), pp. 66–76.
[16] A. K. Noor and S. L. Venneri, senior editors, ''New and Projected Aeronautical and Space Systems, Design Concepts, and Loads,'' in *Flight-Vehicle Materials, Structures, and Dynamics—Assessment and Future Directions*, Vol. 1 (The American Society of Mechanical Engineers, New York, 1994), pp. 15–84.
[17] Sir Graham Sutton, *Mastery of the Air* (Basic Books, Inc., New York, 1965), pp. 157–166.
[18] N. F. Harpur, ''Concorde structural development,'' in *AIAA Commercial Aircraft Design and Operation Meeting* (American Institute of Aeronautics and Astronautics, New York, 1967), No. 67-402, pp. 1–14.
[19] B. Tang, E. G. Henneke II, and R. C. Stiffler, ''Low frequency flexural wave propagation in laminated composite plates,'' in *Acousto-Ultrasonics Theory and Application*, edited by John C. Duke, Jr. (Plenum, New York, 1987), pp. 45–65.
[20] B. Z. Jang, *Advanced Polymer Composites* (ASM International, Materials Park, OH, 1994), p. 93.
[21] I. M. Daniel and O. Ishai, *Engineering Mechanics of Composite Materials* (Oxford University Press, New York, 1994), pp. 34–47.
[22] W. H. Prosser, ''The propagation characteristics of the plate modes of acoustic emission waves in thin aluminum plates and thin graphite/epoxy composite plates and tubes,'' NASA Technical Memorandum 104187, November 1991.

Mode-locking of acoustic resonators and its application to vibration cancellation in acoustic heat engines

P. S. Spoor and G. W. Swift
Condensed Matter and Thermal Physics Group, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

(Received 10 February 1999; accepted for publication 10 June 1999)

Vibration induced in engine hardware by a working fluid can be very significant in high-power, high-amplitude acoustic heat engines, and is a serious impediment to their practical use. This vibration can cause fatigue and destruction of engine components as well as fuel lines, cooling lines, and sensor wires. The forces involved make anchoring such an engine to an ''immovable'' object impractical. Rigidly attaching two such engines together, and acoustically coupling them with a duct of such a length and diameter that the two engines mode-lock in antiphase (thus canceling the longitudinal vibration) appears to be an inexpensive, viable solution. This paper describes in detail experiments demonstrating the feasibility of this idea, and the underlying theory. © *1999 Acoustical Society of America.* [S0001-4966(99)03909-0]

PACS numbers: 43.35.Ud [HEB]

INTRODUCTION

In recent years, considerable progress has been made in designing practical heat engines that have no moving parts, other than acoustic oscillations in the working fluid;[1] such engines work on principles loosely exemplified by such devices as the Sondhauss tube[2,3] and the Rijke tube,[4] where heat suitably applied to a tube causes spontaneous oscillation of the gas inside. Thermoacoustic heat engines have the advantage of no bearings or sliding seals, so they hold out the possibility of extreme simplicity and reliability. In order to achieve useful power densities, however, the engines must contain a working fluid under high pressure, oscillating with an amplitude on order of 10% or more of mean pressure. Thus the working fluid is capable of exerting enormous oscillating force on the structure surrounding it, inducing intense vibration in the engine hardware. Far from being a mere nuisance, this vibration can be downright terrifying in an engine of appreciable size; a one-ton prototype that produces 20 kW of acoustic power[5] experiences several g's of acceleration when operating full tilt.

The need to control this vibration has been understood from the earliest days of thermoacoustic research; in a 1958 patent detailing heat-driven acoustic devices, Marrison[6] describes two heat-driven acoustic resonators, intended to drive small linear alternators, joined end-to-end as shown in Fig. 1, to achieve vibration cancellation. The original patent reads:

> ...a pair of devices...juxtaposed in end-to-end relation for dynamic balance. In order that there shall be adequate coupling between the two vibrating gas columns, a channel is provided that interconnects them...In the operation of this balanced arrangement, the movement of the gas at any instant is either inward toward the central plane of both cylinders, or outward toward the separated ends of the two cylinders. Thus the vibration of each gas column finds a reaction in the vibration of the other gas column so that dynamic balance is secured, and external vibration, shaking of the mount, noise, and the like are minimized.

The interconnecting channel is necessary because thermoacoustic engines, like organ pipes or pendulum clocks, are self-maintained oscillators, and a pair will not naturally run at the same frequency, or in a particular phase relationship, unless they are coupled sufficiently.

The mode-locking of coupled, self-maintained oscillators was first described by Christiaan Huygens in 1665, who reported in a letter to his father[7] that a pair of large pendulum clocks sitting a few feet apart would synchronize their ticks; he further noted that

> If two rhythms are nearly the same and their sources are in close proximity, they will always lock up, fall into synchrony, entrain.

Following Huygens, we suppose that if two acoustic engines are nearly the same in natural frequency and coupled in some way with sufficient strength, they may entrain, thus locking to each other in frequency and phase. This principle is implicit in the Marrison device, and in this work we explicitly aim for an understanding of it, so it may be used to the greatest advantage in minimizing vibration.

Marrison's concept requires the tandem system of two engines to be twice as long as a single engine. Acoustic engines, requiring as they do a resonator which is a significant fraction of a wavelength (usually a half wavelength), tend to be large already, and a vibration-cancellation method that does not require additional length is clearly preferable. We instead consider two engines mounted side-by-side rather than end-to-end, with two neighboring ends connected by a narrow, bent duct a half-wavelength long. The acoustic pressure at the neighboring ends will tend to be in antiphase, thus encouraging the movement of the gas in each engine to be opposite the movement in the other. We show that mode-locking and vibration cancellation in such an arrangement can be achieved by careful consideration of how the locked state depends on engine and coupler geometry, the working

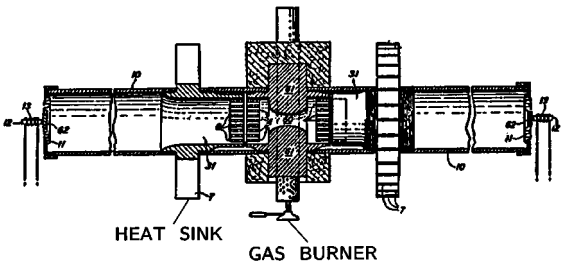


FIG. 1. W. A. Marrison's original concept, from a 1958 patent, for two acoustic engines mounted in opposition for dynamic balance, with a narrow channel coupling the two (Ref. 6).

fluid properties, and the difference in natural frequencies of the two engines.

Our experiments use two small (acoustic power$\approx$5 W, length$=$1.75 m, diameter$=$35 mm) thermoacoustic engines whose cases are welded together. They are coupled acoustically by one of a number of ducts nominally half a wavelength long, with radii ranging from 1/14th to 1/7th of the engine resonator radius. The working fluid for most of the measurements is air at 80 kPa, which is atmospheric pressure in Los Alamos, elevation 2200 m. Mode-locking of the engines, with their acoustic oscillations in antiphase, is readily achieved; measurements of frequency, pressure amplitude, and phase, as well as accelerometer measurements on the cases, indicate that 90% of the case vibration is canceled by using a coupler which is 3.5 mm in diameter, or one-tenth of the diameter of an engine resonator, if the engines are not more than 0.2% apart in natural (uncoupled) frequency. The sensitivity to frequency difference is partly a consequence of our system's small size; numerical simulations indicate that for a 30-bar He system whose engine resonators are 10 times the diameter of those in our small test system, a coupler one-tenth the engine resonator diameter allows cancellation of 95% percent of case vibration, when the engines' natural frequencies are as much as 1% different. This simulation, applicable to the 20-kW engine mentioned earlier, implies that our method should prove useful.

In addition to the measurements and simulations mentioned above, we have developed a simple theory to explain the dependence of the phase of the locked state (and hence the degree of vibration cancellation) on the coupler dimensions, fluid properties, and the difference in the engines' natural frequencies. We find that theory, simulation, and experiment all agree extremely well for our small test system, which provides encouragement that our methods can be relied upon to help design larger and more complex mode-locked acoustic systems.

Our engines can also mode-lock by ''mass coupling'' through their shared structure, with no acoustic coupling at all. This effect, which will be detailed in a separate paper,[8] is too weak to be a major concern in the vibration cancellation problem.[9]

I. THEORY

An excellent introduction to mode-locking can be found in Chapter 12 of A. B. Pippard's impressive book, *The Physics of Vibration*,[10] and we shall draw extensively from his work in developing our own theoretical model. Pippard considers the mode-locking of discrete systems such as LRC circuits; one of the interesting aspects of the present work is that acoustic systems contain the additional complexity of spatial dimension. Therefore we review some properties of passive, coupled acoustic resonators before considering the full mode-locking problem.

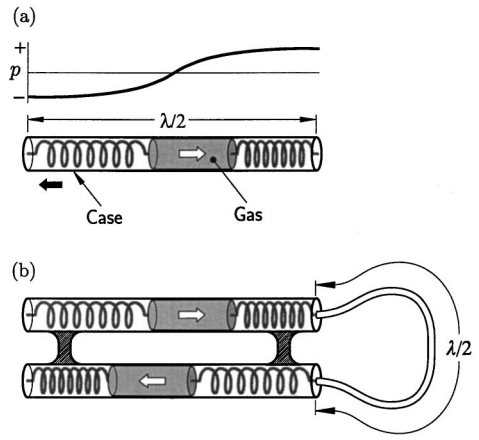


FIG. 2. Mass-and-spring models of half-wavelength acoustic resonators, showing (a) a single resonator, with a plug of gas bouncing off gas springs on the ends, and (b) two such resonators, rigidly attached, and coupled by a narrow half-wavelength duct to invert the phase of the pressure, thus encouraging the two resonators to oscillate in antiphase. Here λ stands for wavelength.

A. Coupled passive half-wave acoustic resonators

Figure 2 illustrates the basic case-vibration problem and our proposed solution. In Fig. 2(a), a single resonator's case moves in response to the motion of the gas inside; the frequency is such that the resonator's length is λ/2, where λ is the wavelength. This case motion can be analyzed either in terms of the axial forces exerted on its ends by the oscillating pressures inside, or in terms of momentum conservation, which requires that the case momentum be equal in magnitude but opposite in direction to the gas momentum at any instant. In Fig. 2(b) a half-wavelength coupler has been added to invert the pressure between two rigidly attached resonators, so that the acoustic oscillations are in antiphase and the net axial force exerted by the oscillating gas on the assembly is zero. In order to achieve such vibration suppression, the system of three coupled ducts must favor a normal mode corresponding to a half-wavelength in each duct, with the two large ducts (the resonators) oscillating in antiphase. Let the radius of the resonators be r_0 and the radius of the coupler be r_c; in the approximation that the ducts are all lossless and have the same length l, and $r_0 \gg r_c$, the normal mode frequencies in the neighborhood of $f=a/2l$ are easily obtained:

$$f_0 = \frac{a}{2l} \tag{1}$$

and

$$f_\pm \approx \frac{a}{2l}\left(1 \pm \frac{\sqrt{2}}{\pi}\frac{r_c}{r_0}\right), \tag{2}$$

where a is the sound speed. One solution, f_0, is the same as for a single duct, and corresponds to exactly a half-wavelength in each duct, with pressure nodes in the center of each duct, and velocity nodes at the two rigid ends and at each of the two junctions. Since velocity is zero at these junctions, matching volume velocities on either side does not require a discontinuity in velocity; hence, the coupler has the same waveshape and magnitude of oscillation as the resonators do. This is the desired antiphase mode, where the pressure oscillations in the resonators are π out of phase.

By contrast, f_+ and f_- correspond to ''in-phase'' modes, which tend to maximize case vibration rather than eliminate it. However, these modes have nonzero velocity at the junctions, and so the amplitude of oscillation in the coupler is much larger than that in the resonators. Although this analysis is based on lossless acoustics, the large oscillation amplitude in the coupler suggests that these in-phase modes are much lossier than the antiphase mode, so the antiphase mode is more likely to be selected. In fact, we have never observed anything but the antiphase mode in our mode-locked engines. In the antiphase mode the resonators and the coupler have the same oscillating pressure and velocity amplitudes, so the ratio of coupler to resonator dissipation in the thermoviscous boundary layer near the wall is simply the ratio of their radii. If this behavior is a good approximation for what actually happens when two acoustic engines are coupled together, a narrow coupler is less lossy than a wide one. Thus if resonator losses are significant in an engine, there is motivation for keeping the connecting duct narrow when coupling two together, to minimize additional loss.

B. Coupled, maintained acoustic resonators

The discussion of coupled duct resonators at the beginning of this section treated the two resonators and coupler as one composite resonator, with well-defined normal modes. A system of two self-maintained oscillators coupled with a narrow duct actually has a much richer behavior, such as that observed by supplying two similar organ flue pipes or slide whistles with compressed air so that they sound stable tones, and allowing them to couple by bringing their mouths in close proximity. When the pipes are far apart, they ''sing'' with their separate natural frequencies; one hears a ''beating'' at the difference frequency. As the pipes are brought closer together, this beat frequency slows, and then abruptly stops—at which point the pipes are locked in frequency and phase. The frequency of an individual pipe is sensitive to conditions at the mouth of the pipe; when the two pipes are close enough to interact, they may alter each other's frequencies enough to eventually phase lock, if their natural frequencies are not too different. Our coupled thermoacoustic engines display roughly the same behavior as these coupled organ pipes, and we will let our analytic approach be guided by this concept of beating slowed or stopped by coupling.

Before proceeding with our theory, it will be helpful to define exactly what information we seek. Mode-locking in acoustic oscillators is a broad enough subject that we cannot present an exhaustive treatment here. Rather, we will limit our inquiry at the outset to those aspects that we expect will be most relevant to the vibration-cancellation problem.

One important difference between coupled passive oscillators and coupled maintained oscillators is that in the passive system, the vibrations of the individual oscillators are either exactly in phase with each other or exactly antiphase. When the oscillators are maintained, however, this is no longer true, and a weakly coupled pair of very closely matched maintained oscillators may lock with essentially arbitrary phase. In addition, the amplitudes in each oscillator may be substantially different (true of a passive system as well, if the coupling is sufficiently weak). This may be seen quite readily in the organ pipe demonstration described above, if pressure sensors are placed in each pipe.

This has important consequences for us—for maximum vibration cancellation, our coupled resonators must be in antiphase *and* the amplitudes in both sides must be matched. If we let

$$\begin{aligned} \phi &= \text{phase difference between the resonators} \\ \zeta &= \text{ratio of pressure amplitudes in the resonators,} \end{aligned} \tag{3}$$

and consider the axial forces exerted by the oscillating fluid on the structure of our coupled system, we can express the degree of vibration cancellation in terms of ϕ and ζ. If we define a vibration-cancellation ratio

$$R_V = 1 - \frac{\text{amplitude of case vibration (coupled engines)}}{\text{amplitude of case vibration (one engine alone)}}, \tag{4}$$

then it is found that

$$R_V = 1 - [(1+\zeta\cos\phi)^2 + \zeta^2 \sin^2\phi]^{1/2}, \tag{5}$$

or

$$R_V \approx 1 - [(\zeta-1)^2 + (\phi-\pi)^2]^{1/2}, \quad \text{if} \quad \zeta \approx 1, \quad \phi \approx \pi. \tag{6}$$

Thus the quantities of most interest to us are ϕ and ζ; in particular, we seek their sensitivities $\partial\phi/\partial\Delta\omega$ and $\partial\zeta/\partial\Delta\omega$ to the mistuning $\Delta\omega$ of the resonators, $\Delta\omega$ being the difference in the resonators' uncoupled natural frequencies. To further simplify our analysis, we will assume that the coupler itself is in resonance. This seems intuitively like a desirable condition, based on the lossless analysis of Sec. I A and since it would appear to allow the maximum transmission of energy from one resonator to the other. This assumption has another interesting, nonintuitive consequence as well. If we treat the coupler as an impedance coupling two acoustic circuits, the resonant coupler corresponds to a coupling impedance which is all real (resistive). A resistively coupled pair of maintained oscillators has some curious properties;[11] among them is that the amplitude in both oscillators must be identical, regardless of the phase angle! This fortuitous result not only makes our vibration-cancellation job easier, but it greatly simplifies the analysis. It implies that if we assume a resonant coupler and let $\phi \simeq \pi$, we can set $\zeta = 1$ without any loss of information. Thus the extent of vibration cancellation is given by ϕ alone, and its rate of change with $\Delta\omega$ is given by $\partial\phi/\partial\Delta\omega$ alone.

It should be noted that for very narrow couplers, where the thermoviscous boundary layer is an appreciable fraction of the coupler radius, experiments and simulations show that

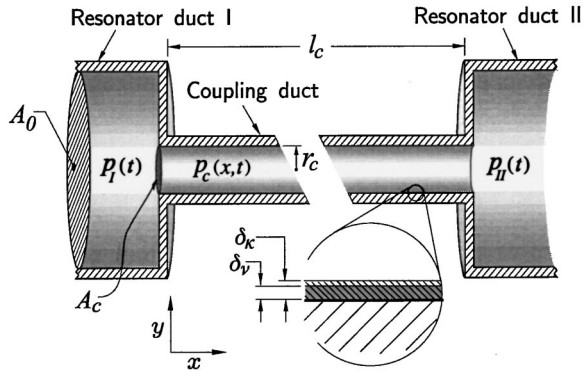


FIG. 3. Details of the junctions between two resonators and the coupling duct that connects them. For vibration cancellation, the system should be ''folded'' as shown in Fig. 2(b).

a coupler shorter than the resonant length provides the optimum vibration cancellation, even though it allows some amplitude imbalance.

Figure 3 shows the details of the interface between the resonators and the coupling duct. The oscillating pressure in resonator I at the junction with the coupler is $p_{\mathrm{I}}(t)$; that in resonator II is $p_{\mathrm{II}}(t)$. We let the end of the coupler at p_{I} be $x=0$, and that at p_{II} be $x=l_c$, where l_c is the length of the coupler. These pressures should be in antiphase in order to cancel vibration in the resonator cases. Note the blowup of the boundary layer near the coupler wall, indicating the viscous and thermal penetration depths δ_ν and δ_κ. We assume the two resonators (and the coupler) contain the same working gas in the same state, and the resonators are identical except for a small difference in natural frequency, due to a slight difference in gas temperature, for instance. Let the natural frequency of resonator I be ω_{I} and that of resonator II be ω_{II}; by ''natural frequency'' we mean the frequency that a resonator would have if it were uncoupled from the other. The mistuning can then be defined as

$$\Delta\omega=\omega_{\mathrm{II}}-\omega_{\mathrm{I}}. \tag{7}$$

We also assume the coupling is weak and perturbs only slightly the uncoupled behavior of each resonator. If the resonators are *not* locked, the oscillations in the resonators are modulated by a relatively slow beat envelope, like the organ pipes previously discussed, and the beating is slowed and eventually stopped by increasing the coupling. Thus we can write:

$$p_{\mathrm{I}}(t)=\Re\{P_{\mathrm{I}}(t)e^{i[\omega t+\phi_{\mathrm{I}}(t)]}\}, \tag{8}$$

$$p_{\mathrm{II}}(t)=\Re\{P_{\mathrm{II}}(t)e^{i[\omega t+\phi_{\mathrm{II}}(t)]}\}, \tag{9}$$

where $\Re\{\}$ indicates taking the real part. We assume $P_{\mathrm{I,II}}$ and $\phi_{\mathrm{I,II}}$ are real and slowly varying compared to ω, and $P_{\mathrm{I,II}}$ are always $\geqslant 0$. The relative phase ϕ and amplitude ratio ζ introduced in Eq. (3) may now be defined as:

$$\phi=\phi_{\mathrm{II}}-\phi_{\mathrm{I}}, \qquad \zeta=\frac{P_{\mathrm{II}}}{P_{\mathrm{I}}}. \tag{10}$$

We let the compromise frequency ω be a constant; the apparent *instantaneous* difference in frequency between resonators I and II is simply $\Delta\omega_{\mathrm{inst.}}=\partial\phi/\partial t$.

To obtain a pair of coupled differential equations, we start with a homogeneous equation for one of the maintained resonators, which it obeys when uncoupled:

$$\frac{d^2p_{\mathrm{I}}}{dt^2}+\omega_{\mathrm{I}}^2p_{\mathrm{I}}+\frac{\omega_{\mathrm{I}}}{Q}\left(1-\frac{P_0^n}{P_{\mathrm{I}}^n}\right)\frac{dp_I}{dt}=0. \tag{11}$$

The feedback that maintains the oscillations at or near frequency ω_{I} and amplitude P_0 is provided by the term $(\omega_{\mathrm{I}}/Q)(1-P_0^n/P_{\mathrm{I}}^n)dp_{\mathrm{I}}/dt$; this causes the amplitude to relax toward P_0 should the amplitude be initially higher or lower than P_0. This type of equation is found in Pippard[12] and in Van der Pol's paper on maintained triode oscillators.[13] The quantity Q is more or less the passive quality factor of the oscillator; this and the exponent n determine the relaxation time. For this work, the value of n is not important;[14] for further discussion, see Ref. 8. For convenience, we choose $n=2$.

Let $u_c(x,t)$ denote the x-component of particle velocity in the coupler, spatially averaged over the coupler cross-section. If, due to coupling, a volume velocity $-A_cu_c(0,t)$ enters the resonator from the coupler, the resonator obeys the inhomogeneous equation[15,16]

$$\frac{d^2p_{\mathrm{I}}}{dt^2}+\omega_{\mathrm{I}}^2p_{\mathrm{I}}+\frac{\omega_{\mathrm{I}}}{Q}\left(1-\frac{P_0^2}{P_{\mathrm{I}}^2}\right)\frac{dp_{\mathrm{I}}}{dt}=-\frac{2\rho a^2}{l_0}\left(\frac{A_c}{A_0}\right)\frac{du_c(0,t)}{dt}. \tag{12}$$

(The ''source term'' on the right-hand side has a minus sign because we define x and u_c positive pointing into the coupler from the resonator.) Here l_0 is the length of a resonator, A_0 is its cross-sectional area, A_c is the area of the coupler, ρ is the (average) density, and a is the (average) sound speed. A similar equation can be constructed for p_{II}. To couple the equations to one another, we solve for the complex particle velocity in the coupler $u_c(x,t)=\Re\{\tilde{u}_c(x)e^{i\omega t}\}$ (where $\sim$ indicates a complex quantity), subject to the boundary conditions $\tilde{p}_c(0)=P_{\mathrm{I}}e^{i\phi_{\mathrm{I}}}$ and $\tilde{p}_c(l_c)=P_{\mathrm{II}}e^{i\phi_{\mathrm{II}}}$, to obtain

$$\tilde{u}_c(x)=P_{\mathrm{I}}e^{i\phi_{\mathrm{I}}}\frac{\tilde{k}_c}{i\omega\rho}\left[\sin\tilde{k}_cx+\frac{\cos\tilde{k}_cl_c-\zeta e^{i\phi}}{\sin\tilde{k}_cl_c}\cos\tilde{k}_cx\right]. \tag{13}$$

Because the coupler is assumed lossy, we use a complex wave number $\tilde{k}_c$ which is approximately:

$$\tilde{k}_c\approx\frac{\omega}{a_c}\left[\left(1+\frac{\delta}{2r_c}\right)-\frac{\delta}{2r_c}i\right], \tag{14}$$

where

$$\delta=\delta_\nu+(\gamma-1)\delta_\kappa; \tag{15}$$

r_c is the radius of the coupler, a_c is the bulk speed of sound in the coupler, and δ_ν and δ_κ are the viscous and thermal penetration depths, respectively. Here we are making the simplifying, although not wholly accurate, assumption that $r_c\gg\delta_\kappa$ or δ_ν. Later we discuss how it might affect our results.

At this stage, our analysis has much in common with the work of Pippard[10] and Fletcher.[17] Were we interested in the dynamic behavior of the system as it approaches a locked state, we could continue the analysis by carrying out the time

derivatives in Eq. (12) and in the corresponding equation for $p_{II}(t)$, discarding small terms, and solving for $\partial\phi/\partial t$ and $\partial\zeta/\partial t$. However, we are interested primarily in the locked states, where $\partial\phi/\partial t=0$ and $\partial\zeta/\partial t=0$, so we assume simply that $d/dt=i\omega$. If we also assume that $\zeta=1$, i.e., $P_I=P_{II}=P$ (resonant coupler), the following pair of equations may be obtained:

$$-(\omega^2-\omega_I^2)+i\omega\frac{\omega_I}{Q}\left(1-\frac{P_0^2}{P^2}\right)$$

$$=-\frac{2a^2}{l_0}\left(\frac{A_c}{A_0}\right)\frac{\tilde{k}_c}{\sin\tilde{k}_c l_c}[\cos\tilde{k}_c l_c-e^{i\phi}] \tag{16}$$

and

$$-(\omega^2-\omega_{II}^2)+i\omega\frac{\omega_{II}}{Q}\left(1-\frac{P_0^2}{P^2}\right)$$

$$=-\frac{2a^2}{l_0}\left(\frac{A_c}{A_0}\right)\frac{\tilde{k}_c}{\sin\tilde{k}_c l_c}[\cos\tilde{k}_c l_c-e^{-i\phi}]. \tag{17}$$

Subtracting Eq. (16) from Eq. (17), while noting that $\omega_I\approx\omega_{II}\approx\omega$ and $|1-P_0^2/P^2|\ll 1$, gives

$$-\omega\Delta\omega\approx\frac{a^2}{l_0}\left(\frac{A_c}{A_0}\right)\frac{\tilde{k}_c}{\sin\tilde{k}_c l_c}2i\sin\phi. \tag{18}$$

A resonant coupler implies $\Re\{\tilde{k}_c l_c\}=\pi$; thus we have $\tilde{k}_c\approx\pi/l_c-i(\omega/a_c)(\delta/2r_c)$. With appropriate use of the approximations a_c, a_I, $a_{II}\approx a$, $l_c\approx l_0$, and $\delta/r_c\ll 1$, taking the real part of Eq. (18) and rearranging gives

$$\sin\phi\approx-\frac{\pi^2}{4\omega}\left(\frac{A_0}{A_c}\right)\frac{\delta}{r_c}\Delta\omega. \tag{19}$$

Finally, we let $\phi\simeq\pi$, and differentiate with respect to $\Delta\omega$; the result can be expressed in dimensionless form as:

$$\left(\frac{\omega}{\pi}\right)\frac{\partial\phi}{\partial\Delta\omega}=\frac{\pi}{4}\left(\frac{A_0}{A_c}\right)\frac{\delta}{r_c}. \tag{20}$$

The factor $1/\pi$ on the left-hand side effectively normalizes the phase angle—most mode-locked systems (including our acoustic engines) become unlocked at $\phi=\pm\pi/2$, so the range of ϕ is about π radians.

This equation, although clearly the result of many approximations, has the advantage of being simple to interpret. It tells us how the phase of the locked state changes as a function of the (fractional) mistuning $\Delta\omega/\omega$, the geometry, and the thermoviscous properties of the working fluid. In the present context, this effectively tells us how the degree of vibration cancellation is affected by all these parameters. The factor A_0/A_c is the ratio of resonator to coupler area; the factor r_c/δ is approximately the quality factor Q of the coupler.[18] In general, we want $\partial\phi/\partial\Delta\omega$ to be small—i.e., we want vibration to increase only slightly as our engines become more mistuned; however, we also want to keep the coupler relatively narrow to reduce losses (see the end of Sec. I A). The only obvious solution, according to Eq. (20), is to keep the thermoviscous penetration depth δ small compared to the radius r_c of the coupler—in other words, use a high-Q coupler. This suggests that if the working fluid and its ambient state are fixed, the best strategy is to increase all the diameters, thus enabling A_0/A_c to be large while keeping r_c/δ small.

To complement the preceding discussion of how to optimize the coupling, we consider briefly the question of when Eq. (20) is a valid approximation. Equation (20) was derived assuming weak coupling between the resonators; it is fair to suppose that if r_c/δ is big enough, i.e., if the coupler has a high enough Q, then the coupling is not weak. Another way of looking at it, of course, is that the larger the system, or the thinner the boundary layer in the working fluid, the more narrow one can afford to make the coupler and still get the same coupling and the same degree of vibration cancellation.

Another important assumption is that the penetration depth δ is small compared to the coupler radius r_c. For our narrowest couplers, δ/r_c is approximately 0.3; thus it is natural to wonder if Eq. (20) still applies. One could carry all computations out to second order; we will not do so here, but we will mention two corrections one might plausibly add. First, the effective cross-sectional area with which the coupler communicates with the resonators is a little less than πr_c^2, because some of the fluid is viscously clamped at the coupler walls. Thus we might suppose that A_c in Eq. (20) should be replaced by some reduced area A_c', where[18]

$$A_c'\approx\pi r_c^2\left(1-\frac{\delta_\nu}{r_c}\right). \tag{21}$$

Second, a similar correction applies to the length l_c. Since the coupler is narrow and lossy, its effective speed of sound is slower than in the resonators; thus the length that makes the coupler resonant is shorter than l_0, and changes with its radius. Recall that requiring that the coupler be in a half-wavelength resonance is equivalent to making the real part of $\tilde{k}_c l_c$ equal to π; an approximate result for l_c is

$$l_c(\text{resonant})\approx\frac{\pi a_c}{\omega}\left(1-\frac{\delta}{2r_c}\right). \tag{22}$$

Had we not assumed $l_c=l_0$ in the derivation of Eq. (20), the factor $l_c A_0/l_0 A_c$ would have appeared in the result instead of A_0/A_c, so the two corrections above nearly cancel one another. We might expect, then, that Eq. (20) is reasonably accurate even for very narrow couplers.

It is also worth noting that Eq. (20) is independent of any characteristics of the mechanism driving the resonators. Thus Eq. (20) should apply equally well to organ pipes or thermoacoustic engines, when they are acoustically coupled by a resonant tube. However, this is generally *not* true for other (e.g., reactive) types of coupling, such as ''mass coupling'' through the shared structure.

In the following sections, we describe experiments that measure $\partial\phi/\partial\Delta\omega$ for a variety of coupler diameters, and a computer model that simulates these experiments. Both experiment and simulation agree well with the predictions of Eq. (20). Experiments and simulations are also presented for

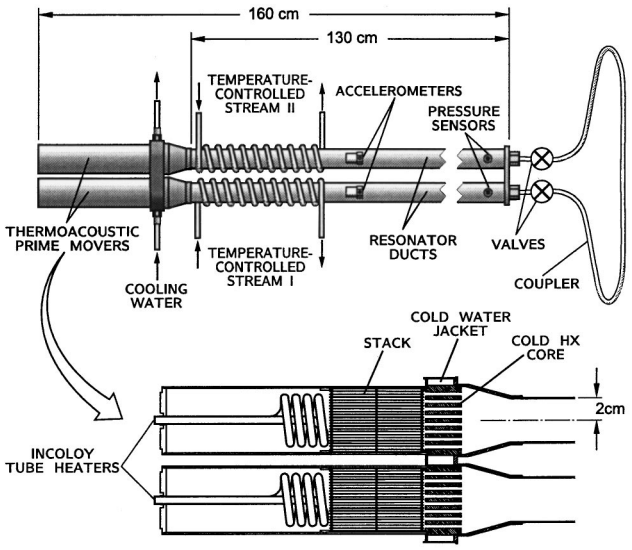


FIG. 4. Experimental setup for exploring the behavior of coupled thermoacoustic engines.

coupled engines whose resonators are not uniform-diameter ducts, in which case Eq. (20) is not quantitatively applicable.

II. APPARATUS

To explore mode-locking of acoustic engines experimentally, we have built two nearly identical thermoacoustic engines which are rigidly attached to each other in a side-by-side fashion. Two neighboring ends of the engines are connected by a coupling duct which can be inserted or removed from the system by means of valves. These are ''straight pattern'' ball valves, which look like a section of straight duct when open, the fluid not following any kind of tortuous path. Each engine consists of a straight, circular duct of uniform diameter (the resonator) connected to a thermoacoustic prime mover, by which we mean a heater, a cold heat exchanger, and a stack of parallel plates in between; these sustain the oscillations in the resonator when the engine is running.

Figure 4 shows the essentials of the experimental setup, and some details of the thermoacoustic prime mover hardware. Note that the engines share a water jacket that cools their cold heat exchangers, and both are welded to a common plate at the other end. A section of each resonator is wrapped with copper refrigeration tubing, which circulates water from a temperature-controlled bath. Thus the gas temperature in each engine can be varied independently, enabling control of $\Delta\omega$, the difference in natural frequencies. A pressure sensor is mounted at the end of each resonator where it joins the coupler. We take the signal from the sensor in resonator I (pressure sensor I) to represent $p_{\mathrm{I}}(t)$, and that from pressure sensor II to represent $p_{\mathrm{II}}(t)$. Using p_{I} as a reference to a lock-in amplifier, and using p_{II} as the input signal, allows a direct measurement of ϕ. Independent monitoring of pressure sensors I and II allows one to obtain $P_{\mathrm{I}}(t)$ and $P_{\mathrm{II}}(t)$. Accelerometers on the cases measure case vibration in the axial (x) direction, and numerous thermocouples (not pictured) provide data on the temperature of the gas near the hot end of the stack, the cooling water temperature, the water jacket temperatures, and the temperature of the gas inside the resonators. Multimeters monitor the input voltage and current to the tube heaters that power the engines. The accelerometers are used to verify that the two engines welded together act like a rigid structure, and that vibrations are indeed canceled when the phase between the engines approaches π—we find that cancellation of case vibration is at least 99.5% complete when $\phi=\pi\pm0.002$.

In order to confirm our analytic and experimental results, we have simulated our system using DELTAE[19] (Design Environment for Low-amplitude ThermoAcoustic Engines). DELTAE integrates the one-dimensional wave equation in the small-amplitude (''acoustic'') approximation, including thermal and viscous effects, according to the thermoacoustic theory of Rott.[20] In our numerical model, we have four geometrically identical engines, two coupled by a narrow duct (which share a common frequency of operation) and two running independently; the gas temperatures in the resonator portion of each of the two coupled engines are varied in order to vary $\Delta\omega$, and the ϕ and ζ values that solve the wave equation in this system are returned by DELTAE. The actual $\Delta\omega$ that exists between the simulated coupled engines is estimated by forcing each uncoupled engine in the model to have the same temperature distribution as one of the coupled engines. For the coupled engines, the input heater power is assumed to be the same in both; however, in order to match gas temperatures at both ends of the stack, the uncoupled engines must use the input heater power as a ''guess'' which can be adjusted to meet the ''targets'' (which include the gas temperatures at the ends of the stack).

Measuring the difference in natural frequencies of the engines is subtle, because it cannot be accomplished when the engines are coupled. The most obvious way to measure $\Delta\omega$ is to measure the beat frequency of the engines when the coupler valves are closed, using a dual-channel oscilloscope to observe p_{I} and p_{II} and timing the beats with a stopwatch. Since $\Delta\omega=\partial\phi/\partial t$ when the engines are uncoupled, where $\partial\phi/\partial t$ is in radians per second, a quick and accurate measurement of $\Delta\omega$ is obtained. Unfortunately, the ''natural frequency'' of an engine is dominated by its mean temperature $T_m(x)$; the temperature at the ''hot end'' of the engine (T_H), in turn, depends on the impedance the prime mover sees looking into the resonator, which depends on whether the coupler valves are open or closed. For instance, assume the input power to an electric heater as shown in Fig. 4 is constant; if the coupler valve, initially closed, is then opened, the load on the engine is increased, causing the pressure amplitude to drop and T_H to rise.[21] Likewise, if the valve is suddenly closed, T_H will begin to fall. The ''natural frequency'' we are trying to measure is, of course, the frequency the engine *would* have with the coupler valves closed, but with the *same* temperature distribution $T_m(x)$ it has in the locked state with the coupler valves open. When we attempt to measure the $\Delta\omega$ corresponding to a given locked state by opening the coupler valves and observing $\partial\phi/\partial t$, we are actually looking at the engines in an ambiguous state where their temperatures, and hence their frequencies, are changing with time.

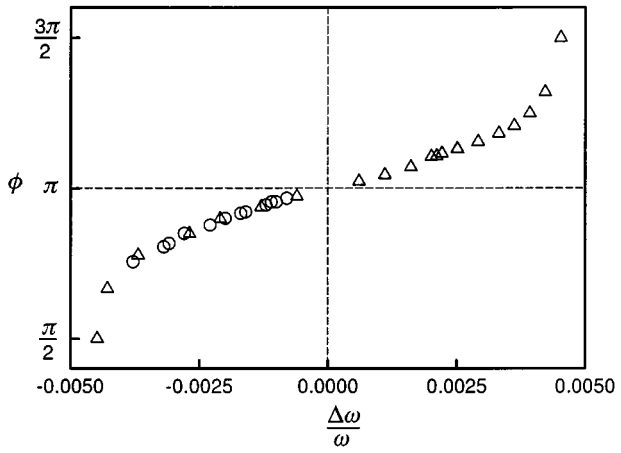


FIG. 5. Relative phase ϕ versus fractional mistuning $\Delta\omega/\omega$ for two acoustically coupled thermoacoustic engines. For the circular data points, the engines were allowed to thermally equilibrate with the coupling valves open; for the triangular points, the engines equilibrated with the coupling valves closed.

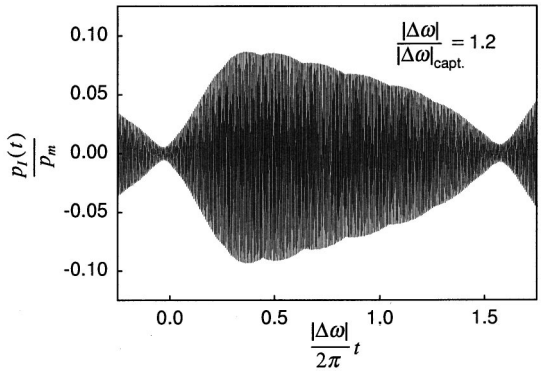


FIG. 6. Oscillating pressure versus time for one of two coupled thermoacoustic engines, when their frequency mismatch $\Delta\omega$ is 20% larger than the capture bandwidth. The oscillating pressure $p_1(t)$ is normalized to the mean pressure p_m, and the time t is normalized to the beat period in the absence of coupling, $2\pi/\Delta\omega$, thus showing the period elongation. (The slight rippling of the beat envelope is due to aliasing.)

In this work, we employ two basic strategies for resolving this ambiguity:

(1) Measure $\phi(\Delta\omega)$ two ways:
   (a) Let engines equilibrate with coupler valves open; measure ϕ. Close valves and immediately measure $\Delta\omega$.
   (b) Let engines equilibrate with coupler valves closed; measure $\Delta\omega$. Open valves and immediately measure ϕ. Compare with (a).
(2) In computer model, force $T_m(x)$ in each engine to be the same in both coupled and uncoupled states; compare simulation results for $\phi(\Delta\omega)$ with 1(a) and 1(b).

We find that for the purposes of the present work, the difference between the results obtained by 1(a) and 1(b) is small; since we are interested in characterizing the locked state of the engines, method 1(a) is used for all the data presented in this paper, unless otherwise specified.

Changes in T_H with load do not interfere too much with the measurement of $\Delta\omega$ for two reasons: $\Delta\omega$ is dominated by the difference in resonator gas temperature, which is set by the separate temperature control coils; and the engines never have a large amplitude difference with a resonant coupler, so any change in T_H with load tends to be the same in both engines. These conditions do not necessarily exist for other types of coupling, such mass coupling.

III. RESULTS

Figure 5 shows typical behavior of the phase ϕ between the engines in the locked state versus the fractional mistuning $\Delta\omega/\omega$, for the system shown in Fig. 4. The coupler has about a 3-mm inner diameter; the working fluid is air at 0.8 bar, and the frequency of operation is about 100 Hz. The circles are data points taken according to method 1(a) mentioned in the previous section, where the engines equilibrate with the coupler valves open; the triangles are data points taken using method 1(b), where the engines equilibrate with the valves closed. Apparently the two methods agree well.

Figure 5 displays a *capture bandwidth* of $(|\Delta\omega|/\omega)_{\text{capt.}} \approx 0.004$, or 0.4%, with band-edges near $\phi = \pi \pm \pi/2$, where the capture bandwidth is defined as the maximum natural frequency difference the engines can have and still achieve locking. It is apparent that $\phi(\Delta\omega)$ has a linear region near $\phi = \pi$; this is where we expect Eq. (20) to apply. Nearer the capture band-edge, ϕ changes very rapidly with $\Delta\omega$, until the engines unlock at $\phi = \pi \pm \pi/2$. The data points shown at $\phi = \pi \pm \pi/2$ are not in fact true, stable locked states; the phase tends to oscillate with a period of about a minute with an amplitude of $\pi/10$ or more. These states are hysteretic, in that locking initially occurs only for $0.7\pi < \phi < 1.3\pi$, but after locking is achieved, it will hold for phase angles closer to the band-edge. Just beyond the band-edge, the engines are unlocked and beat against each other; unlike a linear system, the beat frequency is slower than $\Delta\omega$ and the beat envelope is not sinusoidal. These features are evident in Fig. 6, which shows one beat cycle of the pressure p_1 when the engines have been detuned 20% beyond the capture bandwidth.

One other feature of Fig. 5 worth noting is the gap in the data near $\phi = \pi$, the region of most interest to us. When $\Delta\omega$ is sufficiently small, the ability of the engines to exert forces on each other through their shared structure, what we have called the ''mass coupling,'' is sufficient to elongate the beat period or even cause mode-locking; thus it is impossible to get an accurate estimate of $\Delta\omega$ by the methods mentioned previously, since the engines cannot be considered uncoupled when the coupler valves are closed. We detail mass coupling phenomena in a separate paper;[8] here we simply note that the capture bandwidth for mass coupling depends on how massive the case is compared to the gas inside, the bandwidth being larger when the case is relatively light. In the present work, therefore, we eliminate the mass coupling in subsequent data sets by attaching additional mass to our twin-engine assembly (we add an extra 130 kg of mass, which increases the system's solid mass by more than a factor of 10).

Figure 7 shows $\phi - \pi$ and $\zeta - 1$ for our mode-locked engines near $\phi = \pi$; the working fluid is air at 0.8 bar and 300 K, and the coupler is 3.2 mm in diameter and its length

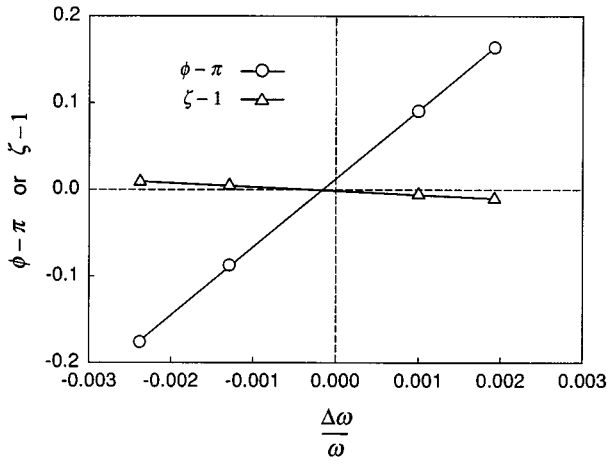


FIG. 7. Plots of $\phi-\pi$ and $\zeta-1$ vs $\Delta\omega/\omega$ for system of Fig. 4, where working fluid is air at 0.8 bar and 300 K; coupler has a diameter of 3.2 mm, and its length is chosen to make it resonant. Note that $\phi-\pi$ and $\zeta-1$ are plotted to the same scale.

is chosen to make it resonant. Recall from Eq. (6) that the cancellation ratio $R_V \approx 1-[(\phi-\pi)^2+(\zeta-1)^2]^{1/2}$; with the resonant coupler in place, it is apparent that the case vibration is dominated by the phase angle ϕ, the amplitude mismatch $\zeta-1$ being relatively negligible, as predicted by theory. Additional measurements verify that if the coupler length l_c is made even a few percent larger or smaller than the resonant length $(\pi a_c/\omega)(1-\delta/2r_c)$, $\zeta-1$ increases dramatically.

The main objective of our experiments is to measure $\partial\phi/\partial\Delta\omega$ as a function of coupling strength, which we vary by changing coupler diameter. Data of the type in Fig. 7 were recorded with air at 0.8 bar, 300 K as the working fluid, for couplers of five different diameters, ranging from 2.7 mm ($A_0/A_c=185$, $\delta/r_c=0.29$) to 4.8 mm ($A_0/A_c=58$, $\delta/r_c=0.16$). In all cases, the resonator radius r_0 remains constant at 1.82 cm, and the coupler length l_c is chosen to make the coupler resonant. The results are summarized in Fig. 8, in which each data point represents the slope of a data set like that of Fig. 7, scaled by $1/\pi$. It is evident that experiment, theory, and simulation are all in excellent agreement, implying that the mode-locking of uniform diameter, circular duct resonators coupled by a resonant, half-wavelength duct is

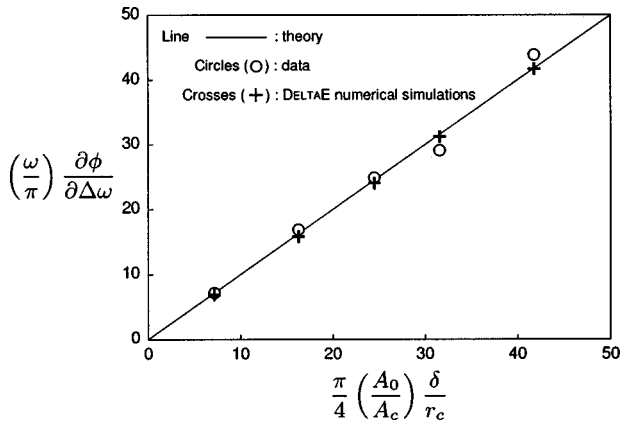


FIG. 8. Summary of $\partial\phi/\partial\Delta\omega$ results for thermoacoustic engines coupled by resonant half-wavelength ducts of various diameters.

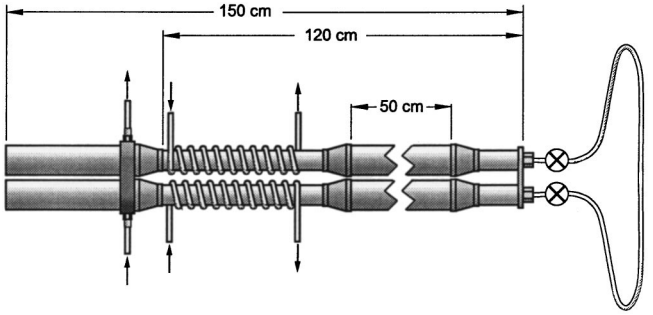


FIG. 9. Engines used for the measurements in Sec. III; these use the same prime movers as the first set of measurements, but the resonator shape has been altered to suppress distortion at high amplitudes.

well understood to first order. The disagreements between experiment and theory are on the order of a few percent, which is about the precision to which the inner radii of the couplers are determined.

Engine resonators are not necessarily cylindrical ducts of uniform cross section, as was assumed when deriving Eq. (20) and other formulas. A resonator with varying cross section may be shorter than a uniform duct of the same natural frequency, it may suppress harmonic generation at high amplitudes, and it may decrease dissipation at low amplitudes. Rather than present a theory[8] for arbitrary resonator shape, we consider one limiting case, to gain insight, and compare its predictions with experiments and simulations.

We consider the limiting case of two coupled Helmholtz resonators, each consisting of two volumes V_1 and V_2 connected by a duct area A_0 and length l. The two V_2's are connected by a coupling duct of radius r_c and length $l_c \sim \lambda/2$. We find that

$$\left(\frac{\omega}{\pi}\right)\frac{\partial\phi}{\partial\Delta\omega} \simeq \frac{\pi}{4}\left(\frac{A_0}{A_c}\right)\frac{\delta}{r_c}\left[4\,\frac{V_1+V_2}{\lambda A_0}\left(\frac{V_2}{V_1}\right)\right] = \frac{\pi}{4}\left(\frac{A_0}{A_c}\right)\frac{\delta}{r_c}F_H, \tag{23}$$

where F_H is what we call the ''Helmholtz factor.'' This formula reverts to Eq. (20), the half-wave resonator result, if $V_1=V_2=V=A_0(\lambda/8)$, and if $l=(4/\pi^2)\lambda$ (which forces the Helmholtz resonator frequency to be the same as the half-wave resonator). Together, l and V describe a plausible mass-spring model of a half-wave resonator. This says the ''effective mass,'' in this context, of the oscillating fluid in a half-wave resonator is about 0.8 of its total mass.

Most importantly, we note that $\partial\phi/\partial\Delta\omega$ has the same dependence on r_c and δ for coupled Helmholtz resonators as for half-wave ducts. This suggests that any data on resonators of arbitrary shape should still lie on a straight line on a graph such as Fig. 8, but with a slope that may differ from unity.

Experiments were done using a setup identical to that shown in Fig. 4, except the resonators were altered as shown in Fig. 9, with a shorter overall length and a sizable section of duct near the coupled ends that is about twice the cross-sectional area of the original duct. This particular design was chosen to make the first and second overtones of the resona-

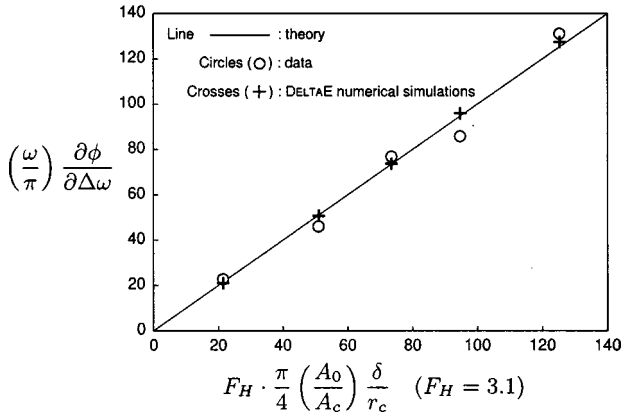


FIG. 10. Summary of $\partial\phi/\partial\Delta\omega$ results for thermoacoustic engines coupled by resonant half-wavelength ducts of various diameters, with the engine resonators having the shape pictured in Fig. 9.

tor lie half-way between harmonics of the fundamental, to suppress distortion at high amplitudes, while keeping the fundamental frequency the same as before. Data of the type shown in Fig. 7 were taken for the same five couplers; the $\partial\phi/\partial\Delta\omega$ results, along with DELTAE simulations of this system, are summarized in Fig. 10. The abscissa has one adjustable parameter, F_H [as defined in Eq. (23)], which is chosen to best fit the data. With F_H properly adjusted, experiment, theory and simulation are in excellent agreement. The general theory for arbitrarily shaped resonators has been derived by us.[8] These results, and those in the previous section, give us confidence that DELTAE can be relied upon to help design systems more complex than those we can readily analyze.

IV. APPLICATION TO VIBRATION CANCELLATION

The application of the results in the previous section to the vibration cancellation problem is straightforward. Typically, one might desire a given level of vibration cancellation, with two engines that are likely to experience a known (or estimated) range of detuning, and one would like to know what length and radius coupler is necessary to achieve the desired cancellation. In other words, given a desired R_V as in Eq. (6), and an expected $|\Delta\omega|$, we want to find l_c and r_c. We assume that the working fluid and its state are fixed; thus δ is known. If we assume a resonant coupler, then $|\phi-\pi| \gg |\zeta-1|$, and $R_V \approx 1-|\phi-\pi|$. Then the slope of $\phi(\Delta\omega)$ at zero is related to the desired vibration cancellation and the maximum estimated $|\Delta\omega/\omega|$ by

$$\frac{1-R_V(\text{desired})}{|\Delta\omega/\omega|(\text{max, est})} = \omega\frac{\partial\phi}{\partial\Delta\omega} = \frac{\pi^2}{4}\left(\frac{\pi r_0^2}{\pi r_c^2}\right)\frac{\delta}{r_c}F_H. \quad (24)$$

This can be solved for r_c to yield

$$r_c = \left[\frac{\pi^2}{4}r_0^2\delta\frac{|\Delta\omega/\omega|}{1-R_V}F_H\right]^{1/3}; \quad (25)$$

the resulting value for r_c can be substituted along with δ into Eq. (22) to obtain l_c.

Suppose, for instance, that we have two engines using helium gas at 30 bar, 300 K, and whose resonators have a radius of 15 cm, with $F_H=1$ (e.g., resonators of uniform diameter). Suppose they may vary as much as 1% in natural frequency during operation, and we want to eliminate 95% of the vibration. Thus we have $\delta=0.3$ mm, $r_0=15$ cm, $|\Delta\omega/\omega|=0.01$, and $R_V=0.95$. According to Eq. (25), the minimum coupler radius is $r_c \approx 1.5$ cm.

An additional constraint one is likely to face is the maximum additional dissipation, due to the coupler, that can be tolerated. According to the simple ideas at the end of Sec. I A, the dissipation in the coupler is simply (r_c/r_0) times the dissipation in a main resonator.[22] This is only strictly true at $\phi=\pi$, where the pressure and velocity profiles in the coupler are the same as in the resonators. When $\phi\neq\pi$, the asymmetry in the boundary conditions of the coupler causes its pressure and velocity amplitudes to increase; this can be seen by looking at Eq. (13). Let the thermoviscous dissipation in a resonator be $\dot{E}_0$, and that in the coupler be $\dot{E}_c$. If we assume a resonant coupler, and $\phi\approx\pi$, hence $\zeta=1$, we find that

$$\frac{\dot{E}_c}{\dot{E}_0} \approx \frac{r_c}{r_0}\left[1+\left(\frac{|\phi-\pi|}{\pi}\frac{r_c}{\delta/2}\right)^2\right]; \quad (26)$$

we can use Eq. (23), the relationship between $|\phi-\pi|$ and $\Delta\omega$, to rewrite this as:

$$\frac{\dot{E}_c}{\dot{E}_0} \approx \frac{r_c}{r_0}\left[1+\frac{\pi^2}{4}F_H\left(\frac{\Delta\omega}{\omega}\right)^2\left(\frac{r_0}{r_c}\right)^4\right]. \quad (27)$$

For a given $\Delta\omega/\omega$, there will exist an optimum r_c/r_0 for which the dissipation in the coupler is a minimum; it is found to be

$$\left.\frac{r_c}{r_0}\right|_{\text{opt.}} \approx \left[\frac{3\pi^2}{4}F_H\left(\frac{\Delta\omega}{\omega}\right)^2\right]^{1/4}. \quad (28)$$

For the large-diameter system considered above, where the maximum $|\Delta\omega/\omega|=1\%$, this formula predicts $(r_c/r_0)_{\text{opt.}}=1/6$ [DELTAE predicts $(r_c/r_0)_{\text{opt.}}=0.17$]; if $r_0=15$ cm, this implies $r_c=2.5$ cm, about 5/3 times the minimum radius needed to achieve the desired vibration cancellation. In other words, because of the increased dissipation that occurs with increasing $\Delta\phi$, one gets less dissipation overall by using a larger-diameter coupler and keeping $\Delta\phi$ small, even though this increases the surface area of the coupler. In this example, the 2.5-cm-radius coupler dissipation is about 15% of resonator dissipation when $|\Delta\omega/\omega|=0$, and increases to a little over 20% when $|\Delta\omega/\omega|=0.01$. On the other hand, the 1.5-cm-radius coupler dissipation is only 10% of resonator dissipation at $|\Delta\omega/\omega|=0$, but rises to over 30% when $|\Delta\omega/\omega|=0.01$.

Thus in choosing the proper coupler radius, one must consider not only how much vibration or dissipation can be tolerated, but also how often and for how long the engines might be expected to operate in the extremes of the expected range of detuning. It may be better to use the minimum r_c specified by Eq. (25), or an even narrower one, if the engines will only rarely experience significant detuning. Equations such as Eq. (23), Eq. (25), and Eq. (28) are but helpful guides in making an intelligent choice.

The above results apply in the low-amplitude limit, where turbulence is unimportant. To achieve a useful power density, a practical device may need to operate at a high amplitude. In the limit of very high Reynolds number, where the Moody friction factor[23] becomes nearly constant with amplitude, we estimate that the same optimum ratio (r_c/r_0) as before will give the minimum dissipation. A more complicated, but not implausible, scenario is one where the narrow coupler has a much higher peak Moody friction factor than the resonator, in which case high velocities in the coupler could be very costly. Widening the coupler would not only reduce $\Delta\phi$, reducing the peak velocity in the coupler, but also reduce its friction factor. Thus the amplitude at which the engines operate is another consideration in optimizing the coupling.

This work is exclusively devoted to couplers which are circular ducts of constant cross-section, of nominal length $\lambda/2$. One may wonder if other coupler shapes might provide the same degree of coupling, with fewer losses. We have little experimental data on alternate coupler shapes, but in our computer simulations, we have explored many different coupler shapes and designs, as well as placement of the coupler at points on the resonator other than the pressure antinodes. We have never found a design that gives as much coupling for as little loss as a uniform-diameter circular duct. It is often advantageous to ''neck down'' a portion of a resonator, because compared to a uniform-diameter duct at the same frequency, it reduces the surface area for dissipation more than it increases the particle velocity, resulting in less dissipation overall.[24] For a coupler, however, necking down the duct decreases the coupling strength; this and other factors make it appear that the uniform-diameter coupler is optimum.

ACKNOWLEDGMENTS

The authors wish to thank Bob Keolian and Bob Ecke for teaching us about mode-locking, Bill Ward, Scott Backhaus, and Bob Hiller for useful discussions, and David Gardner for assistance with the experiments. This work has been supported by the offices of Fossil Energy and Basic Energy Sciences in the U.S. Department of Energy.

[1] G. W. Swift, ''Thermoacoustic engines and refrigerators,'' Phys. Today **48**, 22–28 (July 1995); S. Backhaus and G. W. Swift, ''A thermoacoustic Stirling heat engine,'' Nature **399**, 335–338 (1999).

[2] C. Sondhauss, ''Ueber die schallschwingungen der luft in erhitzten glasröhren und in gedeckten pfeifen von ungleicher weite,'' Ann. Phys. (Leipzig) **79**, 1 (1850).

[3] ''A simple theory of the Sondhauss tube,'' in *Recent Advances in Aeroacoustics*, edited by A. Krothapalli and C. A. Smith (Springer, New York, 1984), p. 327.

[4] P. L. Rijke, ''Notiz über eine neue art, die in einer an beiden enden offenen röhre enthaltene luft in schwingungen zu versetzen,'' Ann. Phys. (Leipzig) **107**, 339 (1859).

[5] G. W. Swift, ''Thermoacoustic natural gas liquefier,'' 1997, in Proceedings of the DOE Natural Gas Conference, Houston TX, March 1997.

[6] W. A. Marrison, U.S. Patent No. 2,836,033 (1958).

[7] C. Huygens, in *Ouevres Completes de Christian Huyghens*, edited by M. Nijhoff, Vol. 5, p. 243 (Societe Hollandaise des Sciences, The Hague, The Netherlands, 1893) (A letter to his father, dated 26 Feb. 1665).

[8] P. S. Spoor and G. W. Swift, ''The Huygens phenomenon in thermoacoustics,'' in preparation for J. Acoust. Soc. Am.

[9] In Ref. 8, we find that mass coupling obeys $\partial\phi/\partial\Delta\omega = 1/(\omega Q\mu^2)$, where μ is the ratio of working fluid mass to case mass $(m_{\rm gas}/m_{\rm case})$. For the realistic, large-diameter system (Ref. 5) discussed in Sec. IV in the text, comparing this result with Eq. (20) shows that the acoustic coupling is of order 10 000 times stronger than the mass coupling.

[10] A. B. Pippard, *The Physics of Vibration* (*Omnibus Edition*) (Cambridge University Press, Cambridge, 1989), Chap. 12.

[11] Reference 10, p. 403.

[12] Reference 10, p. 323.

[13] B. Van der Pol, ''The nonlinear theory of electric oscillations,'' Proc. Inst. Rad. Eng. **22**, 1051 (1934).

[14] The value of n can be determined either from the relaxation time of the oscillator from one amplitude to another, or from changes in steady-state amplitude in response to changes in load. For our thermoacoustic engines, only the steady state is well characterized, for when the engine is relaxing from one state to another, the power input is ambiguous, the thermal mass of the engine absorbing or supplying heat at a variable rate. Changes in steady state amplitude only determine the quantity n/Q, and cannot obtain n or Q independently. Fortunately, for the purposes of this work and that in Ref. 8, this is sufficient.

[15] P. M. Morse, *Vibration and Sound* (American Institute of Physics, Woodbury, NY, 1986), p. 313.

[16] S. Backhaus, ''Search for acoustic Josephson radiation in superfluid ^{4}He,'' Ph.D. thesis, University of California at Berkeley, 1997.

[17] N. H. Fletcher, ''Mode-locking in nonlinearly excited inharmonic musical oscillators,'' J. Acoust. Soc. Am. **64**, 1566 (1978). See also the URL www.csu.edu.au/ci/vol1/Neville.Fletcher/.

[18] A. M. Fusco, W. C. Ward, and G. W. Swift, ''Two-sensor power measurements in lossy ducts,'' J. Acoust. Soc. Am. **91**, 2229–2235 (1992).

[19] W. C. Ward and G. W. Swift, ''Design environment for low amplitude thermoacoustic engines (DELTAE),'' J. Acoust. Soc. Am. **95**, 3671–3672 (1994). Fully tested software and user's guide available from Energy Science and Technology Software Center, US Department of Energy, Oak Ridge, Tennessee. To review DeltaE's capabilities, visit the Los Alamos thermoacoustics web site at www.lanl.gov/thermoacoustics/. For a beta-test version, contact ww@lanl.gov (Bill Ward) via Internet.

[20] N. Rott, ''Thermoacoustics,'' Adv. Appl. Mech. **20**, 135–175 (1980).

[21] J. R. Olson and G. W. Swift, ''A loaded thermoacoustic engine,'' J. Acoust. Soc. Am. **98**, 2690–2693 (1995).

[22] The ratio of coupler to resonator dissipation, even at $\phi = \pi$, is generally a little higher than r_c/r_0, because the engines never have a completely uniform cross-section, which makes them shorter for a given frequency, thus lessening their surface areas. In Fig. 4, the hot ends of the engines are wider than the resonators, to provide more power, and the length of the engines is hence shortened to 0.44λ.

[23] L. F. Moody, ''Friction factors for pipe flow,'' Trans. ASME **66**, 671–684 (1944).

[24] T. J. Hofler, ''Thermoacoustic refrigerator design and performance,'' Ph.D. thesis, Physics department, University of California, San Diego, 1986.

A finite element analysis of an interferometric optical fiber hydrophone with a concentric composite mandrel including a foaming layer

Jong-in Im
Functional Ceramic Division, Research Institute of Industrial Science & Technology, Pohang 790-330, Korea

Yongrae Roh[a)]
Department of Sensor Engineering/Department of Electronic Engineering, Kyungpook National University, 1370 Sankyuk-Dong, Buk-Gu, Taegu 702-701, Korea

(Received 1 October 1998; accepted for publication 3 June 1999)

This paper outlines the design of an optical fiber hydrophone with a concentric composite mandrel that has a fundamental resonance frequency over 15 kHz and which demonstrates a good sensitivity in underwater conditions over 200 m. The composite mandrel consists of double layers (a thin foaming layer on top of a base layer) and a center hole. The foaming layer is coated on top of the base layer, made of a common material (aluminum), so as to radically improve the sensitivity of the sensor through its superior compliance. The optimal structure of the composite mandrel was determined using the finite element method (FEM) to analyze the influence of both the material properties of the foaming layer and the geometry of the composite mandrel on the performance of the hydrophone. The geometrical parameters included the thickness of the foaming layer, the inner and outer diameters, and the length of the mandrel. The analysis results indicated that for a higher sensitivity, a hydrophone needed as thick a foaming layer as possible, made from a material with a relatively low Young's modulus, if possible, less than 1 GPa. The sensitivity also increased with a smaller outer diameter and longer mandrel, and additional sensitivity was achieved by decreasing the ratio of the inner diameter to the outer diameter. The sensitivity of the optimal structure Al composite mandrel was about -82 dB in relation to 1 rad/μPa. This result is 14 dB higher than that of a simple concentric Al mandrel-type hydrophone of the same dimensions, which verifies the effectiveness of the foaming layer.
[S0001-4966(99)02609-0]

PACS numbers: 43.38.Ar, 43.30.Yj, 43.58.Fm [SLE]

INTRODUCTION

Optical fiber sensing is a subject that has attracted considerable attention in recent years.[1–3] A finite element analysis of an interferometric optical fiber hydrophone for underwater acoustic sensing was included in a previous report.[4] However, the hydrophone outlined in the previous work was only a basic format, consisting of an optical fiber coil encapsulated in a homogeneous solid cylindrical mandrel with a thin polymer layer to protect the fiber. Interaction with acoustic fields causes deformation in the mandrel, resulting in a change in the optical path length of the fiber coil which is subsequently detected as a phase change in the light passing through the fiber coil. Therefore, although this design has the advantages of simplicity, reliability, and robustness, its inherent sensitivity to acoustic fields is relatively low.[5–7] Accordingly, this paper outlines a concentric composite mandrel-type optical fiber sensor that can produce a performance superior to the basic cylindrical mandrel sensor.

In the previous report,[4] it was noted that the solid mandrel should be as compliant as possible to enhance the interaction with the acoustic fields. However, a solid mandrel must also support the whole sensor structure; thus, for the sake of structural stability it cannot be too compliant. To overcome this difficulty, several attempts have been made to make the mandrel appear more compliant in its overall properties. Most representative examples include the insertion of an air cavity inside the solid mandrel,[8,9] and the inclusion of a center hole.[5] However, the insertion of an air cavity can cause problems in the structural robustness of the sensor, especially in a deep underwater environment, and involve a higher difficulty in the mechanical realization of the mandrel. Hence, a concentric composite mandrel is proposed as an alternative that can improve sensitivity while maintaining structural stability. A composite mandrel consists of double layers (a thin foaming layer on top of a base layer) and a center hole. The center hole is a commonly employed method for enlarging the deformation of the mandrel. The originality of the proposed design is the addition of a compliant foaming layer. The foaming layer is more flexible than the base layer, and is expected to radically improve the sensitivity of the sensor due to its superior compliance. Figure 1 illustrates the details of the proposed concentric composite mandrel structure.

In a fiber-wound mandrel sensor (see Fig. 1), a thin jacket fiber (layer C) is wrapped around the mandrel (layer A + layer B). This fiber jacket is then coated for pro-

[a)]Electronic mail: yryong@eeg.kyungpook.ac.kr

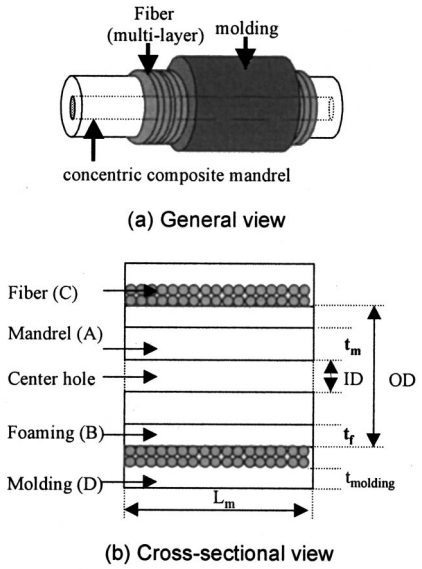


FIG. 1. Schematic structure of an optical fiber hydrophone with a concentric composite mandrel.

tection by a thin polymer (layer D) called a ''molding,'' less than 1 mm in thickness. The proposed mandrel has a composite structure consisting of concentric double layers and a center hole. The double layers consist of a base material layer (layer A) and a softer elastic material layer (layer B) called the ''foaming layer.'' In accordance with the results of previous work, the base material for the mandrel was selected as aluminum.[4] The single mode fiber under investigation in this paper is the same as that used in a simple cylindrical mandrel hydrophone.[10–12] Polyurethane was selected as the material for the molding layer also based on the results of the previous work mentioned above.

The foaming layer works as a real compliant layer that amplifies the deformation of the optical fiber when it is exposed to an external pressure, whereas the base layer, in addition to being a compliant layer, supports the overall hydrophone structure. Hence, the foaming layer normally should be more compliant than the base layer. The foaming layer for the proposed design can be fabricated from plastic foams that consist of a minimum of two phases, a solid polymer matrix, and a gaseous phase derived from a blowing agent. The solid phase can be inorganic, organic, or organometallic. Plastic foams can be produced with a great variety of mechanical properties and are usually classified as flexible, semiflexible (semirigid), or rigid. The physical properties of commercialized plastic foams applicable to the proposed hydrophone are represented in Table I.[13]

The proposed hydrophone, as shown in Fig. 1, is expected to provide a higher sensitivity than a simple cylindrical model. In addition, composite structures can provide more flexible design features, i.e., the structure, shape, and material properties of the mandrel, which in turn implies an improved controllability of the hydrophone performance. This paper outlines the design of an interferometric optical fiber hydrophone with a composite mandrel that has a fundamental resonance frequency over 15 kHz, and superior pressure sensitivity in underwater conditions over 200 m. The optical fiber is assumed to maintain a constant length of 150 m. Basically, the overall phase shift of light propagating through an optical fiber in response to external acoustic pressure is proportionate to the length of the fiber; thus, a longer fiber is desirable. However, a fiber that is too long can produce a constraining effect on the composite mandrel and create the possibility that thermally driven length fluctuations will limit the minimum detectable pressures. Accordingly, the optimum fiber length was set at 150 m. Table II lists the material properties of the constituent parts of the proposed hydrophone. Other design features were selected to increase sensitivity and were associated with the material properties and thickness t_f of the foaming layer. In addition, the geometry of the composite mandrel, including the inner diameter ID (diameter of the center hole), outer diameter OD, and length L_m of the whole mandrel were determined in relation to the effect of the foaming layer. The full three-dimensional behavior of the hydrophone was then analyzed in relation to variations in the design features using the finite element method (FEM), and finally the pressure sensitivity of the hydrophone in terms of its parameters was considered.

TABLE I. Physical properties of the commercial polymer foams.

Types of polymer foams	Density (kg/m^3)	Young's modulus (GPa)
Polystyrene foams	16–850	0.01–5.20
Polyorefin foams	10–965	0.48–5.00
PVC foams	550–1400	1.30–3.10
Epoxy foams	140–400	0.70–0.33
Fluoropolymer foams	889–1954	0.15–0.35
Syntactic polymer foams	200–750	0.15–3.00

I. FINITE ELEMENT ANALYSIS

To analyze the performance of an interferometric optical fiber hydrophone with a composite mandrel in an underwater environment, a finite element model was constructed and analyzed in the manner described in the previous paper.[4] First, a modal analysis of the concentric composite mandrel was completed to maintain the fundamental resonance frequency of the hydrophone above 15 kHz. Next, a static

TABLE II. Material properties of the constituent parts of an optical fiber hydrophone with a concentric composite mandrel.

	Material	Young's modulus, E (GPa)	Density (kg/m^3)	Poisson's ratio
Base layer	Al	73.0	2700	0.33
Foaming layer	To be determined			
Fiber	Si-glass	19.6	1404	0.34
Molding	Polyurethane (LD)	1.3	960	0.45

analysis of the hydrophone was performed using an initial hydrostatic pressure of 2×10^6 Pa to actualize underwater conditions of over 200 m. Finally, a harmonic analysis was conducted using the deformed shape of the hydrophone from the static analysis to analyze the sensitivity of the hydrophone at a predetermined frequency of 5 kHz. The result indicated the optimal structure for the composite mandrel and the corresponding highest sensitivity. The modeling and analysis was carried out using the commercial finite element software ANSYS, which utilizes a substructuring modeling technique for achieving acceptable accuracy with the minimum of computer processing time.

The analysis theory for measuring the sensitivity of an optical fiber hydrophone is the same as in Ref. 4. However, sensitivity is expressed in decibels (dB). As derived in Ref. 4, the magnitude S_M of the sensitivity is written as

$$S_M=\frac{\Delta\phi}{P_{\text{applied}}}=\phi^*\left(\sum \varepsilon_r+\sum \varepsilon_\theta-\frac{n^2}{2}\left[p_{11}^*\sum \varepsilon_r+p_{12}^*\left(\sum \varepsilon_r+\sum \varepsilon_\theta\right)\right]\right), \quad (1)$$

where ϕ is the phase of light propagating through an optical fiber, P is the external acoustic pressure, ε_i are the principal strains of the fiber, n is the refraction index of the fiber core, and p_{ij} are elements of the strain optic tensor. Accordingly, the acoustic sensitivity S of a hydrophone relative to a reference value is written as

$$S(\text{dB})=20\log\frac{S_M}{S_R}, \quad (2)$$

where S_R is the reference sensitivity of 1 rad/μPa.

II. RESULTS AND DISCUSSION

A. Fundamental resonance frequency of the composite mandrel (or geometrical range)

In this study, the mandrel hydrophone was required to have a fundamental resonance frequency over 15 kHz, and the optical fiber was required to maintain a constant length of 150 m. Before conducting a detailed analysis of the hydrophone response, a preliminary modal analysis was performed so that the range of the main calculations could be established beforehand, even through a rough estimate. The maximum allowable dimension of the mandrel was determined when its fundamental resonance frequency reached 15 kHz. However, the optical fiber cannot be wound as a single layer onto the mandrel surface due to its extended length; therefore, it is stacked as multiple layers. Since the fiber has a higher stiffness than the foaming or molding layers, a higher number of fiber layers around the mandrel strengthens the stiffening effect from the optical fiber, which in turn reduces the sensitivity of the hydrophone. Consequently, a lower number of fiber layers is more desirable. Figure 2 shows the

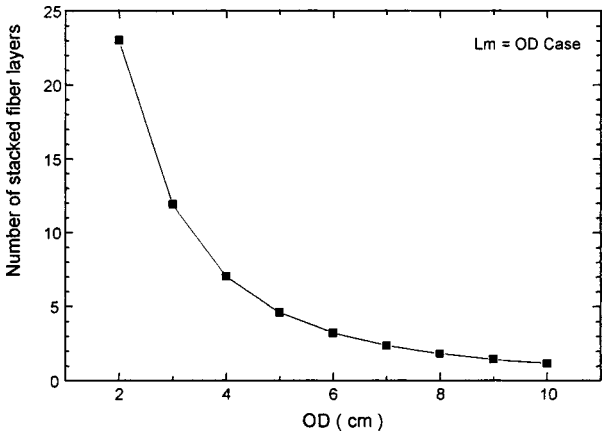


FIG. 2. Variation of the number of optical fiber layers stacked on the surface of the concentric composite mandrel in relation to the OD of the mandrel.

variation of the number of optical fiber layers stacked on the surface of the concentric composite mandrel in relation to the OD of the mandrel.

Figure 3 shows the variation of the fundamental mode frequency of a composite mandrel in relation to its OD, and the ratio of the ID/OD in relation to the representative thickness and material properties of the foaming layer. For convenience, the L_m was assumed to be the same as the OD of the composite mandrel, and the thickness of the foaming layer was set at 1 mm. Most polymers for acoustical applications have a Young's modulus between 0.1 and 6 GPa, and a Poisson's ratio between 0.35 and 0.45.[14–16] The foaming layer selected exhibited softer material properties, that is, a Young's modulus of 0.1 GPa, a Poisson's ratio of 0.45, and a density of 1000 kg/m^3. The results of the modal analysis indicated that in the case of the Al composite mandrel with an L_m and an OD of over 9 cm, and an ID of more than 65 percent of the OD, the fundamental mode frequency was lower than 15 kHz. Therefore, the Al composite mandrel was expected to have an L_m or an OD of less than 9 cm, and an ID of less than 65 percent of the OD. The effect of the thickness of the foaming layer on the resonance frequency of the mandrel was also checked within this range. The maximum allowable thickness of the foaming layer, facilitating a fundamental mode frequency over 15 kHz with an L_m, OD,

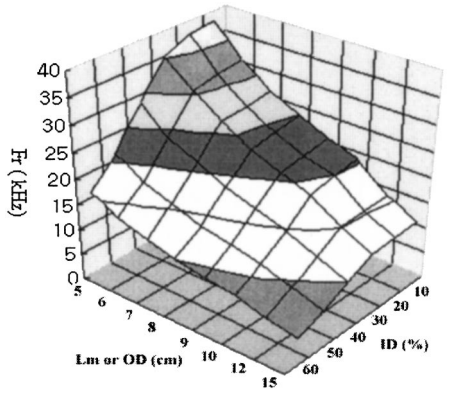


FIG. 3. Variation of the fundamental resonance frequency of the Al concentric composite mandrel with its OD and the ID/OD ratio, where the L_m has been set equal to the OD.

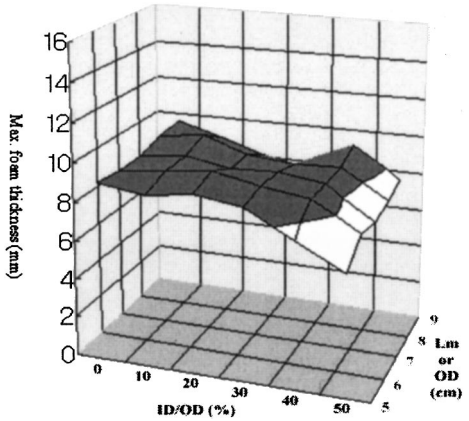


FIG. 4. Maximum allowable thickness of the foaming layer in the Al concentric composite mandrel with a fundamental resonance frequency higher than or equal to 15 kHz.

and mandrel ID as illustrated in Fig. 3, ranged from 8 to 10 mm (see Fig. 4). With a thickness larger than the values in Fig. 4, the fundamental resonance frequency fell below 15 kHz. These results are summarized in Table III.

B. Influence of material properties of foaming layer on hydrophone sensitivity

For the foaming layer, the material parameters that have an effect on the acoustic sensitivity of the hydrophone are the Young's modulus and Poisson's ratio. The Young's modulus of the composite mandrel's base material, Al was about 73 GPa. In a composite mandrel, the foaming layer should be more flexible than the base material in order to improve sensitivity. Therefore, with reference to Table I and the results in Ref. 4, the Young's modulus of the foaming layer had to be within a range from 0.1 to 10 GPa, and its Poisson's ratio was selected within a range from 0.35 to 0.45 in relation to the materials in Table I. The variation of acoustic sensitivity in relation to the material properties of the foaming layer is shown in Figs. 5 and 6. The mandrel had an L_m and OD of 7 cm, and an ID of 4 cm, while the foaming layer had a thickness of 10 mm. In Fig. 5, the acoustic sensitivity decreased as the Young's modulus, E, of the foaming layer increased. Conversely, there was little change in sensitivity when the Poisson's ratio was varied (see Fig. 6). Therefore, for the ranges selected in this study, it can be concluded that for a hydrophone with the highest sensitivity the foaming layer of the composite mandrel must have the lowest possible Young's modulus, regardless of its Poisson's ratio.

TABLE III. Variation range of the geometrical parameters for the Al concentric composite mandrel.

Geometrical parameters	Variation range (cm)
Length L_m	5.0–9.0
Outer diameter OD	5.0–9.0
Inner diameter ID	0–6.0
Thickness of the foaming layer t_f	0–1.0

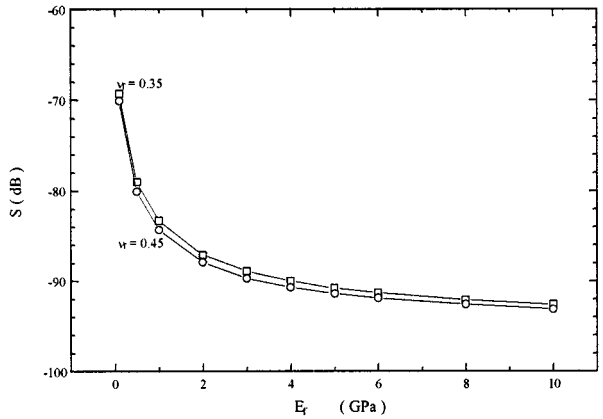


FIG. 5. Sensitivity versus Young's modulus (E) of the foaming layer of the optical fiber hydrophone (ID=4 cm; L_m=OD=7 cm; t_f=1 cm).

C. Influence of mandrel geometry on hydrophone sensitivity

The geometrical parameters of a composite mandrel that affect the sensitivity of a hydrophone are the foaming layer thickness t_f and the L_m, ID, and OD of the mandrel. The effects of the geometrical parameters are described in Figs. 7–9. In all the figures, the material properties of the foaming layer were kept constant, i.e., a Young's modulus of 0.1 GPa and a Poisson's ratio of 0.4. In consideration of the results in Sec. III A for maintaining a fundamental resonance frequency above 15 kHz, the initial L_m and OD were set at 7 cm.

For an Al composite mandrel with no center hole, Fig. 7 shows that the sensitivity increased from −98 to −84 dB as the thickness of the foaming layer increased from 0 to 10 mm. Since the OD was fixed at 7 cm, an increase in the foaming layer thickness resulted in a decrease in the Al layer thickness. The sensitivity of the mandrel with a constant foaming layer thickness also increased with an increase of the ID. However, an increase in sensitivity was not observable until the ID reached about 50 percent of the OD, whereas it was readily observable with even a minor increase in the foaming layer thickness. Furthermore, the sensitivity of a hydrophone with an Al mandrel that had no foaming layer and yet had an ID equal to 85 percent (the maximum

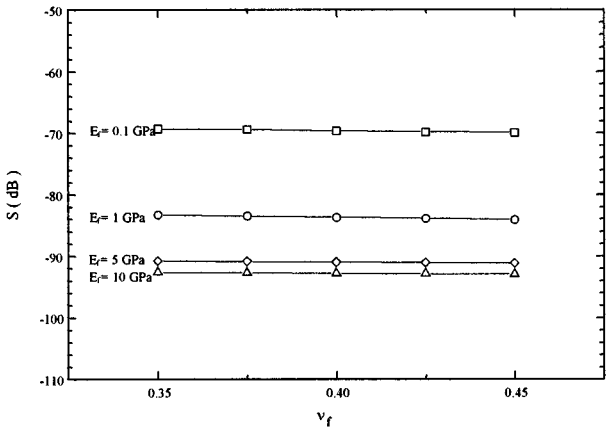


FIG. 6. Sensitivity versus Poisson's ratio of the foaming layer of the optical fiber hydrophone (ID=4 cm; L_m=OD=7 cm; t_f=1 cm).

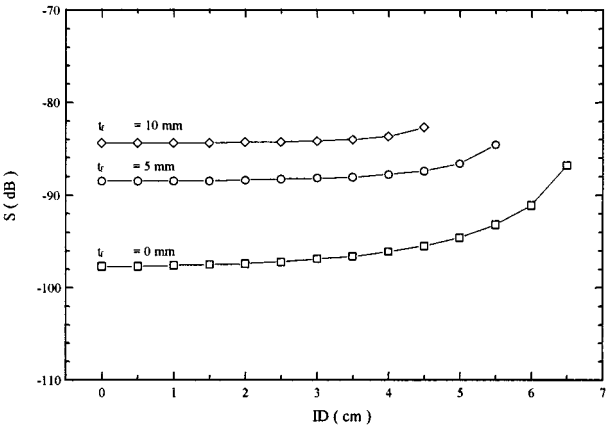


FIG. 7. Sensitivity versus inner diameter (ID) of the optical fiber hydrophone(L_m=OD=7 cm; E_f=1 GPa; ν_f=0.4).

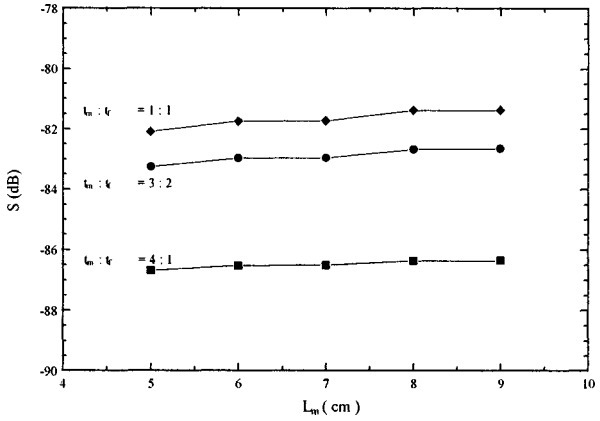


FIG. 9. Sensitivity versus length of the optical fiber hydrophone (ID/OD =0.4; OD=5 cm; E_f=1 GPa, ν_f=0.4).

allowable ID/OD ratio for an L_m of 7 cm) of the OD, i.e., an ID of 6 cm, was only 7 dB higher than that of a hydrophone without a center hole (ID=0). Conversely, when the thickness of the foaming layer increased from 0 to 10 mm (the maximum allowable thickness), the sensitivity increased about 14 dB. Consequently, with a constant L_m and OD, the foaming layer thickness is a more efficient control parameter for the improvement of sensitivity than the ID of the composite mandrel. The variation in sensitivity in relation to the OD and the ID/OD ratio of the mandrel is shown in Fig. 8. In this figure, the ratio of the thickness of the Al base layer to that of the foaming layer (t_m/t_f) was kept constant, i.e., 1:1. With a constant t_m/t_f ratio, an increase in both the OD and the ID/OD ratio resulted in a decrease in sensitivity. With a constant t_m/t_f ratio and ID/OD ratio, a larger OD meant a larger thickness of both the Al base layer and the foaming layer, and thus a smaller stacking number of optical fiber layers on the mandrel surface. The fundamental requirement for a mandrel is that it should be as compliant as possible to amplify the deformation of the optical fiber exposed to external pressure. In consideration of the results in Fig. 7, an increase in the thickness and hence volume of the foaming layer may contribute to a certain improvement of sensitivity. However, Al has a much higher stiffness than the optical fiber. Therefore, the stiffening effect of an Al layer was more prominent, and, consequently, the overall acoustic sensitivity of the hydrophone deteriorated due to the higher stiffness of the composite mandrel. Furthermore, the increased coupling between the radial and longitudinal mode vibrations of a mandrel with a larger OD also contributed to a decrease of sensitivity. These results are consistent with those of previous work.[4] The sensitivity also decreased as the ID of the composite mandrel increased (see Fig. 8). This was due to the decrease in the thickness of the foaming layer with a constant ratio between the OD and the t_m/t_f. On the other hand, the sensitivity slightly improved with an increase of the mandrel length L_m, Fig. 9. The length of the optical fiber was fixed at 150 m, therefore, a longer mandrel for a specific OD resulted in a smaller stacking number of fiber layers on the mandrel surface. This created a more flexible mandrel because the optical fiber was stiffer than the overall composite mandrel, thereby producing a higher sensitivity. In addition, Fig. 9 also confirms the argument that a smaller t_m/t_f ratio improves sensitivity. In conclusion, the highest sensitivity was achieved when an Al composite mandrel had the following geometrical parameters: (1) smaller t_m/t_f ratio, (2) smaller OD, (3) smaller ID/OD ratio, and (4) larger L_m as long as its fundamental resonance frequency remained above 15 kHz. Therefore, the optimal geometry of an Al composite mandrel within the variation range in Table III was determined as having a t_f of 10 mm, OD of 5 cm, ID of 1 cm equal to 20 percent of the OD, and L_m of 9 cm. The optimal Al composite mandrel had a fundamental resonance frequency of 15.5 kHz and a sensitivity of -82 dB in reference to 1 rad/μPa. From Figs. 7 and 8, it can be seen that a simple (without the foaming layer) concentric Al mandrel of the same dimensions had a sensitivity of about -96 dB. Therefore, the addition of a foaming layer with the optimum material properties and thickness improved the sensitivity by 14 dB.

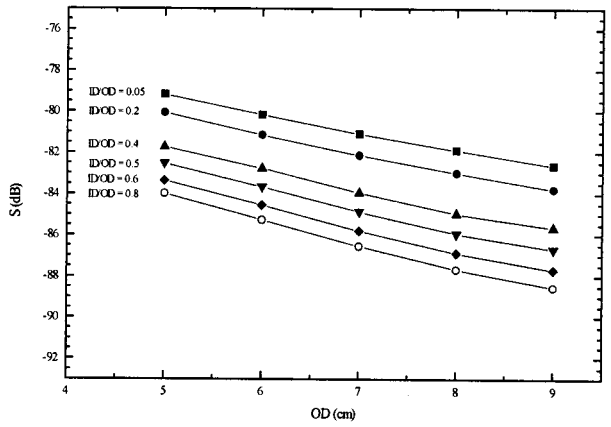


FIG. 8. Sensitivity outer diameter (OD) of the optical fiber hydrophone (L_m=7 cm; E_f=1 GPa; ν_f=0.4; t_m/t_f=1/1).

III. CONCLUSION

This paper outlined the FEM design of an interferometric optical fiber hydrophone with a concentric composite mandrel. The concentric composite mandrel consists of a compliant foaming layer and a base material layer of Al. The

hydrophone was required to have a fundamental resonance frequency over 15 kHz and superior pressure sensitivity in underwater conditions over 200 m. The design parameters investigated included the material properties of the foaming layer and the geometry of the overall composite mandrel. According to the results for the ranges considered in this paper, an improved sensitivity was achieved with an Al mandrel with a thicker foaming layer that had the lowest possible Young's modulus regardless of its Poisson's ratio. With respect to the geometry of a concentric composite mandrel, a higher sensitivity was obtained with a smaller outer diameter OD, a smaller ID/OD ratio, and a longer length L_m as long as the fundamental resonance frequency of the mandrel remained above 15 kHz. The resulting sensitivity of a hydrophone with an optimal structure Al composite mandrel was -82 dB with respect to 1 rad/μPa. This result is a 14-dB increase in sensitivity over a simple concentric Al mandrel type hydrophone of the same dimensions, which verifies the effectiveness of the foaming layer. Therefore, the concentric composite mandrel-type optical fiber hydrophone studied in this paper provides both superior sensitivity and better structural robustness.

[1] J. A. Bucaro, N. Lagakos, J. H. Cole, and T. G. Giallorenzi, ''Fiber optic acoustic transduction,'' Phys. Acoust. **16**, 385–457 (1982).

[2] J. Jarzynski, R. Hughes, T. R. Hickman, and J. A. Bucaro, ''Frequency response of interferometric fiber optic coil hydrophone,'' J. Acoust. Soc. Am. **69**, 1799–1808 (1981).

[3] S. Africk, T. Burton, P. Jameson, and A. Ordubadi, ''Design studies for fiber optic hydrophones,'' Report No. 4658, Bolt, Beranek & Newman, Inc., Cambridge, MA (1981).

[4] J. Im and Y. Roh, ''A finite element analysis of an interferometric optical fiber hydrophone,'' J. Acoust. Soc. Am. **103**, 2425–2431 (1998).

[5] P. J. Nash and J. Keen, ''Design and construction of practical optical fiber hydrophone,'' Proc. Inst. Acoust. **12**, 201–212 (1990).

[6] G. F. Mcdearmon, ''Theoretical analysis of a push–pull fiber-optic hydrophone,'' J. Lightwave Technol. **LT-5**, 647–652 (1987).

[7] M. Vengsarkar, K. A. Murphy, T. A. Tran, and R. O. Claus, ''Novel fiber optic hydrophone for ultrasonic measurements,'' Ultrasonic Symposium, pp. 603–606 (1988).

[8] A. Dandridge, A. B. Tveten, A. M. Yurek, and Y. Y. Chao, ''A fiber optic hydrophone for high frequency applications,'' in *Transducers for Sonics and Ultrasonics*, edited by M. D. McCollum, B. F. Hamonic, and O. B. Wilson (Technomic, Orlando, 1992), pp. 292–297.

[9] G. F. McDearmon, ''Theoretical analysis of a push–pull fiber-optic hydrophone,'' J. Lightwave Technol. **LT-5**, 647–652 (1987).

[10] V. S. Sudarshanam and K. Srinivasan, ''Static phase change in a fiber optic coil hydrophone,'' Appl. Opt. **29**, 855–863 (1990).

[11] J. A. Bucaro and T. R. Hickman, ''Measurement of sensitivity of optical fibers for acoustic detection,'' Appl. Opt. **18**, 938–940 (1979).

[12] N. Lagakos, J. A. Bucaro, and R. Hughes, ''Acoustic sensitivity predictions of single mode optical fibers using Brillouin scattering,'' Appl. Opt. **19**, 3666–3670 (1980).

[13] A. H. Landrock, *Handbook of Plastic Foams: Types, Properties, Manufacture and Applications* (Noyes, Park Ridge, NJ, 1995).

[14] M. G. Grewe, T. R. Gururaja, T. S. Shrout, and R. E. Newnham, ''Acoustic properties of particle/polymer composites for ultrasonic transducer backing applications,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **37**(6), 506–513 (1990).

[15] I. K. Jacqueline, *Encyclopedia of Polymer Science and Engineering*, Vol. 1 (Wiley, New York, 1985), pp. 131–160.

[16] P. J. Hoftyzer, *Properties of Polymer, Their Estimation and Correlation with Chemical Structure*, 3rd ed. (Elsevier Science, New York, 1990), pp. 439–454.

Analysis of lateral vibration behavior of railway track at high frequencies using a continuously supported multiple beam model

T. X. Wu[a)]
Department of Mechanical Engineering, Shanghai Tiedao University, Shanghai 200331, China

D. J. Thompson[b)]
Institute of Sound and Vibration Research, University of Southampton, Southampton SO17 1BJ, United Kingdom

(Received 14 September 1998; accepted for publication 19 May 1999)

The lateral vibration of railway track at high frequencies is a more difficult research topic, compared with the vertical vibration, because the complex cross-sectional deformation of the rail has to be taken into account. The commonly used models for the lateral vibration are based on the FEM/FSM or a derivation of the FEM. It is inconvenient to calculate responses of the rail to external excitation by these models, especially when the discrete supports of the rail are taken into account. To overcome this barrier a new model for studying the lateral vibration behavior of railway track is developed in this paper. It is a simple model consisting of two infinite Timoshenko beams connected by an array of finite beams. Detailed investigation of the lateral vibration for a free rail and a continuously supported rail has been carried out efficiently using this model. The free wave dispersion relation calculated using this model has been shown to be in good agreement with a FE model. Quite good agreement between the predictions and the measurement data has also been reached in terms of accelerances. © *1999 Acoustical Society of America.*
[S0001-4966(99)00409-9]

PACS numbers: 43.40.At [PJR]

INTRODUCTION

The dynamic behavior of railway track in the frequency region 50–5000 Hz is of great importance in relation to the rolling noise generated by the interaction of the wheel and rail. Compared with the vertical vibration, it is more difficult to study the lateral vibration behavior of railway track at higher frequencies because the cross-sectional deformation of the rail has to be taken into account. The most important cross-sectional deformation includes bending and torsion of the rail head and foot and relative motion between the head and foot, for example, bending of the web. Conventional beam models (from Refs. 1 and 2) are not appropriate to study the lateral vibration of railway track because they cannot represent such cross-sectional deformation. Therefore, more comprehensive models that allow the rail cross-sectional deformation are needed. These models are usually from the finite element method (FEM) and thus need to be treated numerically.

Among these models there are three typical ways to deal with an infinite rail. The first method is to consider a piece of rail up to about 1 m long and mesh this length of rail with different elements (beam and plate elements or solid elements), applying symmetric and/or antisymmetric boundary conditions at both ends in order to simulate the waves in an infinite rail. Thompson[3] obtained the dispersion relation of the waves in an infinite rail using this method. The second method is also to deal with a length of meshed rail based on periodic structure theory (PST, Ref. 4) which allows an infinite structure of constant cross section, such as the rail, to be considered as a periodic structure of arbitrary period. Knothe *et al.*[5] obtained the wave-number–frequency relation of a free rail through this method, with different element meshing models. Thompson[3] predicted wave numbers and receptances in a continuously supported rail using a PST model. The third method is by meshing only the cross-sectional area of the rail instead of meshing of the whole rail in three dimensions, while along the rail the wave propagation factor $e^{i(\omega - kz)}$ is introduced. The node displacements on the cross-sectional area can be either two-dimensional (2-D) (only displacements in plane) or 3-D (including cross-sectional warping and longitudinal displacement). Gavric[6] and Gry[7] calculated the frequency–wave-number relation and cross-sectional deformation wave shapes (including cross-sectional warping) using this third method. Furthermore, Gry introduced a span wave coordinates method to calculate the point accelerance of the discretely supported rail. There is also a model from the finite strip method (FSM) which was developed by Strzyzakowski and Ziemanski.[8] In another approach a simple layered plane model was used by Scholl[9,10] to investigate wave propagation in a free rail. Although this latter model is analytical, it has limited applicability to the lateral vibration of rail.

Most of these models cited above are based on the FEM/FSM or a derivation of the FEM (the third method). They are complicated models and thus numerical treatment is essential. For identifying the wave-number–frequency relation in a free rail these models are straightforward. However, it is

[a)]Currently on leave at ISVR, University of Southampton, UK.
[b)]Corresponding author. Electronic mail: djt@isvr.soton.ac.uk

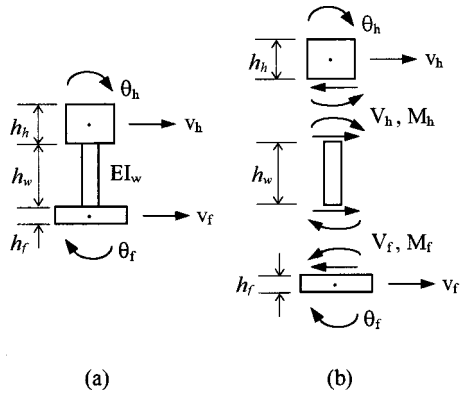


FIG. 1. Multiple beam model for lateral vibration of rail.

very inconvenient to calculate responses of the rail to external excitation by these models, especially when the discrete supports of the rail are taken into account, because a large number of degrees of freedom on the boundaries have to be dealt with. This may be the main barrier to studying the lateral vibration behavior of the supported rail.

To overcome this barrier a new model for studying the lateral vibration behavior of railway track is developed in this paper. This model consists of two infinite Timoshenko beams, which represent the rail head and foot, and their connections—an array of many finite beams, which represent the web of rail. Although simple, this model includes all the essential cross-sectional deformation up to at least 5 kHz such as bending and torsion of the rail head and foot and web bending. Although the twisting stiffness of the web in the axis of the rail is omitted in this model, its effects may be compensated by adjusting some parameters of the model. The predictions from this model are shown to have a quite good agreement with measurements in the frequency region 50–5000 Hz. Since it is simple, some more detailed investigation of the lateral vibration properties of railway track can be achieved using this model.

I. MODELING OF THE RAIL

To investigate the lateral vibration properties of a rail at high frequencies the cross-sectional deformation should be taken into account. The essential cross-sectional deformation types are rail head bending and torsion, rail foot bending and torsion, and the relative motion between the rail head and foot, web bending. A model for studying the lateral vibration of a rail should allow all these kinds of deformation.

A. Multiple-beam model

The cross section of the multiple-beam model is shown in Fig. 1(a). In this model the whole rail is divided into three parts: the head and the foot are represented by two infinite Timoshenko beams which can be subjected to both bending and torsion and the web is replaced by numerous beams along the rail which connect the head and foot. In this simplified model the twisting stiffness and the bending stiffness of the web in the direction of the rail axis vanish. The effects of the web twisting stiffness can be compensated by adjusting the head and foot torsional stiffnesses. Ignoring the bending stiffness of the web in the rail axis direction will cause no problem because it is much smaller than the lateral bending stiffnesses of the rail head and foot. As this model consists of only beams, it is analytical rather than numerical and thus the calculation is relatively simple and detailed study for the lateral vibration behavior becomes possible.

B. Equations of motion for free rail

The material properties of the model are represented by E, the Young's modulus, G, the shear modulus, and ρ, the density. The geometric properties of each cross section are characterized by A, the cross-sectional area, I, the area moment of inertia, J, the torsional moment of inertia, I_p, the polar moment of inertia and κ, the shear coefficient of the rectangle. The subscripts h, f, and w are applied to represent the rail head, foot, and web, respectively. The isolated diagrams for the rail head, foot, and web are shown in Fig. 1(b), where ν_h and θ_h represent the head lateral displacement and rotation about the rail axis respectively, ν_f and θ_f are the foot lateral displacement and rotation about the rail axis, and their positive direction is pointed out in Fig. 1. In Fig. 1(b) V_h and M_h represent the force and moment per unit length between the head and web and V_f and M_f represent the force and moment per unit length between the web and foot. They appear in pairs.

Applying Newton's second law of motion to the rail head gives

$$\rho A_h \ddot{\nu}_h + GA_h\kappa_h(\psi_h' - \nu_h'') = -V_h, \tag{1a}$$

$$\rho I_h \ddot{\psi}_h + GA_h\kappa_h(\psi_h - \nu_h') - EI_h\psi_h'' = 0, \tag{1b}$$

$$\rho I_{hp}\ddot{\theta}_h - GJ_h\theta_h'' = -M_h + V_h\frac{h_h}{2}. \tag{1c}$$

Similarly for the foot,

$$\rho A_f \ddot{\nu}_f + GA_f\kappa_f(\psi_f' - \nu_f'') = -V_f, \tag{2a}$$

$$\rho I_f \ddot{\psi}_f + GA_f\kappa_f(\psi_f - \nu_f') - EI_f\psi_f'' = 0, \tag{2b}$$

$$\rho I_{fp}\ddot{\theta}_f - GJ_f\theta_f'' = -M_f - V_f\frac{h_f}{2}, \tag{2c}$$

and for the web

$$\mathbf{M}_w\begin{pmatrix} \ddot{\nu}_h - \dfrac{h_h\ddot{\theta}_h}{2} \\ 0 \\ \ddot{\theta}_h \\ \ddot{\nu}_f + \dfrac{h_f\ddot{\theta}_f}{2} \\ 0 \\ \ddot{\theta}_f \end{pmatrix} + \mathbf{K}_w\begin{pmatrix} \nu_h - \dfrac{h_h\theta_h}{2} \\ 0 \\ \theta_h \\ \nu_f + \dfrac{h_f\theta_f}{2} \\ 0 \\ \theta_f \end{pmatrix} = \begin{pmatrix} V_h \\ 0 \\ M_h \\ V_f \\ 0 \\ M_f \end{pmatrix}, \tag{3}$$

where ψ_h and ψ_f represent the lateral slopes of the rail head and foot, respectively, and $'$ indicates the derivative with respect to z, the coordinate along the axis of the rail. Here $\mathbf{M}_w$ and $\mathbf{K}_w$ are the mass and stiffness matrices of the web, respectively (beam, per unit length of rail).

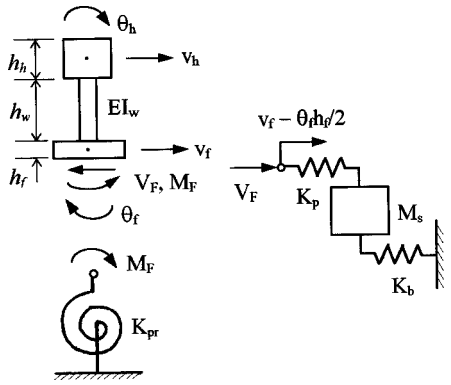


FIG. 2. Isolated diagram of continuously supported rail.

Equations (1)–(3) can be combined to give a compact formula as follows:

$$\mathbf{M}\ddot{\mathbf{q}}-\mathbf{D}\mathbf{q}''-\mathbf{G}\mathbf{q}'+\mathbf{K}_R\mathbf{q}=\mathbf{0}, \tag{4}$$

where **M**, **D**, **G**, and $\mathbf{K}_R$ are constant coefficient matrices related to the rail cross-sectional parameters, and

$$\mathbf{q}=(\nu_h \quad \psi_h \quad \theta_h \quad \nu_f \quad \psi_f \quad \theta_f)^{\mathrm{T}}. \tag{5}$$

C. Equations of motion for a continuously supported rail

The isolated diagram of a continuously supported rail is shown in Fig. 2, where V_F and M_F are the reaction force and moment per unit length from the railway track foundation, respectively. If Z_t and Z_r represent the translational and rotational dynamic stiffnesses per unit length of the foundation, respectively, the equation of motion for the continuously supported rail can be given as follows:

$$\mathbf{M}\ddot{\mathbf{q}}-\mathbf{D}\mathbf{q}''-\mathbf{G}\mathbf{q}'+\mathbf{K}_R\mathbf{q} = -\left[0\ 0\ 0\ Z_t\left(\nu_f-\frac{h_f\theta_f}{2}\right)\ 0\ Z_r\theta_f-Z_t\left(\nu_f-\frac{h_f\theta_f}{2}\right)\frac{h_f}{2}\right]^T. \tag{6}$$

Here Z_t and Z_r are given by

$$Z_t=\frac{K_p(K_b-M_s\omega^2)}{K_p+K_b-M_s\omega^2}, \tag{7a}$$

$$Z_r=K_{pr}, \tag{7b}$$

where K_p and K_b are the lateral stiffnesses per unit length of the pad and ballast respectively, K_{pr} is the rotational stiffness per unit length of the pad, and M_s is the sleeper mass per unit length. In Eq. (7b) the rotational intertia of the sleeper and the rotational stiffness of the ballast are not taken into account because they are much larger than the rotational inertia of the rail cross section and the rotational stiffness of the pad, respectively. Equation (6) can be simplified to

$$\mathbf{M}\ddot{\mathbf{q}}-\mathbf{D}\mathbf{q}''-\mathbf{G}\mathbf{q}'+\mathbf{K}\mathbf{q}=\mathbf{0}, \tag{8}$$

where

$$\mathbf{K}=\mathbf{K}_R+\mathbf{K}_F \tag{9}$$

and

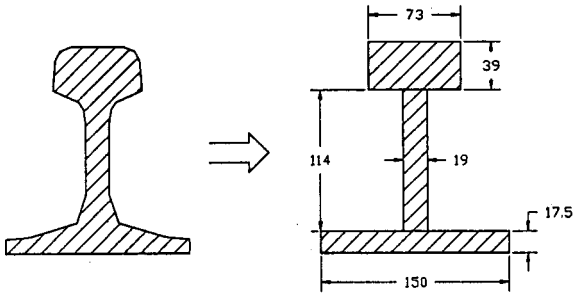


FIG. 3. Simplified cross-section for UIC 60 rail (from Ref. 9).

$$\mathbf{K}_F=\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & Z_t & 0 & -\dfrac{Z_t h_f}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\dfrac{Z_t h_f}{2} & 0 & Z_r+\dfrac{Z_t h_f^2}{4} \end{bmatrix}. \tag{10}$$

D. Cross-sectional parameter determination

In this paper UIC 60 rail is chosen as an example. The cross-sectional data (from Ref. 9) of a simplified version of UIC 60 rail are shown in Fig. 3. In the multiple-beam model developed in this section there are two infinite beams representing the rail head and foot and an array of finite beams representing the web. The characteristic data for these beams can be calculated from the dimensions given in Fig. 3. Since the twisting stiffness of the web vanishes in the model, the torsional stiffness of the whole cross section of the rail has become softer. To overcome this drawback, the torsional moment of inertia of the web cross section is added to the rail head and foot. One-third of the torsional moment of inertia of the web is added to the rail head, whereas two-thirds are added to the rail foot. This is because the rail head is already much stiffer than the rail foot in terms of torsional stiffness. Thus the total torsional stiffness of the rail cross section remains unchanged. In addition, I_w, the web area moment of inertia per unit length about the rail axis, is increased by 10% in order to offset the difference between the beam model and a plate model for the web. This results in an increase in the cut-on frequencies of web bending. By these adjustments the results from the simpler multiple beam model will be much closer to those from a FE model.

The following parameters are used in the multiple beam model:

$$E=2.1\times10^{11}\ \mathrm{N/m^2}, \quad G=0.77\times10^{11}\ \mathrm{N/m^2},$$
$$\rho=7850\ \mathrm{kg/m^3},$$
$$A_h=2.847\times10^{-3}\ \mathrm{m^2}, \quad \kappa_h=0.85,$$
$$I_h=1.264\times10^{-6}\ \mathrm{m^4}, \quad I_{hp}=1.625\times10^{-6}\ \mathrm{m^4},$$
$$J_h=0.9549\times10^{-6}\ \mathrm{m^4},$$

$$A_f = 2.625\times10^{-3}\,\mathrm{m}^2, \quad \kappa_f = 0.85,$$
$$I_f = 4.921\times10^{-6}\,\mathrm{m}^4, \quad I_{fp} = 4.988\times10^{-6}\,\mathrm{m}^4,$$
$$J_f = 0.2471\times10^{-6}\,\mathrm{m}^4,$$
$$A_w = 2.166\times10^{-3}\,\mathrm{m}^2, \quad I_w = 0.5716\times10^{-6}\,\mathrm{m}^4,$$
$$J_w = 0.2338\times10^{-6}\,\mathrm{m}^4.$$

II. PROPAGATING WAVES IN THE RAIL

In this section the dispersion relation of the waves in the rail are explored. First, the waves in a free rail are calculated and the results from the multiple beam model are compared with those from a FE model. Then the waves in a continuously supported rail without damping are calculated. Last, the dispersion relation of propagating waves and their decay rate in a damped rail with a damped continuous foundation are studied.

A. Equation of dispersion relation

Assuming the displacement vector $\mathbf{q}$ has the form of

$$\mathbf{q} = \mathbf{q}(z)e^{i\omega t}, \tag{11}$$

where

$$\mathbf{q}(z) = [\,\nu_h(z) \quad \psi_h(z) \quad \theta_h(z) \quad \nu_f(z) \quad \psi_f(z) \quad \theta_f(z)\,]^T, \tag{12}$$

and substituting Eq. (11) into Eq. (4) and taking derivatives with respect to time only gives

$$\mathbf{D}\mathbf{q}''(z) + \mathbf{G}\mathbf{q}'(z) + (\omega^2\mathbf{M} - \mathbf{K}_R)\mathbf{q}(z) = \mathbf{0}. \tag{13}$$

Assuming $\mathbf{q}(z) = \mathbf{Q}e^{-ikz}$ and substituting it into Eq. (13) gives

$$[-k^2\mathbf{D} - ik\mathbf{G} + (\omega^2\mathbf{M} - \mathbf{K}_R)]\mathbf{Q} = \mathbf{0}. \tag{14}$$

This is a quadratic eigenvalue problem. The solutions for wave number k are usually complex and have the form of $k = \pm(a \pm ib)$, where a and b are real. They appear in opposite pairs when k is purely real or purely imaginary or in double pairs when k is complex. The imaginary part b represents the wave decay. When $b = 0$, the wave is propagating. The eigenvector $\mathbf{Q}$ represents the wave shapes and it is also complex. When $k = 0$, Eq. (14) becomes

$$(\omega^2\mathbf{M} - \mathbf{K}_R)\mathbf{Q} = \mathbf{0}. \tag{15}$$

Then the cut-on frequencies can be obtained from Eq. (15). At cut-on frequencies each cross section in the rail vibrates in phase over the whole length and no wave is produced.

To solve the quadratic eigenvalue problem, it is appropriate to convert Eq. (13) into the state space form. Taking

$$\mathbf{X} = \begin{bmatrix} \mathbf{q}(z) \\ \mathbf{q}'(z) \end{bmatrix} \tag{16}$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{D}^{-1}(\omega^2\mathbf{M} - \mathbf{K}_R) & -\mathbf{D}^{-1}\mathbf{G} \end{bmatrix}. \tag{17}$$

Equation (13) can be written in the state space form:

$$\mathbf{X}' = \mathbf{A}\mathbf{X}. \tag{18}$$

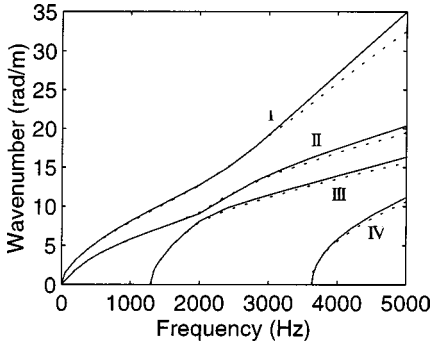


FIG. 4. Dispersion relation of waves in a free rail for lateral vibration: ——— from the multiple beam model and ··· from the FE model.

This leads to a normal eigenvalue problem

$$\det(\lambda\mathbf{I} - \mathbf{A}) = 0, \tag{19}$$

where $\lambda = -ik$.

B. Waves in a free rail

From Eq. (19) the dispersion relation of propagating waves in the rail can be determined. For comparison, a FE model representing the idealized UIC 60 rail of Fig. 3 is also used. It is a short length of rail (1 m) with symmetry or antisymmetry boundary conditions at the ends to simulate the waves in an infinite rail. The rail head and foot are each represented by 30 Timoshenko beam elements and the web is meshed with 6×30 plate elements. To consider the distances from the rail head axis to the web and from the rail foot axis to the web, two sets of 31 short stiff and massless beam elements are placed between the head and the web and between the foot and the web. These beams are used for transmitting force and moment only, and almost no elastic deformations occur on them.

The results from the analytical and FE models are shown in Fig. 4. As conventional,[3] the first (I) and the second waves (II) are called a lateral bending wave and a torsional wave, respectively, though at high frequencies these two waves have a compound deformed shape. The third wave (III) involves the web bending and torsion of both the head and the foot out of phase with each other. The fourth wave (IV) involves a double bending of the web, so that the head and foot are again each in torsion, but in phase with each other. For further features of these waves and wave shapes, see Refs. 3 and 5–7. The important thing to be emphasized here is that the results from the analytical model have a very good agreement with those from the FE model. The agreement in terms of the dispersion relation of waves in the rail is crucial, especially for those near the cut-on frequencies, because rail vibration can be regarded as the superposition of different waves in the rail.

C. Waves in a continuously supported rail

To calculate the frequency–wave-number relation in a supported rail, the dynamic stiffness of the foundation is employed and it is added to the system stiffness matrix, see Eqs. (9) and (10). The parameters of the railway track foundation

$$K_p = 83.3\,\mathrm{MN/m}^2, \quad K_{pr} = 1.09\,\mathrm{MN},$$

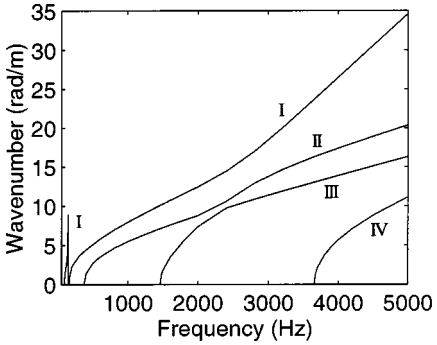


FIG. 5. Dispersion relation of waves in a continuously supported rail without damping for lateral vibration.

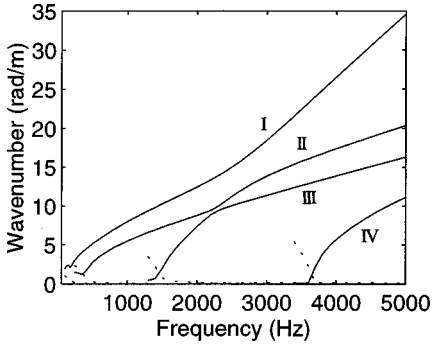


FIG. 6. Dispersion relation of waves in a continuously supported rail with damping for lateral vibration: ——— real part of the wave number and ··· imaginary part of the wave number.

$$K_b = 133.3\ \text{MN/m}^2, \quad M_s = 270\ \text{kg/m},$$

are derived from the values of track C from Ref. 11 and initially have no damping.

The results are shown in Fig. 5. Compared with the case of the free rail, the cut-on frequencies of waves I–III are increased because of the foundation dynamic stiffness, but for wave IV it is almost unchanged. In addition, the first wave, the lateral bending wave, has a branch at about 100–150 Hz. In this frequency region the rail vibration is coupled with the foundation mass. Apart from the regions close to cut-on frequencies, the dispersion relation curves in the supported rail are close to those for the free rail at higher frequencies. This implies that the railway track foundation mainly affects the low-frequency vibration of the rail, apart from the effects of damping.

If damping is introduced into both the rail and the foundation, the wave number k is always complex and has the form of $k = \pm(a - ib)$ which represents two waves propagating in two opposite directions with exponential decay. The decay rate Δ (in dB/m) of each wave is determined by the imaginary part of k and can be calculated using

$$\Delta = 20 \log_{10}(e^b) = 8.686\ \ b. \tag{20}$$

Damping is added through loss factors: $\eta_p = 0.25$ for the pad and $\eta_b = 0.6$ for the ballast. No damping is added to the rail.

Since all wave numbers have the form of $k = \pm(a - ib)$, it is important to distinguish the conventional propagating waves from the near-field waves. This can be achieved without difficulty according to their characteristics: the near-field waves always have a large imaginary part, whereas the propagating waves have a very small imaginary part above their cut-on frequencies.

The frequency–wave-number relation in the supported rail with damping is shown in Fig. 6. The curves of the real parts can be seen to be almost the same as those in the case without damping except near the cut-on frequencies. The imaginary parts can be seen to be higher before the cut-on frequencies and much lower after the cut-on frequencies; they decrease dramatically near their cut-on frequencies.

III. RECEPTANCE OF CONTINUOUSLY SUPPORTED RAIL

Although in practice a rail is discretely supported by pads and sleepers, a continuously supported rail model can avoid the complication of the periodic support nature and, to some extent, the results from the continuously supported rail model are still acceptable. In this section, the continuously supported model is employed for calculating receptances and thus the pad, sleeper, and ballast are replaced by equivalent continuous layers of mass and stiffness. The parameters of these layers used in this section are the same as above.

A. Equation of motion

In all subsequent work, it will be assumed that forces and responses are harmonic, so the $e^{i\omega t}$ term will usually be omitted. Considering the continuously supported rail subjected to a lateral harmonic excitation $Fe^{i\omega t}$ acting on the top of the rail head at the point $z = 0$, the equation of motion can be given as follows:

$$-\mathbf{D}\mathbf{q}'' - \mathbf{G}\mathbf{q}' - (\omega^2\mathbf{M} - \mathbf{K})\mathbf{q} = \delta(z)\left(F \quad 0 \quad F\frac{h_h}{2} \quad 0 \quad 0 \quad 0\right)^{\mathrm{T}}. \tag{21}$$

Equation (21) can be written in the state space form:

$$\mathbf{X}' = \mathbf{A}\mathbf{X} + \mathbf{P}\delta(z), \tag{22}$$

where $\mathbf{A}$ has the same form as in Eq. (17) but $\mathbf{K}_R$ is replaced by $\mathbf{K}$ which is the sum of rail stiffness matrix $\mathbf{K}_R$ and foundation dynamic stiffness matrix $\mathbf{K}_F$, and

$$\mathbf{P} = [\mathbf{0} \quad -\mathbf{D}^{-1}\mathbf{F}]^{\mathrm{T}}, \tag{23a}$$

$$\mathbf{F} = \left[F \quad 0 \quad \frac{Fh_h}{2} \quad 0 \quad 0 \quad 0\right]^{\mathrm{T}}. \tag{23b}$$

B. Point and cross receptances

Taking the Laplace transform of Eq. (22) gives

$$(s\mathbf{I} - \mathbf{A})\mathbf{X}(s) = \mathbf{P}. \tag{24}$$

The Laplace transforms of the rail head displacement and rotation, respectively, have the forms

$$V_h(s) = \frac{\Delta_1(s)}{\Delta(s)}, \tag{25}$$

$$\Theta_h(s) = \frac{\Delta_3(s)}{\Delta(s)}, \tag{26}$$

where $\Delta(s)$ is the determinant of matrix $s\mathbf{I}-\mathbf{A}$. $\Delta_1(s)$ and $\Delta_3(s)$ are the determinants of the matrices which are obtained from $s\mathbf{I}-\mathbf{A}$ by replacing its first and third columns by $\mathbf{P}$, respectively. The lateral responses of the rail head are found by performing the Laplace inverse transform using contour integration:

$$\nu_h(z)=\frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} V_h(s)e^{sz}\,ds$$

$$=\sum_{\substack{k \text{ with } \mathrm{Re}(s_k)<0 \text{ or} \\ \mathrm{Im}(s_k)<0 \text{ if } \mathrm{Re}(s_k)=0}} \mathrm{Res}[V_h(s_k)e^{s_k z}],\quad z>0, \tag{27}$$

$$\theta_h(z)=\frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} \Theta_h(s)e^{sz}\,ds$$

$$=\sum_{\substack{k \text{ with } \mathrm{Re}(s_k)<0 \text{ or} \\ \mathrm{Im}(s_k)<0 \text{ if } \mathrm{Re}(s_k)=0}} \mathrm{Res}[\Theta_h(s_k)e^{s_k z}],\quad z>0, \tag{28}$$

where the residues at the poles s_k are given by

$$\mathrm{Res}[V_h(s_k)e^{s_k z}]=\frac{\Delta_1(s_k)}{\Delta'(s_k)}e^{s_k z}, \tag{29}$$

$$\mathrm{Res}[\Theta_h(s_k)e^{s_k z}]=\frac{\Delta_3(s_k)}{\Delta'(s_k)}e^{s_k z}. \tag{30}$$

In a similar way the lateral responses of the rail foot can be obtained. The point receptance at the top of the rail head at $z=0$ is given by

$$\alpha(\omega)=\frac{\nu_h(0)+h_h\theta_h(0)/2}{F}. \tag{31}$$

C. Numerical results

The amplitude and phase of the lateral point receptance are shown in Fig. 7. Three well-damped resonances can be seen at about 100, 160, and 380 Hz. At 100 Hz the whole track bounces on the lateral stiffness of the ballast, whereas at 160 Hz the rail vibrates laterally on the stiffness of the pad. These two resonances occur at the cut-on frequencies of the lateral bending wave which can be seen in Fig. 5. At 380 Hz the rail rotates on the pad rotational stiffness. This can be identified as the cut-on of the torsional wave. The other two peaks at about 1400 and 3600 Hz can be identified as the cut-on of the web bending waves III and IV, respectively. These are damped by the pad loss factor (rotational), although the second one is much sharper.

Various ''cross'' receptances are shown in Fig. 8. (The term cross receptance is used here to mean the response at some other positions at $z=0$ to a force at the top of the head. The term transfer receptance is reserved for the response at $z\neq 0$.) Figure 8(a) and (b) shows the lateral displacement (in the center of the head or foot) and rotation, respectively. The solid lines are for the head and the dashed lines for the foot. It can be observed that the responses of the rail head usually are higher than those of the foot. At low frequencies the lateral displacement and rotation curves have similar tendencies, whereas at the web bending cut-on frequencies the peaks are more noticeable in the rotation than in the displacement. This is because the web bending strongly involves the rotation of the head and foot. The rail head displacement curve can be seen to be quite smooth at high frequencies and there is almost no peak appearing at the second web bending cut-on frequency.

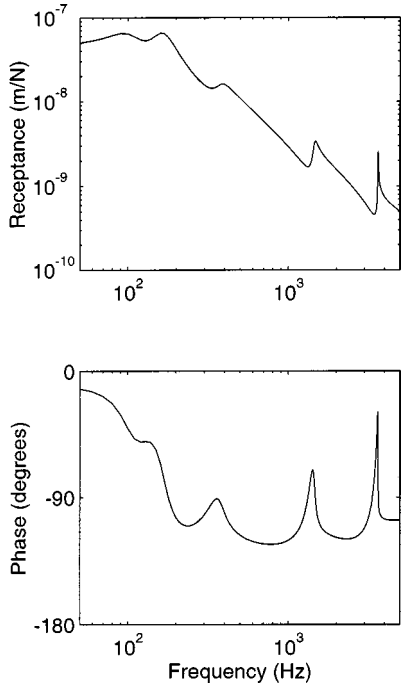


FIG. 7. Amplitude and phase of the point receptance for lateral vibration of the continuously supported rail.

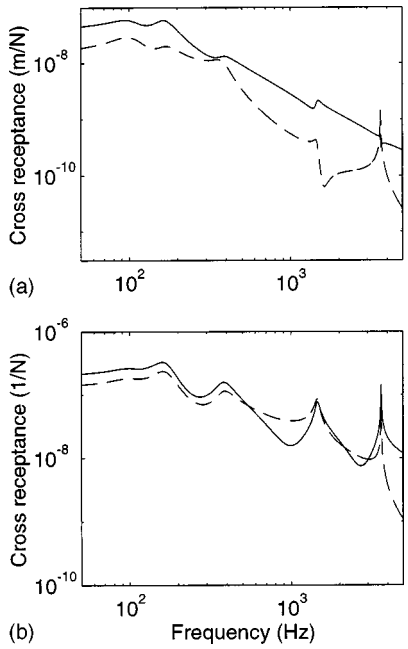


FIG. 8. Cross receptances for lateral vibration of the continuously supported rail. Force acts at the top of the rail head (a) Translational response and (b) rotational response: ——— for rail head and --- for rail foot.

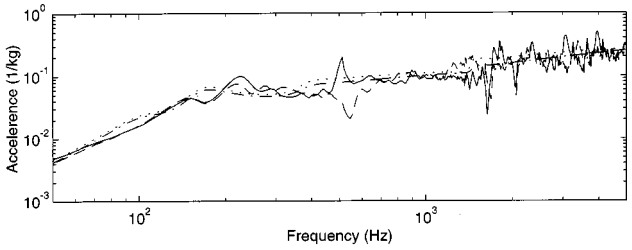


FIG. 9. Comparison of predictions from the multiple beam model and measurements (from Ref. 11): ——— measured at mid span, --- measured above sleeper, -·-·- predicted at the center of head, and ··· predicted at the top of head.

D. Comparison with experimental results

While no new measurements have been made for this study, some experimental results are available in Ref. 11. One of them was obtained from UIC 60 rail with concrete monobloc sleepers. The measurements were carried out on unloaded tracks by means of impact excitation using an instrumented hammer. For lateral vibration the excitation acted in the center of the rail head and the accelerance at the same place on the opposite side was measured. The measured and predicted lateral accelerances in the center of the rail head are compared in Fig. 9. The input parameters for the track are as in previous sections. A quite good agreement can be seen from Fig. 9 throughout the frequency region from 50 to 5000 Hz. The main deviation is that the pinned–pinned resonance peaks vanish in the predictions from the multiple beam model which is continuously supported. These pinned–pinned resonance characteristics can only be obtained using a discretely supported model.

IV. TRANSFER RECEPTANCE—VIBRATION DECAY ALONG THE RAIL

The lateral vibration decay along the rail can be investigated through the transfer receptance of the rail, which represents the combined wave (the lateral bending wave, the torsional wave, and the web bending waves) propagation decay along the rail. Figure 10 shows the transfer receptances for the continuously supported rail at selected frequencies and contains six parts for excitation frequencies at 200, 500, 1200, 1600, 3200, and 3800 Hz. Both the excitation and the response are at the top of the rail head in each case. In Fig. 10 the upper line represents the receptance amplitude, while the lower line represents the real part of the receptance. From the real part the variation of wavelength with frequency can be observed.

It can be seen from Fig. 10 that the lateral vibration decay becomes more and more complicated with increasing frequency because after each cut-on frequency one more propagating wave will be added to the response and each wave has its own decay rate (see Figs. 6 and 11). The superposition of different waves can lead to a ''beating'' effect; see, for example, the result at 1200 Hz in Fig. 10. At this frequency the difference between the wave numbers of the lateral and torsional waves is about 2.7 rad/m (see Fig. 6) which corresponds to a ''beating wavelength'' of 2.3 m which is observed in the receptance amplitude. In general,

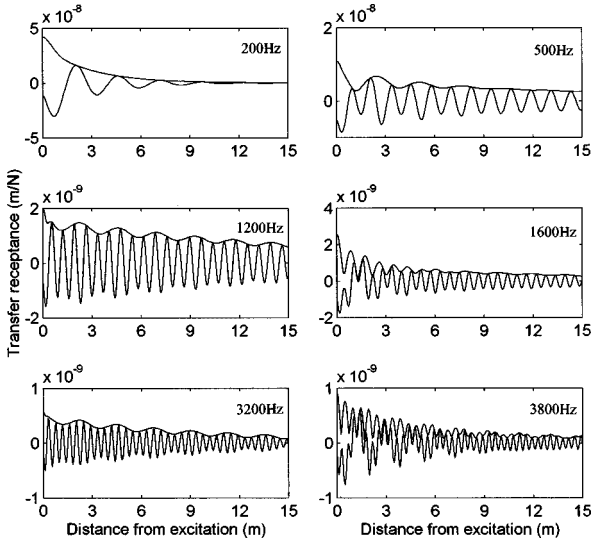


FIG. 10. Combined wave propagation decay in the continuously supported rail. Upper lines: amplitude of transfer receptance; lower lines: real part of transfer receptance.

the vibration decay gradually reduces from low frequencies to high frequencies, except at the cut-on frequencies. Around the cut-on frequencies the transfer receptances experience a noticeable change. After each cut-on frequency a new wave joins the response with its initial long wavelength (low wave number). This new wave combined with other relatively short waves forms a compound wave. On the other hand, the new wave has a higher decay rate near its cut-on frequency (see Fig. 6), so that it decays more quickly than other short waves and at larger distances the short waves are dominant. This can be seen in the graphs for 1600 and 3800 Hz in Fig. 10. For example, it can be seen that at 3800 Hz the second web bending wave is a new contributed wave with long wavelength and decays quickly, so that the short waves are dominant at larger distances. Because of the higher decay rate of the new wave near the cut-on frequency the initial decay rate of the compound wave also becomes higher shortly after the cut-on frequency.

The decay rates of different waves in the continuously supported rail are compared with the measurement data from Ref. 11 in Fig. 11. The measured results represent the average decay rate of the compound wave whereas the predic-

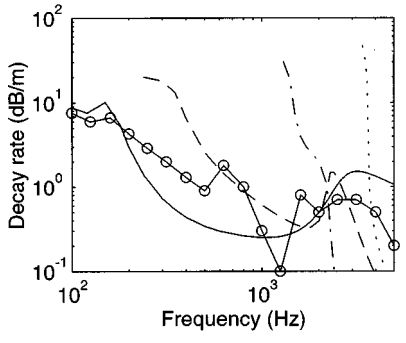


FIG. 11. Decay rates of rail vibration: ○—○ measured (from Ref. 11), ——— wave I, --- wave II, -·-·- wave III, ··· wave IV (from continuously supported rail model).

tions show the result for each wave separately. The measurement will tend to be dominated by the wave with the lowest decay rate. It can be seen that the trends in the decay rates from the theoretical model are consistent with the measurement, apart from the high-frequency region. Above 3500 Hz the theoretical model is under-damped because the loss factor for the rail has been chosen as $\eta_r = 0$. However, setting $\eta_r = 0.01$ led to very high decay rates in this high-frequency region.

V. CONCLUSIONS

In this paper a multiple beam model has been presented for studying the lateral vibration behavior of railway track. This simplified model is less complex than the commonly used FE models, but all essential cross-sectional deformations of the rail can be included. Detailed investigation of the lateral vibration for a continuously supported rail has been carried out efficiently using this model. The results of the free wave dispersion relation from this model have shown a good agreement with a FE model which justifies the neglect of the web twisting and bending stiffnesses in the direction of the rail axis. Quite good agreement between the predictions and measurement data has also been reached in terms of accelerance.

Because the lateral vibration causes complex cross-sectional deformations in the rail, various peaks of the point receptance occur in the frequency region 50–5000 Hz. Among them three peaks appear at low frequencies, which are related to the track foundation dynamic stiffness, whereas two other peaks appear at high frequencies, which correspond to the rail web bending. The decay of the lateral vibration with distance is complicated by the presence of multiple waves. In general, the lateral vibration decay rate decreases from low to high frequency, with some temporary increases near cut-on frequencies.

A more realistic model for railway track vibration should be an infinite rail with periodic supports. Some important characteristics, for example, the pinned–pinned resonance for the lateral vibration, can only be obtained through a discretely supported rail model. The authors will extend the use of the multiple beam model from the continuously supported rail to the discretely supported rail.

[1] S. L. Grassie, R. W. Gregory, D. Harrison, and K. L. Johnson, ''The dynamic response of railway track to high frequency vertical excitation,'' Journal Mechanical Engineering Science **24**, 77–90 (1982).

[2] M. Heckl, ''Acoustic behaviour of a periodically supported Timoshenko beam,'' Report to the ORE Committee C163 (1991).

[3] D. J. Thompson, ''Wheel-rail noise generation, part III: rail vibration,'' J. Sound Vib. **161**, 421–446 (1993).

[4] D. J. Mead, ''A general theory of harmonic wave propagation in linear periodic systems with multiple coupling,'' J. Sound Vib. **27**, 235–260 (1973).

[5] K. Knothe, Z. Strzyzakowski, and K. Willner, ''Rail vibrations in the high frequency range,'' J. Sound Vib. **169**, 111–123 (1994).

[6] L. Gavric, ''Computation of propagating waves in free rail using a finite element technique,'' J. Sound Vib. **185**, 531–543 (1995).

[7] L. Gry, ''Dynamic modelling of railway track based on wave propagation,'' J. Sound Vib. **195**, 477–505 (1996).

[8] Z. Strzyzakowski and L. Ziemanski, ''On the application of the finite strips method to dynamical analysis of vehicle-track systems,'' Z. Angew. Math. Mech. **71**(4), 221–224 (1991).

[9] W. Scholl, ''Darstellung des Körperschalls in Platten durch Übertragungsmatrizen und Anwendung auf die Berechnung der Schwingungsformen von Eisenbahnschienen,'' Fortschritt-Berichte VDI Reihe 11, No. 93 (1987).

[10] W. Scholl, ''Two theoretical models for wave propagation in rails,'' in Proceedings of the Second International Conference on Recent Advances in Structural Dynamics, Southampton, 1984, Vol. II, pp. 699–707.

[11] N. Vincent and D. J. Thompson, ''Track dynamic behaviour at high frequencies. Part 2: Experimental results and comparisons with theory,'' Vehicle System Dynamics Supplement **24**, 100–114 (1995).

Anderson localization of vibration on a framed cylindrical shell

Douglas M. Photiadis and Brian H. Houston
Naval Research Laboratory, Washington, DC 20375-5000

(Received 27 February 1998; revised 7 October 1998; accepted 30 April 1999)

Analysis of the measured response of a near-periodic framed shell to a point force reveals that the higher-order azimuthal modes ($n>10$) on the shell are significantly localized. A quantitative analysis of the data yielding the spatial attenuation rates, the measured group speeds, and the total measured damping factors for various azimuthal orders is presented. The two principal mechanisms which could account for the spatial attenuation rates, damping and Anderson localization, are investigated. Predictions based on these underlying phenomena are obtained and compared with the measurements. The spatial attenuation rates resulting from damping are typically a factor of 2.3 times too small to account for the measurements, while the spatial localization due to irregularity and the data are in reasonable agreement. This evidence, along with the strongly fluctuating nature of the response, indicates that Anderson localization is the dominant mechanism at work in this system. [S0001-4966(99)00209-X]

PACS numbers: 43.40.Ey [CBB]

INTRODUCTION

In a recent publication,[1] Nearfield Acoustic Holography (NAH)[2] measurements of the response of a nearly periodic, framed cylindrical were presented. In that paper, the focus was on the extended modes and, in particular, the wave number space response of the system was analyzed. In this paper, the response of the system in real space is considered, with particular emphasis on the observed spatial attenuation of the relatively large n normal modes. The measurements demonstrate for the first time the prevalence of highly localized resonances in the response of such systems. Based on comparisons with theoretical predictions, we have concluded that Anderson localization[3] is the mechanism producing the observed spatial confinement.

Anderson localization is not simply the confinement of vibration in the vicinity of an external source; rather, Anderson localization has to do with the properties of the normal modes of the system. In a simple system, a system which is spatially uniform or periodic, the normal modes or structural resonances consist of standing waves extending over the entire system, global modes. These modes contain a single dominant length scale, the inverse wave number k_s^{-1} of the natural traveling wave of the system, or, if the system is periodic, a few additional scales resulting from interaction with the periodicity.[4] The modes will also typically contain an additional scale, an attenuation length arising from a combination of structural and radiation damping. This scale will be manifest as a simple exponential decay of the modal envelope. In such a system, the acoustic signature in the far field will show a clear and simple pattern reflecting in large part the dispersion of the natural waves supported by the structure.[5,6] The great simplicity of the resonances in this case is a very general result, arising from the translational symmetry of the system.

In a complex system, a system whose properties vary spatially on many different scales, the mode shapes need not have such a simple nature. In some systems, the resonant modes may continue to be extended but may gradually lose coherence due to the irregular spatial variation in the system properties. In other systems, modes may be localized in the neighborhood of ''defects,'' i.e., deviations in the spatial properties of the system from the mean. These resonances, which may occur throughout the structure with a fairly irregular frequency distribution, are termed localized modes.[7] The mode shapes associated with these resonances on average possess an exponentially decaying envelope, in a similar manner as occurs in the case of attenuation due to damping. The length scale of the spatial envelope is called the localization length, and arises from interference phenomena, not damping. When ''almost all'' of the resonant modes of the system are localized in this way, Anderson localization[3] is said to occur.

One may regard the occurrence of localized modes from two points of view. In the first picture one regards the system as being composed of many dissimilar subsystems coupled to one another. Each subsystem possesses its own resonance spectra in isolation, and, while modified by the coupling to the other subsystems, these individual spectra persist in a perturbed form in the fully coupled system. In this view, local modes are not at all unexpected; instead, extended global modes are a bit difficult to understand. In the second, complementary view, one regards the system as a modified uniform system. In this viewpoint, the irregularity in the system is considered to scatter the natural waves of the underlying uniform system. When this scattering becomes strong enough, local standing waves are formed. In this view, large-scale propagation is easy to understand, but the formation of the local resonances is not as natural.

The main difficulty in experimentally observing the Anderson localization of structural vibration lies not in the measurement of the localization of vibration in the vicinity of a source, but rather, in the elimination of competing mechanisms which can give rise to similar behavior. So in addition to comparing theoretical predictions of the localization resulting from irregularity with experiment, we also consider the following mechanisms: geometrical spreading, in-

ternal damping, and acoustic radiation damping.

Typically, in a direct spatial measurement of a two-dimensional system with such a large density of states,[8] one would be unable to resolve individual resonant modes. As a result, one would observe a significant spatial falloff arising merely from geometrical spreading as one moves away from the source. This spatial behavior is not reflective of the individual mode shapes, but instead is the result of a coherent superposition of many modes in the neighborhood of the source. In the framed shell system, because of the axisymmetry, we may examine motion only in the axial direction simply by Fourier transforming the measurements in θ. Each individual azimuthal mode examined in this way allows motion only in one dimension. Hence in this experiment spreading is not an issue. In addition to eliminating geometrical spreading, the density of states is greatly reduced by this signal processing and one typically observes only a few resonances of the system at each frequency. Constructive interference between these modes will result in an enhancement of the response near the drive point, but at distances greater than a single correlation length,[9] the spatial decay does not arise from such modal interference phenomena; the decay rates associated with forced solutions involving several modes coincide with the fundamental modal localization factors.[10]

Internal damping and acoustic radiation damping both give rise to an exponentially attenuated response as one moves away from the drive. We do not take extraordinary measures to reduce this phenomenon. Instead, utilizing the measured damping factor and group speed of the modes, we obtain from a straightforward analysis the spatial attenuation factor which arises from all sources of dissipation. Comparing this quantity, the experimentally derived ''spatial attenuation due to damping,'' to the direct spatial measurements of attenuation, we determine whether damping accounts for the observed localization. When the modal damping is large, as is the case for some small n modes as a result of radiation into the surrounding fluid, the damping indeed accounts for the observed localization, and the effects of the small amount of irregularity in the system cannot be observed. When the modal damping is small, however, as is the case for a large group of modes which are acoustically cut off in the frequency band of interest, the experimentally derived ''spatial attenuation due to damping'' does not account for the observed localization of the response. And further, the very strong fluctuations of the observed attenuation factor of these modes strongly suggest that a simple dissipation mechanism is not the source of the observed localization phenomena. Based on this analysis, and the comparison of the theoretical predictions of the localization lengths due to irregularity with experiment, we have concluded that Anderson localization is the dominant spatial attenuation mechanism in the framed shell system.

The term Anderson localization is here meant to convey the observation that most of the modes of the system in this frequency range, almost all the large n modes, have become localized resonant states as a result of the irregularity. Recall from the above discussion that in the analogous exactly periodic system, the corresponding states, the Bloch or Floquet modes,[4] are all extended. Thus as a result of the small amount of irregularity in our system, the qualitative character of the measured response of the system to a localized excitation has dramatically changed. The excitation levels near the drive point have become greatly enhanced and the resonant local modes dominate the drive point admittance and the nearby response. A spatial measurement of the ''effective damping'' of the system yields very large attenuation rates. But damping is in fact not the mechanism behind this spatial falloff in the response. Instead, one is simply observing a buildup of trapped reverberant vibrational energy.

The observations described above are not in fact sufficient to indicate Anderson localization in a strict sense—the localization of all the resonant modes of the system by irregularity. The low-frequency modes whose wavelengths are of the order of the size of the system are of course never localized by weak irregularity. Nor in the case which we examine are the majority of the small n extended modes substantially localized; the influence of irregularity on these modes is typically too weak to produce localization lengths smaller than either the length of the structure or the radiation damping length. Nevertheless, the physical mechanism producing the localization of most of the resonant states of the system in the mid-frequency range is localization by irregularity. We loosely describe this as Anderson localization.

The outline of the remainder of the paper is as follows. In Sec. I, we first present some theoretical background and previously obtained results. In particular, the general formulas for the localization lengths of oscillator chains obtained by Herbert and Jones[11] and Thouless[12] are given. We then employ an approximation technique which retains only those resonances of the framed shell system which lie within a particular frequency pass band of interest, the single band approximation. This approximation results in describing the system effectively as a linear chain of coupled oscillators, and the formulas due to Herbert and Jones[11] and Thouless[12] may be immediately applied.

In Sec. II, the experimental results are presented. The measurements consist of NAH measurements of the vibratory response of a framed cylindrical shell to a local excitation. A substantial portion of the mid-frequency range was covered, roughly $\omega_c < \omega < 4\omega_c$, with $\omega_c = c_p/a$ the ring frequency of the cylinder (a is the radius of the shell and c_p is the longitudinal wave speed). As mentioned above, an investigation of the response of this system in wave number space was recently published by the authors.[1] Here, we examine the spatial response of the system and focus on the localized modes, primarily those modes with relatively large azimuthal mode number, $n \geqslant 11$.

In Sec. III, the results are discussed and analyzed, and the theoretical approach described in Sec. I is applied to the case at hand. The predictions of the theory are compared to the measurements along with predictions of the attenuation arising from the measured dissipation. We conclude with a discussion of the generality of these results and some applications to current research efforts in this area.

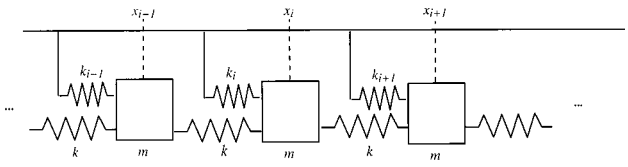


FIG. 1. A schematic of the oscillator chain.

I. THEORY

The goal of this section is to provide a simple theoretical framework to explain our observations of the spatial localization of the large order azimuthal modes of the shell.

A. The oscillator chain

The approximations we describe below rely on a number of known theoretical results concerning the behavior of simple oscillator chains. Here, we review some of the important facts we will need later. The one-dimensional coupled oscillator system is governed by

$$-m\omega^2 x_i = -k(2x_i - x_{i+1} - x_{i-1}) - k_i x_i . \tag{1}$$

The individual oscillators have mass m and stiffness k_i, while the springs coupling the oscillators to one another are taken to have the uniform stiffness value k. A schematic of the oscillator chain is shown in Fig. 1. One quickly finds

$$(\omega_i^2 - \omega^2)x_i = \frac{k}{m}(x_{i+1} + x_{i-1}) = V(x_{i+1} + x_{i-1}), \tag{2}$$

where we have written the coupling of the adjacent oscillators as $V = k/m$. In a statistical sense, an irregular single degree of freedom coupled oscillator chain is characterized by the distribution of resonance frequencies ω_i and the coupling V. To leading order in a perturbative analysis, only the center frequency and the variance of the ω_i^2 are relevant, and to this level of approximation there are therefore only two dimensionless parameters in the system. Thus a number of theoretical results concerning the behavior of this system are known.

1. Spectra of the uniform chain

The uniform chain is quite simple to analyze. Taking the ω_i to be equal, $\omega_i = \omega_0$, the eigenfrequencies can be immediately found by substituting a traveling wave solution $\exp(ik_n x)$ into Eq. (2) above,

$$\omega_n^2 = \omega_0^2 - 2V\cos(k_n a), \tag{3}$$

where a is the mean spacing of the oscillators. Here, we have chosen $k_n \approx 2n\pi/L$, with L the length of the system, to mimic the effect of boundary conditions at the end of the system. Note that the wave number k_n must be restricted to the range $|k_n| < \pi/a$ to avoid overcounting, so there will be the same number of resonances as there are oscillators. The resonances form a single band with a squared frequency width,

$$\Delta(\omega^2) = \omega_+^2 - \omega_-^2 = 4V, \tag{4}$$

as shown in Fig. 2. The frequency bandwidth of the periodic system is hence a direct measure of the coupling between adjacent elements.

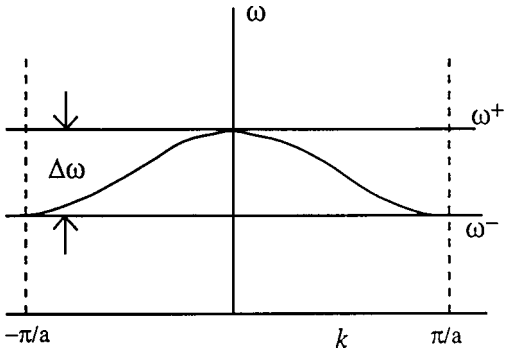


FIG. 2. The band structure of a uniform linear chain of oscillators.

2. Estimates of the localization length

In the presence of irregularity, the modes of almost all such chains are localized standing waves, apart from finite length effects. The spatial extent of these standing waves is characterized for a particular mode by the localization length l or localization factor γ, where

$$\phi_\omega(z) \to \exp(-|z|/l(\omega)) = \exp(-\gamma(\omega)|z|) \quad \text{as } z \to \infty. \tag{5}$$

Thouless[12] and Herbert and Jones[11] give the following results for a linear chain with coupling V and a uniform distribution $\rho(\omega_i^2)$ of ω_i^2 over the range $[\omega_0^2 - W/2, \omega_0^2 + W/2]$. For the case in which the frequency spread due to irregularity is smaller than the bandwidth, one has

$$\gamma(\omega)a \approx \frac{1}{24}\frac{W^2}{4V^2 - (\omega^2 - \omega_0^2)^2} \quad (W < V). \tag{6}$$

For the opposite case in which the frequency spread due to irregularity is greater than the bandwidth, the result is slightly more complicated. Letting $E = (\omega^2 - \omega_0^2)$, Thouless gives

$$\gamma(\omega)a \approx \frac{1}{2}\ln\frac{W^2 - 4E^2}{4e^2V^2} + \frac{E}{W}\ln\frac{W + 2E}{W - 2E} \quad (W > \Delta\omega^2). \tag{7}$$

The important point to notice regarding the above formulas is that in addition to the frequency, there are only two dimensionless parameters, the frequency spread due to irregularity W and the coupling V. In applying the formulas from the references to the vibrational chain, we have exploited the analogy between the steady-state quantum and classical equations of motion: $E \to (\omega^2 - \omega_0^2)$.

At midband, the situation is even simpler, with only the ratio of these parameters being needed to predict the localization length. The results are

$$\gamma(\omega)a \approx \frac{1}{96}\frac{W^2}{V^2} \quad (W < V, \omega \approx \omega_0) \tag{8}$$

and

$$\gamma(\omega)a \approx \ln\frac{W}{2eV} \quad (W > V, \omega \approx \omega_0). \tag{9}$$

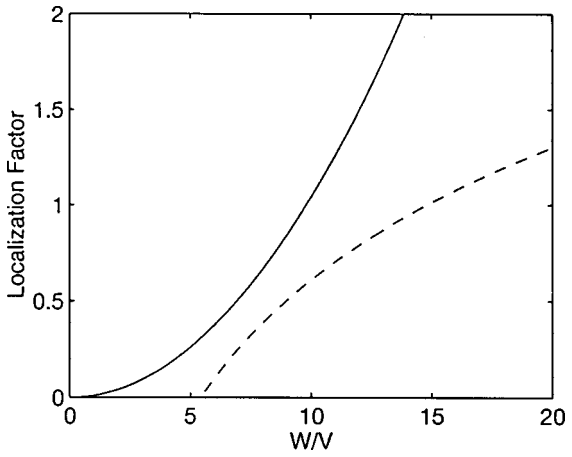


FIG. 3. The localization factor as a function of irregularity; the solid line is the small W approximation, while the dotted line is the large W approximation.

The midband approximations of the formulas for the localization length above are shown in Fig. 3. One may ascertain from the figure that for $W/V \leqslant 5$ the ''weak localization'' formula, Eq. (8), is a reasonable approximation, while for $W/V \geqslant 10$ the ''strong localization'' formula, Eq. (9), is reasonable. In the transition range, a reasonable estimate of the localization length can be obtained by interpolating between the two curves since the localization factor is known to be a monotonic function of irregularity.

The model analyzed by Thouless may appear to be too restrictive, requiring in addition to a linear chain of coupled oscillators a uniform distribution of resonance frequencies. But in fact, the result for the small irregularity case, Eq. (6) above, depends on the shape of the distribution only through the variance of the squared frequencies, $\sigma^2(\omega_i^2)$. This quantity is simply related to W for a uniform distribution via

$$\sigma^2(\omega_i^2) = \frac{W^2}{12}, \tag{10}$$

and therefore the above $W < V$ formulas are quite general. The large W formulas above are dependent on the form of the distribution of the resonance frequencies and would need to be evaluated in a case-by-case fashion. This modification requires that an integral of the form $\int \rho(\omega'^2)\ln(\omega^2 - \omega'^2)d\omega'$ be evaluated; for further details see Ref. 12. In this paper, we need only the small W formulas so detailed information regarding the frequency distribution is not required, only the variance $\sigma^2(\omega^2)$.

We note that the application of these concepts to vibration problems was introduced by Hodges,[13] who derived, employing somewhat different techniques than Thouless, approximate results which are in accord with the above formulas. A number of other investigators later derived related formulas in the context of vibrations, but most notable in this context is the work due to Pierre[14] and collaborators,[15–17] who investigated a number of aspects of one-dimensional vibrational systems including the coupled effects of damping and irregularity. In Sec. IV, we shall employ one of the simplest results obtained in these studies; in the weak localization regime, $W < V$, one may obtain the net localization factor as a simple sum of the localization factors resulting from irregularity and dissipation independently.[15] For a review of additional theoretical research in this area, see the recent review in Ref. 18.

B. The single band approximation (SBA)

Consider now a one-dimensional irregular wave-bearing system, such as a single azimuthal mode of the framed shell in our experiment. In the absence of irregularity, a typical spectrum of the resonant Bloch wave solutions of the system is shown in Fig. 4. In this case, instead of just a single frequency band of resonant states as for a chain of oscillators, there are multiple bands. The single band approximation (SBA) consists of assuming that the effects of irregularity can be estimated in a particular frequency band by including only the states within the band of interest, i.e., in a theoretical sense, replacing the actual system with a chain of oscillators which reproduces the resonance spectrum within the band.[19] This has been indicated pictorially in the figure. One would expect that this approximation would be most valid for bands which are not too close to one another.

Note that the single band approximation technique is not intrinsically associated with one-dimensional systems. In the case of two or three dimensions, such an approach leads to a

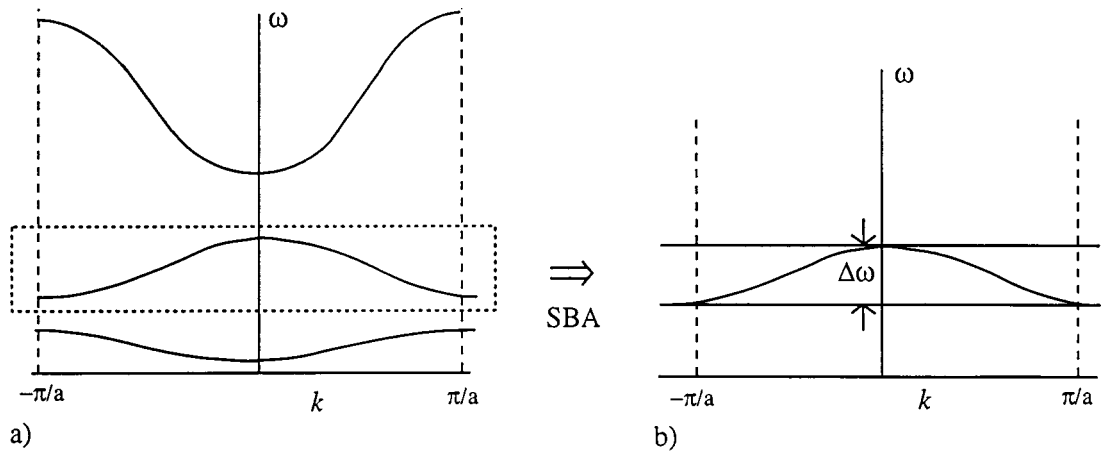


FIG. 4. (a) A typical full spectrum of a single azimuthal helical flexural mode; (b) the spectrum of the azimuthal mode in the single band approximation.

model consisting of an array of coupled oscillators, a great simplification to be sure, but still a much more complex system than the linear chain which arises in the one-dimensional wave-bearing system.

Applying the single band approximation to a vibrational system entails determining the appropriate linear chain to use to describe the system in the band of interest. This correspondence requires two pieces of information: the coupling of the adjacent oscillators V and the distribution of local oscillator frequencies. The first piece of information, the coupling V, is relatively easy to estimate since it is directly related to the pass bandwidths of the system. One estimates via measurement or theory the pass bandwidth in the frequency range of interest, and then chooses the coupling V for the oscillator chain via $\Delta(\omega^2)=4V$ as given in Eq. (4) above. The distribution of local oscillator frequencies is not so simple. One must first deduce what modes the local oscillators represent and then determine how much the irregularity shifts the associated resonance frequencies. One might make an intuitive ansatz concerning the nature of the oscillators, as we do in this paper to simplify the analysis, but generally a simple approximation such as this may not suffice.

An analysis to obtain the effective oscillators can in general be done systematically by decomposing the system into individual subsystems consisting of the repeating element, i.e., the single cell. One must then compute (or measure if possible) the resonance frequencies of the single cell. The ''oscillator'' will correspond to the particular resonance of this spectrum which gives rise to the pass band of interest, and the single band approximation will consist of omitting the presence of all other resonances which occur in the spectrum of the single cell. The distribution of frequencies would then be predicted by determining the change of the single-cell resonance frequency as the relevant physical parameters are varied. Pursuing this analysis in the current context would be quite interesting but is beyond the scope of this paper.

The impact of the single band approximation is immediate, for it focuses our attention on determining the two dominant parameters of the system, the coupling V and the variance of the local resonance frequency distribution resulting from irregularity in the system. We will defer the determination of these parameters in our model until the analysis of the experimental data in Sec. III.

C. Effects of damping

The possibility of strong localization in a lightly damped periodic system was suggested by Langley.[20] We employ a different and more general approach to obtain the same result as was found there. As a result of spatially uniform dissipation, the system possesses a damping factor ζ, and a corresponding damping time for modes near frequency ω,

$$\tau_d^{-1}=\omega\zeta/2. \tag{11}$$

Now consider a wave packet consisting of a group of modes within a small range of frequencies. In time τ_d this packet has propagated a distance $c_g\tau_d$ (c_g is the group speed) and hence the spatial attenuation factor is the inverse of this length,

$$\gamma_d=(c_g\tau_d)^{-1}=\frac{\omega\zeta}{2c_g}. \tag{12}$$

This result, in agreement with Ref. 20 for the case of a periodic system, demonstrates how even a small amount of damping can give rise to short attenuation lengths if the group speed of the system becomes small. This always occurs near cutoff frequencies and so damping will generally be a significant source of attenuation in these situations.

Attenuation, just as localization resulting from irregularity, is more important in narrow bands than in broad bands because the attenuation factor given by Eq. (12) is inversely proportional to the group speed of the modes. For a Bloch pass band of width $\Delta\omega$, one may estimate the group speed at the band center by noting that the change in the Bloch wave number in traversing the band is π/a and therefore

$$c_g=\frac{d\omega}{dk}\approx\frac{\Delta\omega a}{\pi}. \tag{13}$$

Thus the spatial attenuation factor due to damping at midband is approximately

$$\gamma_d a\approx\frac{\pi\omega\zeta}{2\Delta\omega}, \tag{14}$$

and will be larger for narrower pass bands. In the case of Anderson localization, the attenuation factor at midband for small irregularity is, from Eq. (8),

$$\gamma(\omega)a\approx\frac{1}{96}\frac{W^2}{V^2}\approx\frac{\sigma^2(\omega_i^2)}{2\omega_0^2(\Delta\omega)^2}, \tag{15}$$

and also increases with decreasing bandwidth. Here we have used the relation between the coupling and the bandwidth, Eq. (4), and also approximated $\omega_+^2-\omega_-^2\approx 2\omega_0\Delta\omega$.

Notice from the above formulas that the attenuation factor due to irregularity depends much more strongly on the pass bandwidths than the attenuation factor due to damping, varying as $(\Delta\omega)^{-2}$ as opposed to $(\Delta\omega)^{-1}$. Typically then, fairly narrow pass bands corresponding to relatively weakly coupled resonances are influenced more easily by irregularity than by damping.

Castanier and Pierre arrived at the same comparison [see Eq. (37) in Ref. 15] enroute to showing that the net attenuation factor for a damped disordered system was simply $\gamma=\gamma_{\text{disorder}}+\gamma_{\text{damping}}$ in the weak localization regime, $W\leqslant V$. Assuming that both the damping factor and the irregularity are of the same order, they concluded that damping effects would typically dominate in this regime. However, no general conclusions of this sort can be made without a knowledge of the damping and the irregularity, which may in fact differ by several orders of magnitude. Later, Bouzit and Pierre[16] carried out an analysis involving coupled beams, and arrived at the same conclusion we have arrived at here: Narrow, weakly coupled bands tend to be more sensitive to irregularity than damping, while the broader bands tend to be more sensitive to damping than irregularity. The precise crossover at which one or the other phenomena dominate

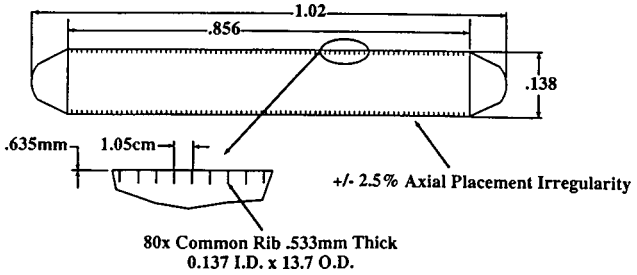


FIG. 5. The experimental model.

depends on the specifics of the system. Indeed, the detailed analysis of our experimental results bears out this prediction.

II. MEASUREMENTS

A. Experimental setup

The experimental structure consists of a thin, nickel cylinder, $h/R \approx 0.01$, with a diameter of about 5 in and a length of about 40 in. The shell is reinforced with 80 ring stiffeners. A drawing of the shell is given in Fig. 5. The rings are identical to within fabrication tolerances and are attached in a near-periodic fashion with mean spacing a along the shell. The detailed positions of the rings along the shell are distributed uniformly and randomly within a small region about the nominal periodic location; i.e., the probability density that the position of a particular ring near $x = na$ has position x is

$$p_n(x) = \begin{cases} 1/\Delta x, & |x - na| < \Delta x/2; \\ 0, & \text{otherwise.} \end{cases} \tag{16}$$

The maximum deviation of the position, $\Delta x/2$, of a ring from its nominal periodic location is about 2% of the rib spacing in the physical model, $\Delta x/a \approx 0.04$. This pseudorandom variation in position was designed using a random number generator and then measured as built on the actual model.

The results we present here consist of NAH measurements of the response of the nearly periodic framed cylindrical shell to a broadband point force. The experiment has been described previously and we refer the reader to Ref. 1 for the details. To summarize briefly, the shell was driven at the center by a point shaker with a broadband pulse. The resulting input force was measured, along with the acoustic response in the surrounding fluid, at an appropriate number of sampling points on a cylindrical surface. These measurements, after appropriate signal processing,[2] enable the reconstruction of the normal surface velocity field of the structure as a function of position and frequency, $v(\theta, z, \omega)$. Scaling this result to the applied force, one obtains direct measurement of the normal component of the Green's function of the fluid-loaded shell to a normal excitation force. The frequency band of the experiment covers the range $8 < f < 40$ kHz, with the ring frequency of the shell given by $f_c \approx 12$ kHz, a reasonable piece of the mid-frequency range.

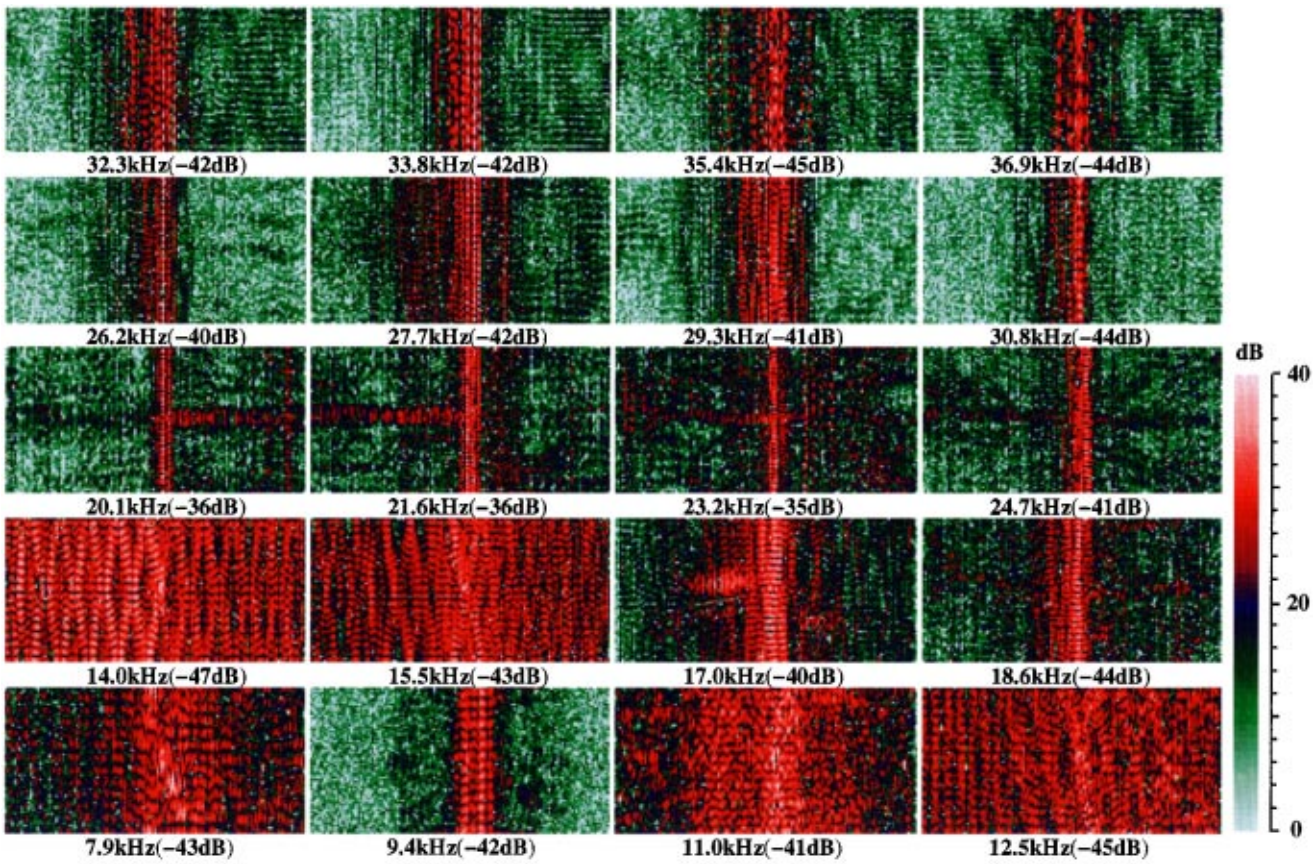


FIG. 6. The response of the shell in real space. The horizontal axis of each subplot is the axial position along the straight section of the shell, while the vertical axis is angle. The shell has been unwrapped.

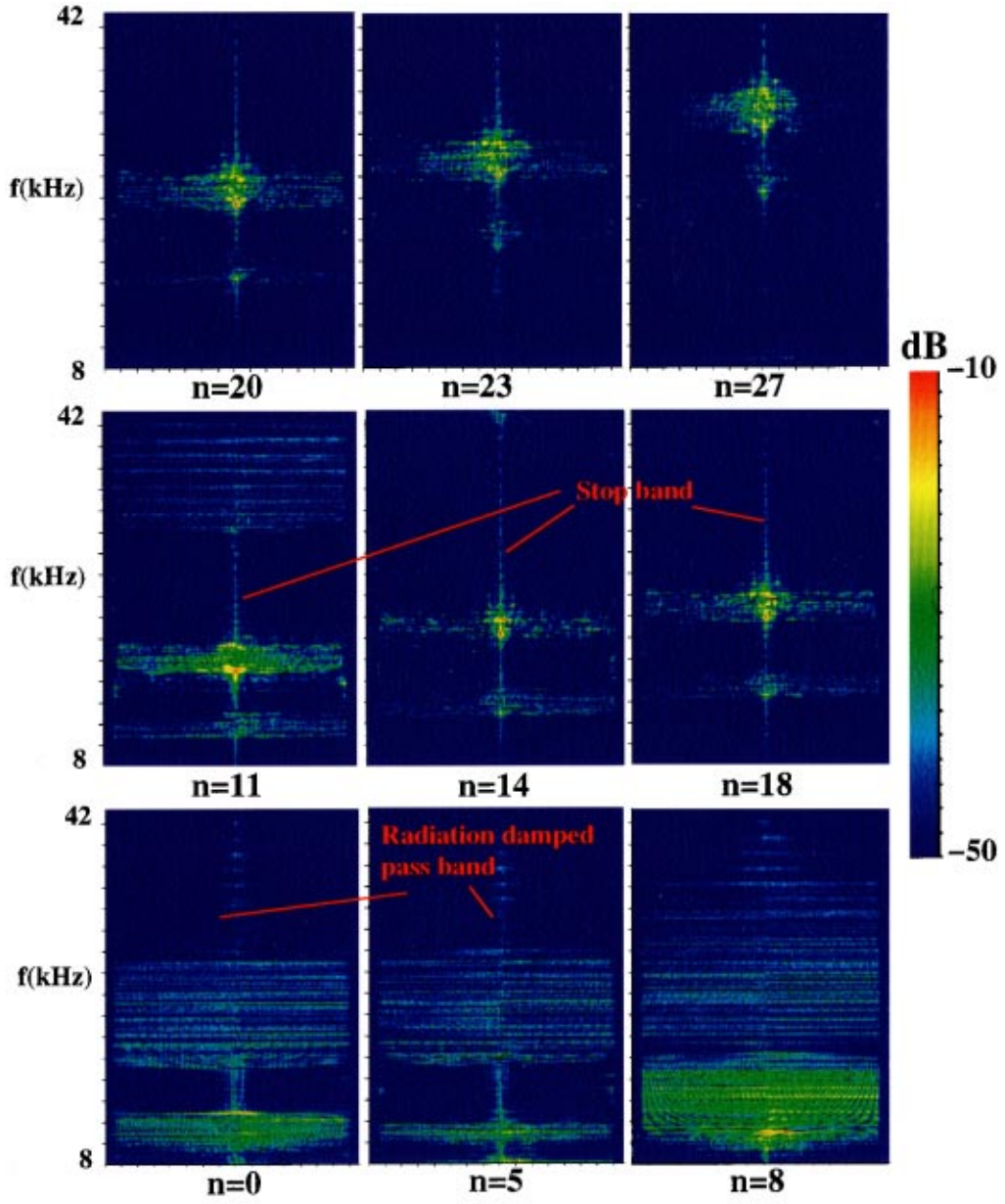


FIG. 7. The response of various azimuthal modes of the shell in real space. The horizontal axis of each subplot is axial position along the shell, while the vertical axis is frequency.

B. Observations

In Fig. 6 is shown the real space response of the shell for several discrete frequencies. It is apparent that the response of the system is much larger near the drive location than elsewhere, though significant global response is also evident. The strong frequency dependence of the response of the system is also apparent. Without further analysis, it is difficult to further characterize the normal modes of the system, and to determine the extent to which we are just observing nearfields associated with either normal stop band/pass band behavior or damping. In particular, attributing the rate of decay of the nearfields to damping would result in estimates of the damping factor which are quite significant.

Fourier transforming the response in θ results in a dramatic simplification because of the axisymmetry of the system; the resonances of the shell are all composed of a single azimuthal order. Further, because the resulting system one examines in this way is effectively one dimensional, with propagation only along the z-direction, the resonances are discrete in frequency. One may therefore directly observe the spatial behavior of only a few resonant modes of the system at a time by plotting the response of the system as a function of space and frequency for individual azimuthal modes. Shown in Fig. 7 is the response of the system analyzed in this manner. The horizontal axis on the plots is position along the length of the shell, while the vertical axis is frequency.

The pass/stop band behavior of the response is very clear. The resonances of the system are clustered into bands because of the presence of the near-periodic array of ribs. In the stop bands, the response of the system is entirely local, consisting only of a nonresonant nearfield. The spatial char-

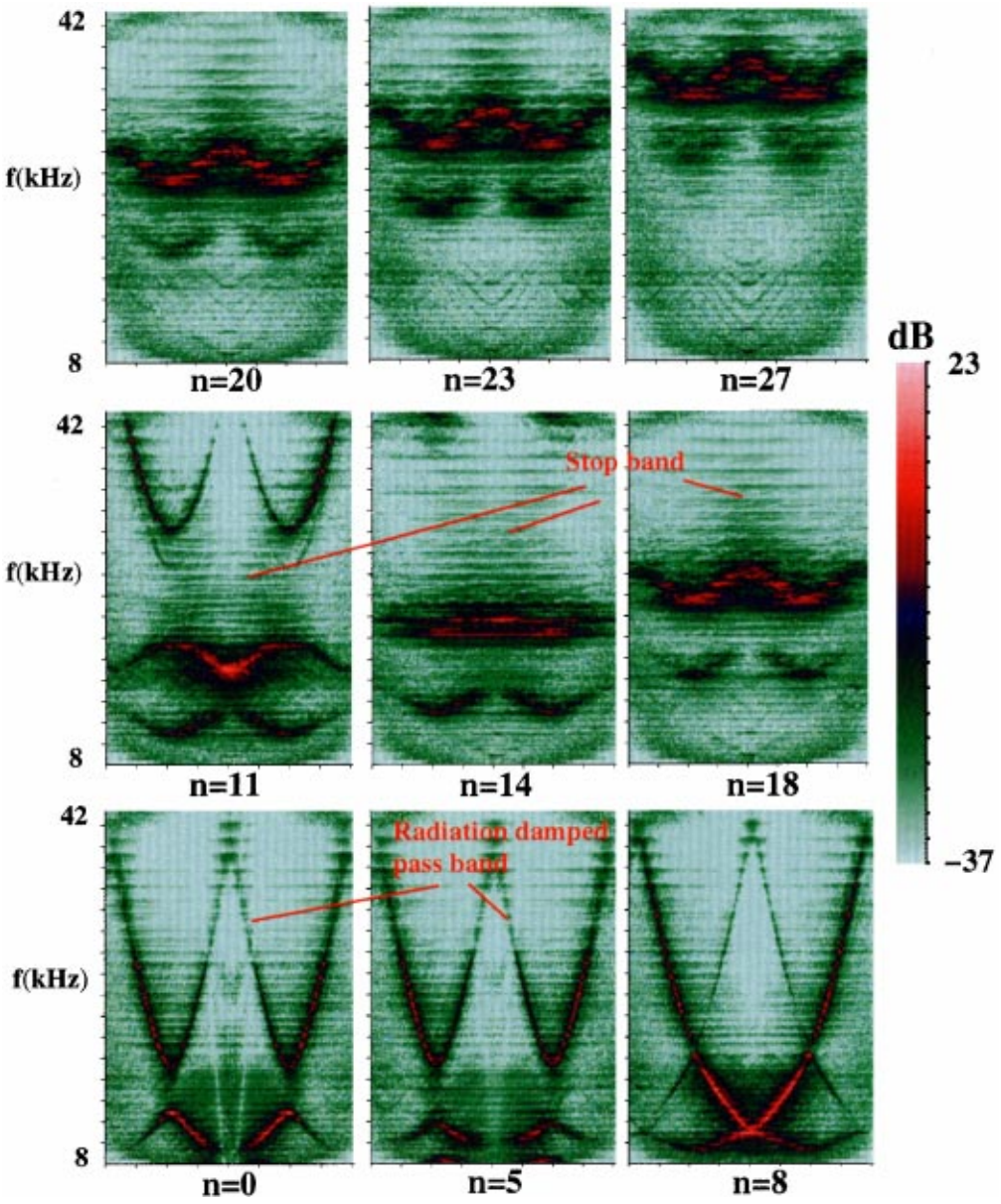


FIG. 8. The response of various azimuthal modes in wave number space. The horizontal axis is axial wave number, while the vertical axis is frequency.

acter of the resonances of the system is immediately apparent from Fig. 7. The low order modes are nearly all extended; that is, have a nearly uniform amplitude along the length of the shell, just as one would expect for the modes of a periodic system. The effects of the irregularity on these modes have not produced significant localization within the dimensions of the shell. The higher azimuthal modes, beginning at roughly $n=11$, are significantly localized. As the azimuthal order n increases, the localization increases, in accord with the theoretical predictions of Anderson localization resulting from irregularity in this system.[21] This general observation, requiring no further analysis, is one of our principal results.

The identification of the onset of the stop bands is not always so obvious for the small n modes, because when the resonant modes within a band develop a supersonic component, the radiation damping is very strong and the spatial response of the modes becomes highly localized. In addition, however, the response at the drive point is also strongly suppressed, so rather than appearing to be local resonances, the onset of radiation damping within a pass band resembles the onset of a stop band. A band illustrating this is indicated in Fig. 7.

This interpretation is made possible by using a wave number space representation. In Fig. 8 is the frequency response of the individual modes displayed in wave number space. Examining the $n=0$ mode, for example, one may observe the dispersion curves for the Bloch wave number of the third pass band converging toward $k_z=0$ as the frequency increases. At a frequency of about 26 kHz, the Bloch dispersion curves enter the sonic cone and become strongly reduced in level. Referring back to the $n=0$ mode shown in real space in Fig. 7, one may observe the sudden strong overall reduction in the response at this frequency, an aspect which might otherwise be mistaken for the onset of a stop band.

An interesting aspect of the wave number space results

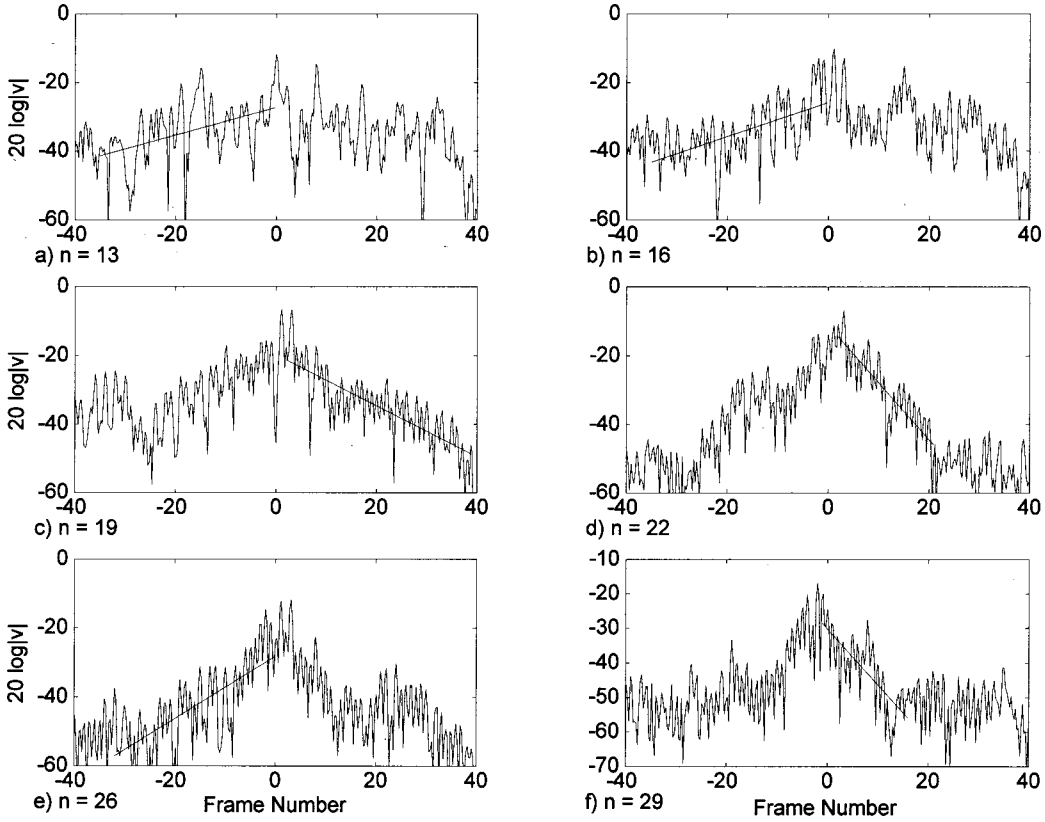


FIG. 9. Line plots giving the response of various azimuthal modes of the shell.

is the appearance of well-defined dispersion curves for the Bloch wave number.[1] For the small n modes, the resolution is quite high, owing to the extended nature of the modes. But even the large n modes usually exhibit a well-defined dispersion structure of the underlying periodic structure, despite being broadened significantly by the spatial localization. This provides several important pieces of information about the system: the group speeds of the various modes via the slopes $d\omega/dk$, and the bay to bay coupling V via the bandwidths of the associated periodic system which can be discerned from the figure.

III. THEORY VERSUS EXPERIMENT

In this section, the techniques employed to obtain the attenuation factors in the context of the specific measurements reported here are described. The experimental results are then compared to the theoretical predictions which result from models based on damping and irregularity, respectively.

A. Obtaining localization factors from measurement

Very near the source, strong fluctuations related to the structural nearfield are expected and observed. As one moves farther from the source, we expect to see an exponential decay of the response on average. Some illustrative plots of the response of particular frequencies within a passband are shown in Fig. 9. An approximately exponential decay of the response is typically observed on average. Our approach therefore is to seek an least mean squares (LMS) fit of the function

$$|\psi(z)|^2 = \exp(-2z/l_{\text{exp}}) = \exp(-2\gamma_{\text{exp}}z), \tag{17}$$

with the data across either the left or right side of the structure excluding the source region. The observed decay rate is the localization factor $\gamma_{\text{exp}}(\omega)$ and may be interpreted[10] as either the decay rate of the particular solution under forcing or, provided the number of frames is large enough, the localization factor of the eigenmode(s) very near frequency ω. The only remaining issue in this regard is to choose the spatial window over which to do the LMS fit so that the noise floor has not been reached. Figure 9(d) and (f) show clearly the wrong estimate one would obtain if an automated procedure using the entire right(left) half of the structure were used rather than cutting off the spatial window when the noise floor is reached.

In Fig. 10, the spatial attenuation factor versus frequency obtained for the dominant resonance band of the $n=18$ mode is shown. While some trends in the result can be discerned, larger attenuation lengths near the edges of the band, for example, very large fluctuations in the attenuation factor are apparent. In order to proceed, some averaging of the data must be performed. One must therefore determine which experimental quantity related to the mode shape is more or less normally distributed; i.e., a quantity whose average value is close to a typical value. Anderson *et al.*[22] have shown that the quantity γ, as opposed to $l=\gamma^{-1}$, for example, is normally distributed and we therefore focus on the spatial attenuation factor γ.

The averaging of the attenuation factor is performed in the center part of each band over the frequency range $(\omega_0 - \Delta\omega/4, \omega_0 + \Delta\omega/4)$ where ω_0 is the center frequency of the

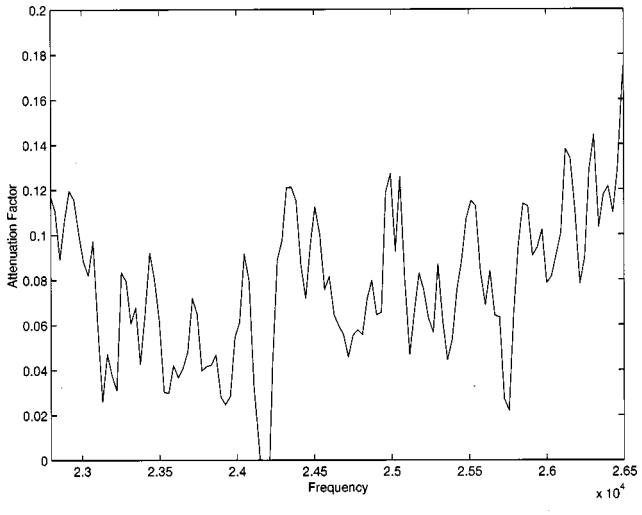


FIG. 10. The measured attenuation factor versus frequency for the dominant resonance band of the $n=18$ mode.

band. Assuming a form of ergodicity, the narrowband frequency average is assumed to be equivalent to an ensemble average and to produce a result suitable for comparison with the theoretical localization factor at the band center. The variation in the result arising from moving closer to the edges of the bands when performing the average gives rise to a small increase in the measured attenuation factor [a factor of $\ln(3)\approx 1.1$ if the attenuation factor follows Eq. (6)] which is much smaller than the remaining statistical uncertainty. This procedure was carried out for a number of experimentally observed localized modes. In the case shown above for the $n=18$ mode, the results were $\gamma=0.085\pm 0.03$; i.e., the magnitude of the fluctuations for this case are about one-third of the average value. This provides us with some sense of the limitations of our ability to make detailed predictions about this sort of system.

B. Attenuation factors from damping

Recall the formula Eq. (14) for the attenuation factor resulting from damping at midband,

$$\gamma_d a \approx \frac{\pi\omega\zeta}{2\Delta\omega}. \tag{18}$$

A theoretical prediction of this quantity requires a knowledge of the loss factor ζ and the bandwidth $\Delta\omega$. We obtain both of these quantities from the measurements.

The bandwidth $\Delta\omega$ is easy to obtain. A subtle point important both here and for the prediction of the attenuation factor resulting from irregularity is that the bandwidth $\Delta\omega$ refers to the bandwidth of the associated periodic system, not the bandwidth of the actual ''irregular'' system. The irregularity produces a spreading of the pass bands and tends to produce overestimates of the bandwidths. In real space, this confusion can lead to significant errors. The ''best'' estimate of the bandwidth of the associated periodic system is obtained by examining the wave number spectra of the data. Examining the wave number space results shown in Fig. 8 carefully, the imprint of the periodic dispersion relation can be clearly discerned and used to determine the correct bandwidth. Nevertheless, a reasonable degree of uncertainty in the bandwidths obtained from the measurements remains and gives rise to a corresponding uncertainty in the theoretical predictions.

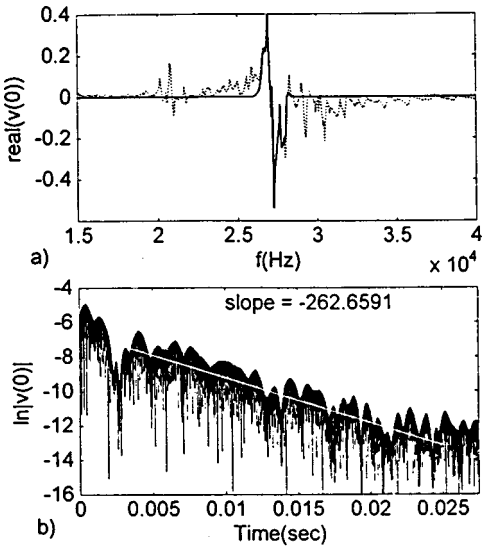


FIG. 11. (a) The frequency response of the $n=22$ mode of the shell. The dotted line is the full measured response of the system while the solid line is the windowed region. (b) The response of the windowed region in time domain. The decay constant can be simply obtained from the slope.

Were the density of states small enough, one could simply examine individual resonance peaks of the admittance and read off the loss factors from the widths. Since individual peaks cannot be resolved, time domain is more convenient. The admittance in time domain, windowed over a bandwidth B, is given in terms of a normal mode expansion as

$$G(x,x,t)=\frac{\pi i}{m}\sum_{\omega_n\in B}\frac{|\psi_n(x)|^2}{2\omega_n}\exp(-i\omega_n t-\zeta_n\omega_n t/2). \tag{19}$$

Notice that the use of the admittance is important here so that neighboring modes cannot cancel and give rise to erroneously small widths. Provided the ζ_n do not vary significantly over the group of frequencies under examination, the slope of the logarithm of $|G(x,x,t)|$ is given by $\zeta_n\omega_n/2$ and yields the loss factor. We see no reason that the loss factor should vary substantially from mode to mode provided no significant radiation damping is involved, and assume this evaluation to give a reasonable estimate of the loss factor. (Recall that the relatively large n modes are acoustically cut off in this frequency range.)

In Fig. 11, a typical result of this signal processing, simply a windowed Fourier transform of the measured admittance, is shown. The time-domain response decays as a simple exponential over a large time window and the loss factors are easily obtained. In performing this analysis, we used a Fourier transform with a cosine window to reduce the side lobes and varied the center frequency and width of the window in order to uncover changes in the loss factor which might arise from the signal processing or variations of the

loss factor across the various modes. The loss factors obtained in this way were reasonably constant not only within a single resonant band but also among the various bands we examined. Generally, we found for this shell that the nonradiating modes have a loss factor $\zeta \approx .002$. This loss factor, when combined with the measured bandwidths of the various groups of resonant modes, results in a spatial attenuation factor which is nearly constant (as n varies) and given by

$$\gamma_d \approx 0.03. \tag{20}$$

Note that the empirical determination of the damping factor in this fashion includes all sources of modal damping, i.e., any mechanism which could lead to a loss of vibrational energy from a given mode.

C. Attenuation factors from irregularity

Recall from Eq. (8) the formula for the attenuation factor at midband resulting from irregularity in the system for $W < V$,

$$\gamma(\omega)a \approx \frac{1}{96}\frac{W^2}{V^2}. \tag{21}$$

Rewriting this formula using Eq. (10), which relates the width W to the variance σ^2, and Eq. (4), which relates the coupling V to the bandwidth $\Delta\omega$, we have

$$\gamma(\omega)a \approx \frac{2\sigma^2(\omega_i^2)}{(\Delta(\omega^2))^2} \approx \frac{\sigma^2(\omega_i^2)}{2(\omega_0 \Delta\omega)^2}. \tag{22}$$

The bandwidths $\Delta\omega$ have already been obtained in the previous section and we may thus focus on determining the single parameter $\sigma^2(\omega_i^2)$, the variance of the local resonance frequencies of the system.

Within the frequency range of interest, there are three principal resonance bands which one might seek to model using the single band approximation (SBA). The lowest band, not quite within the frequency range of the measurement for small value of n, is narrow and relatively weak. The middle band is still relatively narrow but quite strong, and generates the dominant spectral feature for large n within this frequency range. The third band is very broad and gradually moves out of the measurement frequency window as n increases. We focus here on the dominant large n feature, the second band.

In order to apply the single band approximation, we need to establish a relationship between the ''local resonance frequency'' associated with the second band and the spacing, the physical parameter which is varied experimentally. This resonance is a coupled rib/shell mode, and from finite-element calculations,[23] a significant fraction of the energy always resides within radial motion of the shell. We therefore assume the local resonance frequency associated with this band to be determined by a standing wave ansatz of the form

$$k_{fz}a = C, \tag{23}$$

where k_{fz} is the axial component of the wave number of the fluid-loaded helical flexural wave, and C is an arbitrary constant which will turn out to be irrelevant. Notice that the condition, Eq. (23), is all we need in order to ''specify'' the oscillators of the equivalent chain in the single band approximation; the shift in frequency due to a change in spacing a can be obtained from Eq. (23) and hence the variance of the local resonance frequencies may be determined from the known variation of the rib spacing. We note that an identical condition is also applicable to third pass band, albeit with a different constant C.

The variance of the distribution of squared frequencies is given by

$$\sigma^2(\omega_i^2) = \int d\omega P(\omega)(\omega_0^2 - \omega^2)^2 = \int da' P(a')(\omega_0^2 - \omega^2(a'))^2, \tag{24}$$

where we have utilized the equality $d\omega P(\omega)|_{\omega=\omega(a)} = daP(a)$ which holds[24] by virtue of ω being a single valued function of the spacing a. Taking

$$\omega^2(a) \approx \omega_0^2 + (d\omega^2/da)(a - a_0) \approx \omega_0^2 + 2\omega(d\omega/da)(a - a_0), \tag{25}$$

one quickly finds

$$\sigma^2(\omega_i^2) \approx (2\omega(d\omega/da))^2 \sigma^2(a), \tag{26}$$

where $\sigma^2(a)$ is the variance of the distribution of the spacings between the frames.

Using the standing wave condition, Eq. (23), the derivative $(d\omega/da)$ in Eq. (26) can be written in terms of the helical flexural wave number as

$$\frac{d\omega}{da} = \frac{d\omega}{dk_{fz}}\frac{dk_{fz}}{da} = -\frac{d\omega}{dk_{fz}}\frac{k_{fz}}{a}, \tag{27}$$

while the variance of the spacings may in this case be shown to be simply twice the variance of the distribution of the coordinates; see the Appendix. Hence, we have the result for the variance of the distribution of squared resonance frequencies,

$$\sigma^{(2)}(\omega_i^2) \approx \left(2\omega_0 \frac{d\omega}{dk_{fz}}\frac{k_{fz}}{a}\right)^2 \frac{(\Delta x)^2}{6} \approx \frac{8}{3}\omega_0^4 \cos^4\theta \left(\frac{\Delta x}{a}\right)^2, \tag{28}$$

involving only the center frequency of the band ω_0, the helical angle θ of the flexural wave, and the spread of the spacings. In obtaining the final expression for the variance, Eq. (28), we have used the following facts: (1) the angular dependence of a helical flexural wave is simply $k_{fz} = k_s \cos\theta$, $c_{gz} = c_g \cos\theta$, etc., and (2) for flexural waves the group speed c_g is approximately equal to twice the phase speed.[25]

Gathering together the results, the attenuation factor resulting from irregularity is

$$\gamma(\omega)a \approx \frac{4}{3}\left(\frac{\omega}{\Delta\omega}\right)^2 \left(\frac{\Delta x}{a}\right)^2 \cos^4\theta. \tag{29}$$

The bandwidth $\Delta\omega$ is directly measured as discussed above and the width Δx is known, but the helical angle angle θ must be at least partly derived from theory. This formula is,

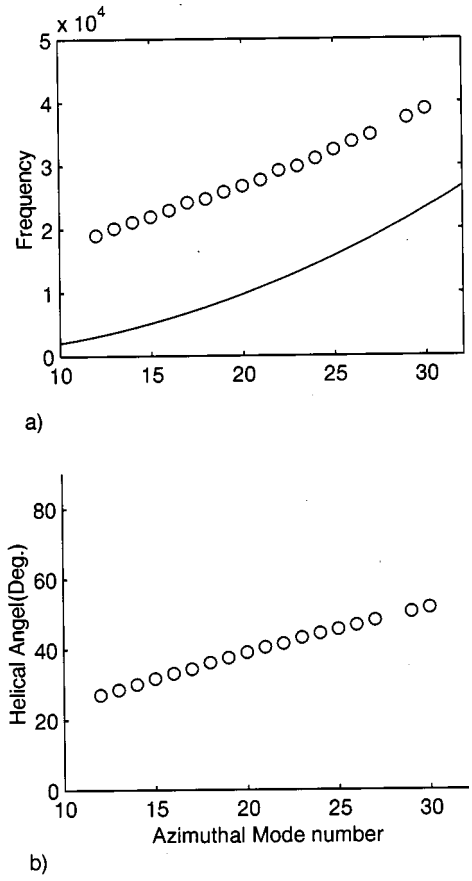


FIG. 12. (a) Band center resonance frequencies versus azimuthal mode number n. The solid line shows the theoretical cutoff frequencies for the various values of n based on a fluid-loaded version of shell theory. (b) Helical angles of the localized flexural modes associated with the various resonances versus n.

as we noted above, applicable to both the second and third pass bands observed in the experiment.

In applying the above formula to the results, we obtain the helical flexural angle as follows. Starting from the thin shell equations of motion including fluid loading, one may, for a particular azimuthal mode n and frequency ω, solve[25] for the axial component of the wave number k_{nz}. Then the magnitude of the wave number is simply $k_n^2 = k_{nz}^2 + (n/a)^2$ and $\cos\theta_n = k_{nz}/k_n$. It is interesting to note that in this frequency range, an ''easy'' technique assuming that the helical flexural angle may be determined using flat plate theory gives the simple analytical expression

$$\cos\theta_n = k_z/k_s \approx \left(1 - \left[\frac{n\pi}{k_s R}\right]^2\right)^{1/2}, \tag{30}$$

in fairly good agreement, better than 5%, with the prediction of shell theory.

The results of carrying out this procedure are shown in Fig. 12. In Fig. 12(a), the center frequency for each band is shown plotted versus n. Also shown on the plot is the cutoff frequency for each flexural mode. The resonances are well above the cutoff frequency, consistent with finite element analysis (FEA) results[23] showing the modes to contain a significant amount of flexural shell motion. In Fig. 12(b) are shown the corresponding helical angles of the various modes. In accord with the resonance frequency being well above cutoff, these helical angles are reasonable in size, but not near the critical angle $\pi/2$.

D. Results and discussion

The experimental results analyzed in the previous section present a clear picture of the behavior of the normal modes of a near-periodic framed shell. It is evident from the measurements that a significant degree of localization is occurring. The localization can be as large as 20–30 dB over a distance of ten frames. There are basically two possible mechanisms which could account for these observations; modal damping phenomena and Anderson localization. Included within the notion of ''modal damping'' is the possibility of the scattering of energy out of a particular azimuthal mode and into another, effectively damping the energy in the initial mode, a mechanism which could arise from small deviations from axisymmetry.

While the response is most often on average exponentially attenuated within the spatial range of the measurement, strong fluctuations, as large as 10–20 dB, about this average occur. Strong peaks occurring away from the drive point are not unusual; this behavior can be seen in Fig. 9(a), (b), and (d). Further, the attenuation rate itself also strongly fluctuates within a particular band as illustrated by Fig. 10. These strong fluctuations, both spatially and with frequency, indicate the presence of local resonances, and must be related to the irregularity in the system; otherwise, the spatial symmetry in the system would prohibit the appearance of such phenomena. Hence, simply from a more or less qualitative standpoint, the occurrence of significant, large fluctuations in the system response is evidence of a form of Anderson localization.

The localized modes occur most typically within clusters of resonances with a fairly narrow bandwidth. The broad pass bands for $n = 0,5,8$ exhibit extended rather than localized response, while the much narrower pass bands for the larger values of n exhibit localized response. As discussed in Sec. I C, this strong dependence on bandwidth is to be expected for localization due to irregularity. This trend is true throughout the data, though the relatively narrow bands for $n = 0,5$ modes are not strongly localized. We have not analyzed these resonances in detail, but the much lower frequency will account for much of this difference; the narrow $n = 0$ pass band at about 10 kHz is roughly the same width as the pass band for $n = 20$ at about 30 kHz, and thus, according to Eq. (29), will have an attenuation length which is about ten times longer. Note also that the localization phenomenon becomes stronger at the pass band edges. Once again, this behavior is well predicted by a theoretical model based on Anderson localization.

The attenuation factors for a number of azimuthal modes as obtained from measurement and theory are shown in Fig. 13. Shown in the figure is the attenuation factor obtained from the measurements along with ''error bars'' ($\gamma \pm \sigma/2$) plotted versus the total attenuation factor resulting from irregularity and damping,[15]

$$\gamma = \gamma_{\text{disorder}} + \gamma_{\text{damping}}, \tag{31}$$

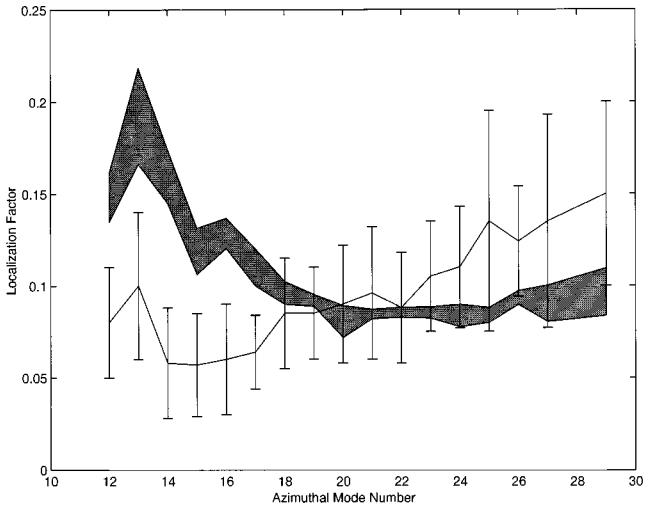


FIG. 13. Experimental measurements versus theoretical predictions of the spatial attenuation factor versus azimuthal mode number. The error bars represent a range of experimental measurements while the filled-in region is the theoretical prediction.

where $\gamma_{\text{damping}} \approx 0.03$ and γ_{disorder} is obtained from Eq. (29). It is important to realize that the ''error bars'' associated with the experimental results are not indicative of experimental error, but instead indicate an actual spread of results as illustrated by Fig. 10. The theoretical estimate shown is the expected ensemble average based on the simple model given in the previous sections. The spread in the theoretical estimate of the localization length arises as mentioned above from the uncertainty in the bandwidths of the system.

It is apparent that the predictions based on localization phenomena yield the correct order of magnitude of the observed attenuation factor. This is indeed significant since no length or time scale of this order is an input of the theory. Instead, the localization length is a result of the dynamics of the system, an interference phenomenon, and could certainly produce a result differing from the measurements by several orders of magnitude. Further, the detailed formula for the localization factor due to irregularity, Eq. (29), depends quite strongly on the parameters; varying in proportion to the variance of the distribution of the positions, inversely to the square of the pass band widths, and to the fourth power of the cosine of the angle of the helical flexural wave. Apart from the variance of the positions, these quantities are not known precisely, and even small changes in the parameters can have a reasonable impact on the predictions. But the order of magnitude of the predictions is not greatly effected by these details and the following is certainly true: For the typical observed pass bandwidths and the amount of known irregularity in the structure, the order of magnitude of the attenuation factors resulting from the combination of irregularity and damping are in accord with the experimentally observed spatial attenuation rates.

The predictions based on a theoretical model involving damping alone simply cannot account for our observations on this shell. The very large spatial fluctuations (Fig. 9) in a system which is nominally periodic cannot be accounted for. The very large fluctuations in the attenuation rate cannot be explained (Fig. 10). The magnitude of the damping, $\gamma_d \approx 0.03$, is too small even in the range $n \approx 15$ where the measured spatial attenuation is smallest: For note that while theoretical predictions of localization factors refer to a mean value, and deviations from this value are expected, predictions of damping are not allowed to fluctuate. It is for this reason that we did not show damping in Fig. 12; predictions based on damping alone are wholly inadequate to account for our experimental results.

The theoretical model does not capture all the features of the measurements, but does exhibit similar trends, increasing as n increases for $n > 20$, and showing a marked peak at $n = 12,13$. It is in this range of n, $12 \leqslant n \leqslant 17$ that the model is most inadequate. Part of this failure results from the experimentally measured mode shapes not exhibiting a clear, convincing exponential decay; e.g., Fig. 9(a) and (b). Hence, a clear determination of the measured spatial attenuation rates is uncertain for these values. Even in this case, however, it would appear that something is missing theoretically. Perhaps the culprit consists of long-range effects of fluid loading,[26] which have been shown in numerical simulations[27] to give rise to unexpected fluctuations, or perhaps our modeling of the frame shell resonances needs to be improved for these values of n.

IV. CONCLUSION

Nearfield Acoustic Holography (NAH) measurements of the response of a nearly periodic, framed shell to a point excitation have revealed the prevalence of localized resonances in the modal spectrum of the shell. In the mid-frequency range, roughly from $1 < \omega R/c_p < 4$, almost all of the large n resonances ($n > 10$) of the shell tend to be highly localized, while because of the very small amount of irregularity in the structure, the small n modes ($n < 10$) are still extended on length scales smaller than or of the order of the size of the shell. The localized modes tend to dominate the real space response of the shell near the source because of their small spatial extent, and thus we observe a significant amount of spatial attenuation in the shell response.

We have analyzed two theoretical mechanisms which could account for the observations. Within the first model, in which attenuation resulting from damping is predicted, it is straightforward to estimate the spatial attenuation factor given a knowledge of the attenuation rate and the group speed, both of which are directly measured. The resulting spatial attenuation factor, when compared to NAH measurements, is too small to account for the experimental observations. For large values of n, the discrepancy is nearly an order of magnitude in this experiment. Further, dissipation phenomena cannot account for the large observed fluctuations in the measured results.

The second mechanism we have analyzed, localization resulting from the irregularity in the system in combination with damping, is considerably more involved. We have developed a simple approximation to the localization factor in this case by modeling each azimuthal mode of the cylindrical shell as a simple oscillator chain. Then, using known results for oscillator chains, theoretical predictions for the localization factors of the various modes are obtained. These predic-

tions are in order of magnitude agreement with the experimental results, and in most cases are within the measurement accuracy.

Based on these comparisons, and the highly fluctuating character of the results, we have concluded that the observed spatial attenuation is primarily due to localization by irregularity, Anderson localization. Further, the phenomena appear to be quite robust; preliminary results[28] indicate that the localized resonances we have observed here persist even in the presence of a significant degree of non-axisymmetry.

This result has a significant impact on a number of aspects of our current efforts to understand and control the acoustic behavior of complex structures. In a big picture context, the discovery that localized resonances are prevalent in framed shells represents a real shift in our overall physical picture of the structural acoustics of this system. Typically, research in the past has focused on global resonances, wave phenomena, and the effects of a few isolated discontinuities. When localized modes are dominant, however, many of the complex problems in these arenas simply cease to be important. Instead, the issues are in many ways simpler because the structure may be considered one piece at a time. Some applications to current issues are the following.

On the theoretical side, a very basic problem in the numerical prediction of the vibration of structures in the mid-frequency range is simply the increase in the number of degrees of freedom N one must include as the frequency increases. Solution times then increase more or less in proportion to N^3, unless a successful banding of the matrices can be achieved. The fact that most of the resonances are local, however, indicates that one need not solve for the entire motion of the structure simultaneously. Instead, each small section may be independently analyzed, and then combined using either a coherent or stochastic technique. This ''divide and conquer'' approach may well prove to greatly reduce the computation required for mid-frequency predictions; a preliminary model[29] of this sort has already been fairly successful in explaining some of the acoustic data of a significantly more complex structure than the one we have considered here.

In a practical context, it is clear that irregularity can be used to achieve acoustic isolation along framed shell structures. Thus in the context of self noise, rather small design changes in the geometry of the supporting elements could have a large effect—and would almost certainly be beneficial. Further, even more gains could likely be achieved by increasing the damping of the structure: By first localizing the vibration as a standing wave using irregularity, and then introducing a small amount of damping, a large payoff from even a small amount of increased damping is likely. In the case of radiation and scattering, similar gains could probably be achieved using the same techniques.

ACKNOWLEDGMENTS

We would like to acknowledge the contributions of Dr. Harry Simpson, who played a critical role in the NAH analysis, and Martin Marcus, who provided supporting computations using the code SARA. This research was supported by the Office of Naval Research, Code 334.

APPENDIX

Let us suppose that the position of the nth rib is x, with probability

$$p_n(x)=\begin{cases} 1/\Delta x, & |x-na|<\Delta x/2; \\ 0, & \text{otherwise}, \end{cases} \tag{A1}$$

as given by Eq. (16). Then the probability that the nth bay has a spacing in the interval $(\delta, \delta+d\delta)$ is

$$P(\delta)d\delta=\left(\int dx' p_n(x') p_{n+1}(x'+\delta) \right) d\delta. \tag{A2}$$

The above integral yields

$$P(\delta)=(\Delta x)^{-2}(\delta x-|a-\delta|)H(\delta x-|a-\delta|), \tag{A3}$$

where H is the Heaviside function. It is then trivial to show that the variance of the distribution of the coordinates is $(\Delta x)^2/12$ while the variance of the distribution of the spacings is $(\Delta x)^2/6$.

[1] D. M. Photiadis, E. G. Williams, and B. H. Houston, ''Wavenumber space response of a near periodically ribbed shell,'' J. Acoust. Soc. Am. **101**(2), 877 (1997).
[2] E. G. Williams, H. D. Dardy, and K. B. Washburn, ''Generalized nearfield acoustic holography for cylindrical geometry: Theory and experiment,'' J. Acoust. Soc. Am. **81**(2), 389–407 (1987).
[3] P. W. Anderson, ''Absence of diffusion in certain random lattices,'' Phys. Rev. **109**(5), 1492–1505 (1958).
[4] L. Brillioun, *Wave Propagation in Periodic Structures* (Dover, NY, 1946).
[5] M. L. Rumerman, ''Contribution of membrane wave reradiation to scattering from finite cylindrical steel shells in water,'' J. Acoust. Soc. Am. **936**, 55–65 (1993).
[6] B. H. Houston, J. A. Bucaro, and D. M. Photiadis, ''Broadband acoustic scattering from a ribbed shell,'' J. Acoust. Soc. Am. **98**(5), Pt. 1, 2851–2853 (1995).
[7] P. Sheng, *Introduction to Wave Scattering, Localization, and Mesoscopic Phenomena* (Academic, San Diego, 1995).
[8] The spacing between the resonant modes is roughly 2 Hz at a frequency of 20 kHz. As a brief point of nomenclature, the term state refers to a natural resonant mode of the system and will be used interchangably with the term mode or naturalmode. Occasionally, the term mode will also refer to a particular azimuthal degree of freedom of the system; e.g., the $n=20$ mode. Hopefully, the particular meaning will be clear from context.
[9] The correlation length of a one-dimensional system is of the order of the localization length—somewhat smaller in fact; see Ref. 7, p. 202.
[10] K. Ishii, ''Localization of eigenstates and transport phenomena in the one-dimensional disordered system,'' Suppl. Prog. Theor. Physics. **53**, 77–138 (1973).
[11] D. C. Herbert and R. Jones, ''Localized states in disordered systems,'' J. Phys. C **4**, 1145–1161 (1971).
[12] D. J. Thouless, ''Percolation and localization,'' in *Ill-Condensed Matter* (North Holland, Amsterdam, 1979), pp. 5–60.
[13] C. H. Hodges, ''Confinement of vibration by structural irregularity,'' J. Sound Vib. **82**(3), 411–424 (1982).
[14] C. Pierre and P. Cha, ''Strong mode localization in nearly periodic structures,'' AIAA J. **27**(2), 227–241 (1989).
[15] M. P. Castanier and C. Pierre, ''Individual and interactive mechanisms for localization and dissipation in a mono-coupled nearly-periodic structure,'' J. Sound Vib. **168**(3), 479–505 (1993).
[16] D. Bouzit and C. Pierre, ''Localization of vibration in disordered multi-span beams with damping,'' J. Sound Vib. **187**(4), 626–648 (1995).
[17] D. Bouzit and C. Pierre, ''An experimental investigation of vibration localization in disordered multi-span beams,'' J. Sound Vib. **187**(4), 649–669 (1995).
[18] H. Benaroya, ''Localization and the effects of irregularities in structures,'' Appl. Mech. Rev. **49**(2), 57–135 (1996).
[19] A similar approximation is often used in condensed matter physics. See N. Ashcroft and D. Mermin, *Solid State Physics* (Saunders College, Philadelphia, 1976), pp. 175–213.

[20] R. D. Langley, ''On the forced response of one-dimensional periodic structures: Vibration localization by damping,'' J. Sound Vib. **178**(3), 411–428 (1994).
[21] D. M. Photiadis, ''Localization of helical flexural waves by irregularity,'' J. Acoust. Soc. Am. **96**(4), 2291–2301 (1994).
[22] P. W. Anderson, D. J. Thouless, E. Abrahams, and D. S. Fisher, ''New method for a scaling theory of localization,'' Phys. Rev. B **22**(8), 3519 (1980).
[23] M. Marcus, B. H. Houston, and D. M. Photiadis, ''Wave localization on a submerged cylindrical shell with rib aperiodicity,'' J. Acoust. Soc. Am. (submitted).
[24] F. Reif, *Fundamentals of Statistical and Thermal Physics* (McGraw-Hill, NY, 1965), pp. 30–31.
[25] M. C. Junger and D. Feit, *Sound, Structures, and their Interaction*, 2nd ed. (MIT, Cambridge, MA, 1972), pp. 235–272.
[26] D. G. Crighton, ''Transmission of energy down periodically ribbed elastic structures under fluid-loading,'' Proc. R. Soc. London, Ser. A **394**, 405–436 (1984).
[27] D. M. Photiadis, ''Anderson localization of one-dimensional wave propagation on fluid-loaded structures,'' J. Acoust. Soc. Am. **91**, 2 (1992).
[28] B. H. Houston and D. M. Photiadis, ''The radiation and vibratory response of a fluid loaded structure with high internal complexity,'' J. Acoust. Soc. Am. **102**(5) Pt. 2, 3072 (1997).
[29] J. A. Bucaro, A. J. Romano, A. Sarkissian, D. M. Photiadis, and B. H. Houston, ''Local admittance model for acoustic scattering from a cylindrical shell with many internal oscillators,'' J. Acoust. Soc. Am. **103**(4), 1867 (1998).

Implications of using colocated strain-based transducers for output active structural acoustic control

Jeffrey S. Vipperman
Department of Mechanical Engineering, 531 Benedum Engineering Hall, University of Pittsburgh, Pittsburgh, Pennsylvania 15261

Robert L. Clark
Department of Mechanical Engineering and Materials Science, Box 90300, Duke University, Durham, North Carolina 27708-0300

(Received 25 January 1999; accepted for publication 16 April 1999)

Piezoceramic transducers have become popular elements of smart structures that are used for active vibration control and active structural acoustic control. A spatial differentiation is performed by the piezoceramic transducers since they couple into the strain field of the piezostructure. This differentiation causes higher-frequency modes to be emphasized more heavily, causing the effective compliance of the structure as seen by the piezoceramic transducer to increase with frequency. This nonuniform compliance has significant impact on the performance that can be achieved through colocated direct rate feedback control. It is shown that the rectangular piezoceramic transducer is a low-pass wave number filter with a cutoff frequency inversely proportional to the aperture size. Thus DRFB performance can be greatly improved simply by making the size of the piezoceramic transducer large relative to the size of the structure. The resulting increase in coupling to the lower-frequency modes, which are generally targeted by the control system, results in a much reduced control effort. In the event that a large aperture is not practical, it is shown that dynamic compensation can be used to obtain good performance at the cost of much increased computational complexity. Analytical open and closed loop results for an acoustically radiating simply supported plate piezostructure are presented. [S0001-4966(99)01908-6]

PACS numbers: 43.40.Vn [PJR]

INTRODUCTION

Piezoelectric transducers have become popular elements of smart structures[1–3] that incorporate active vibration control (AVC)[4] and active structural acoustic control (ASAC).[5] They are lightweight, compact, mostly linear, and relatively inexpensive. Two popular types of piezoelectric materials used in smart structures are lead-zirconate titanate (PZT) ceramic (used for sensors and actuators) and poly-vinylidene fluoride (PVDF) polymer (mostly used for sensors). Rather than applying a point or distributed force to the structure, the piezoelectric transducer couples to the strain field of the structure. When two piezoelectric patches of the same shape and spatial aperture are placed on either side of the structure and wired out of phase electrically, theoretically pure bending can be achieved.

It is well known that colocation of sensors and actuators for feedback control greatly improves stability robustness due to the minimum phase characteristic that arises in the structural transfer function.[6] Colocation for a transducer generally requires both the same physical aperture as well as the same coupling mechanism (i.e., point force with point sensor or strain actuator with strain sensor).[7] The ideal way to colocate piezoceramic transducers is to use the same element as both a sensor and an actuator or *sensoriactuator*. A practical means of implementing piezoelectric sensoriactuators using adaptive signal processing was recently developed.[8,9]

Previous work considered discrete arrays of rectangular piezoelectric transducers that could be optimized in location for harmonic feedforward AVC or ASAC[10] and later broadband feedback AVC and ASAC.[11] Others considered single, distributed strain transducers that were shaped,[12,13] shaded,[14] or a combination of both[15] to achieve observability and/or controllability of particular structural modes. Similar aperture design techniques have been employed to approximate the volume velocity, and hence radiation of structures.[16,17] In addition, aperture shaping has been used to achieve certain frequency roll-off characteristics in the control path transfer functions which are amenable to the application of direct rate feedback (DRFB) control.[18] It was realized near the debut of the use of piezoceramics in smart structures that the transfer functions of typical piezostructures exhibit a frequency response function (FRF) magnitude that rolls upward with increasing frequency[1] (i.e., the compliance of the structure increases with frequency), a condition that destabilizes and degrades DRFB control performance. Since active control targets low-frequency modes, an FRF which exhibits decreasing compliance with frequency is desirable and provides a natural bandwidth for the control system. McCain[18] used higher order functions (i.e., linear, quadratic, cubic) in the transducer aperture shape to successfully achieve an FRF magnitude that has decreasing compliance with frequency. This technique was limited to one-dimensional structures (beams), proved difficult in realizing quadratic or cubic shaped apertures, and only considered a single size aperture for the strain transducer (6.7% of the length of the beam).

In this paper, a simple technique for achieving desirable control path frequency response characteristics which are amenable to the application of DRFB control is presented for two-dimensional structures. The technique involves using a rectangular aperture that covers at least one-half the area of the structure. A desirable frequency response function similar to that achieved with a point sensor/actuator is realized. It will be shown that performance of DRFB active structural acoustic control systems increases greatly with large-aperture strain transducers. For cases where the transducer cannot be made large, dynamic compensators obtained from linear quadratic Gaussian (LQG) designs can be used to obtain good control performance at the expense of complexity and loosing robust stability guarantees. In addition to better DRFB performance, the voltage requirements of the large-aperture actuators are much reduced due to the higher coupling into the low-frequency modes which are targeted by the control system.

I. THEORY

A. Piezostructure model

The *piezostructure*, defined to be the combination of the structure and its surface mounted or embedded piezoelectric components, can be described by the second-order MDOF equations which are derived from Hamilton's Principle for a coupled electromechanical system:[2]

Actuator Eqn.

$$(\mathbf{M}_S+\mathbf{M}_P)\,\ddot{\boldsymbol{\eta}}+\mathbf{C}_s\,\dot{\boldsymbol{\eta}}+(\mathbf{K}_s+\mathbf{K}_p)\,\boldsymbol{\eta}=\mathbf{B}_f\mathbf{f}+\boldsymbol{\Theta}\mathbf{v}, \tag{1}$$

Sensor Eqn. $\qquad \mathbf{q}=\boldsymbol{\Theta}^T\boldsymbol{\eta}+\mathbf{C}_p\mathbf{v}, \tag{2}$

where $\mathbf{M}_s$ is the structural mass matrix, $\mathbf{M}_p$ is the piezoceramic mass matrix, $\mathbf{K}_s$ is the structural stiffness matrix, $\mathbf{K}_p$ is the piezoceramic stiffness matrix, $\mathbf{C}_s$ is the structural damping matrix, $\mathbf{B}_f$ is the forcing matrix, $\mathbf{f}$ are the applied structural forces, $\mathbf{q}$ is the charge on the piezoelectric and $\mathbf{v}$ the voltage drop, η is a vector of generalized displacements for the structure, $\boldsymbol{\Theta}$ is the electromechanical coupling matrix, and $\mathbf{C}_p$ is the patch capacitance matrix.[2] The minute backcoupling[19] between the direct and inverse piezoelectric effects given by Eqs. (1) and (2) are ignored for simplicity. Note that sensor equation (2) can be differentiated to give the rate of charge output (current) which is then proportional to the rate of strain in the structure.

Equations (1) and (2) can be put in a state-space form[2]

$$\dot{\mathbf{x}}(t)=\mathbf{A}\mathbf{x}(t)+\mathbf{B}\mathbf{u}(t), \tag{3}$$

$$\mathbf{y}(t)=\mathbf{C}\mathbf{x}(t)+\mathbf{D}\mathbf{u}(t), \tag{4}$$

where $\mathbf{A}$ is the dynamics matrix of the piezostructure, $\mathbf{B}$ is the input influence matrix, $\mathbf{C}$ is the output influence matrix, $\mathbf{D}$ is the feedthrough matrix and is the zero matrix for properly compensated sensoriactuators, $\mathbf{x}(t)$ is the state vector, $\dot{\mathbf{x}}(t)$ is the derivative of the state vector, t is the temporal variable, $\mathbf{y}(t)$ is the output vector, and $\mathbf{u}(t)=\{\mathbf{f}/\mathbf{v}\}$ is the input vector. Equations (3) and (4) were obtained by first normalizing Eq. (1) with respect to the total mass matrix, $(\mathbf{M}_s+\mathbf{M}_p)$, and choosing the natural modal basis as the state vector:

$$\mathbf{x}(t)=\left\{\begin{array}{c}\boldsymbol{\eta}(t)\\ \dot{\boldsymbol{\eta}}(t)\end{array}\right\}. \tag{5}$$

The matrices comprising the state equations (3)–(4) are defined as follows:

$$\mathbf{A}=\left[\begin{array}{cc}\mathbf{0} & \mathbf{I}\\ (\mathbf{M}_s+\mathbf{M}_p)^{-1}(\mathbf{K}_s+\mathbf{K}_p) & (\mathbf{M}_s+\mathbf{M}_p)^{-1}\mathbf{C}_s\end{array}\right], \tag{6}$$

$$\mathbf{B}=\left[\begin{array}{cc}\mathbf{0} & \mathbf{0}\\ (\mathbf{M}_s+\mathbf{M}_p)^{-1}\mathbf{B}_f & (\mathbf{M}_s+\mathbf{M}_p)^{-1}\boldsymbol{\Theta}\end{array}\right], \tag{7}$$

$$\mathbf{C}=[\boldsymbol{\Theta}^T\quad\mathbf{0}], \tag{8}$$

$$\mathbf{D}=[\mathbf{0}\quad\mathbf{C}_p], \tag{9}$$

where $\mathbf{I}$ is the identity matrix, and $\mathbf{0}$ represents the zero matrix. Note that $\mathbf{C}_p=\mathbf{0}$ since the patch capacitance matrix will be compensated (removed).[8,9,20,21]

B. Spatial signal processing

The piezoceramic mass and stiffness and electromechanical coupling matrices in Eqs. (6)–(8) are defined as:

$$\mathbf{M}_p=\int_\chi \boldsymbol{\Psi}_r^T S_p(\chi)\rho_p\boldsymbol{\Psi}_r\,d\chi, \tag{10}$$

$$\mathbf{K}_p=\int_\chi (L_u\boldsymbol{\Psi}_r)^T S_p(\chi)\boldsymbol{\kappa}_p(L_u\boldsymbol{\Psi}_r)d\chi, \tag{11}$$

$$\boldsymbol{\Theta}=\int_\chi (L_u\boldsymbol{\Psi}_r)^T S_p(\chi)\mathbf{e}^T(L_\phi\boldsymbol{\Psi}_v)d\chi, \tag{12}$$

where χ is the domain of the structure, $\boldsymbol{\Psi}_r$ is the assumed displacement distributions from the Ritz model, $S_p(\chi)$ is a spatial aperture of the transducer, $\boldsymbol{\kappa}_p$ is the flexural stiffness of the piezoelectric, L_u is the differential operator for the particular elastic problem, $\mathbf{e}$ is a matrix of piezoelectric constants, L_ϕ is the electrical differential operator, and $\boldsymbol{\Psi}_v$ is the admissible function used for the electrical potential. The piezoelectric constants $\mathbf{e}$ are usually calculated from the piezoelectric strain constants, $\mathbf{d}$, and piezoceramic flexural stiffness matrix taken at constant field, $\mathbf{c}^{\mathbf{E}}$, as $\mathbf{e}=\mathbf{d}\mathbf{c}^{\mathbf{E}}$. The admissible function $\boldsymbol{\Psi}_r$ must obey the geometric boundary conditions[22] and for the model $\boldsymbol{\Psi}_r$ are taken to be the known eigenfunctions of the structure with no attached mass or stiffness. Typical examples for a spatial aperture in two dimensions ($\chi\in\mathcal{R}^2$) are a point force, S_p^{pnt}, and the rectangular distributed strain input, S_p^{dist}, characteristic of a piezoceramic, respectively:[14]

$$S_p^{\mathrm{pnt}}(x,y)=\delta(x-x_i)\,\delta(y-y_i), \tag{13}$$

$$S_p^{\mathrm{dist}}(x,y)=[\mathcal{U}(x-x_2)-\mathcal{U}(x-x_1)][\mathcal{U}(y-y_2)-\mathcal{U}(y-y_1)], \tag{14}$$

where δ is the Dirac delta function, $\mathcal{U}$ is a Heaviside step function, $\{x_i,y_i\}$ are input points and $\{x_1,x_2,y_1,y_2\}$ define the boundaries of a distributed actuator in the x and y dimensions, respectively. Last, the spatial differential operators for

the point transducer and distributed strain transducer, respectively, are:

$$L_u^{\text{pnt}}(x,y)=1, \tag{15}$$

$$L_u^{\text{dist}}(x,y)=\nabla^2. \tag{16}$$

Equations (10)–(12) demonstrate the effect of the transducer spatial aperture on the overall system model of the piezostructure. A very minor impact on the **A** and **B** matrices of the piezostructure model is produced by Eqs. (10) and (11) since typically $M_s \gg M_p$ and $K_s \gg K_p$. However, when considering Eq. (12), a distributed versus point transducer can significantly impact the electro-mechanical coupling matrix $\boldsymbol{\Theta}$, and thus the input and output matrices of the state-space system, Eqs. (7) and (8). The input/output matrices in turn affect the zeros [roots of $\mathbf{N}(s)$ below] of the system transfer function matrix which is computed as

$$\mathbf{H}(s)=\mathbf{C}(s\mathbf{I}-\mathbf{A})^{-1}\mathbf{B}+\mathbf{D}, \tag{17}$$

$$\mathbf{H}(s)=\frac{\mathbf{N}(s)}{\mathbf{D}(s)}, \tag{18}$$

where s is the Laplace variable, $\mathbf{N}(s)=\mathbf{C}\cdot\text{Adjoint}(s\mathbf{I}-\mathbf{A})\cdot\mathbf{B}+\mathbf{D}$ are the numerator polynomials and $\mathbf{D}(s)=|(s\mathbf{I}-\mathbf{A})|$ is the characteristic polynomial of the system, where $|\cdot|$ represents the determinant of a matrix.

1. Simply supported beam piezostructure

A beam or one-dimensional structure with an attached piezoceramic sensoriactuator will first be discussed since the physics are simpler. As previously discussed, the effects of the spatial aperture are primarily observed in the electro-mechanical coupling given by Eq. (12). For the simply supported beam, $\chi \in \Re$ over the interval $(0 \leqslant x \leqslant L_x)$, $L_u = -z(-\partial^2/\partial x^2)$ for pure bending, $\Psi_r(x)=\sin(k_r,x)$, $k_r=(r\pi/L_x)$ for $r=1,2,3,...$ and Eq. (12) becomes

$$\boldsymbol{\Theta}_r = 4\boldsymbol{\Theta}_0 k_r^2\left(\frac{\cos(k_r x_2)-\cos(k_r x_1)}{k_r}\right), \tag{19}$$

$$\boldsymbol{\Theta}_r = 4\boldsymbol{\Theta}_0 k_r^2\left(\sin(k_r X_c)\cdot\frac{\sin\left(k_r\frac{W_x}{2}\right)}{k_r}\right), \tag{20}$$

where x is the spatial coordinate along the beam, L_x is the length of the beam, $\{x_1,x_2\}$ define the edges of the PZT patch, $\boldsymbol{\Theta}_0$ is a constant based primarily on the constitutive properties of the piezoceramic, $X_c=(x_1+x_2)/2$ is the center of the PZT patch and $W_x=x_2-x_1$ is the width of the PZT patch. The k_r^2 term results from the spatial differentiation performed by the distributed strain sensoriactuator. Consequently, the response of the piezostructure increases with increasing frequency or wave number, k_r. This is partially offset by the integral in Eq. (12), which results in a $1/k_r$ term, for a net increase of k_r in Eq. (20).

Next, the first sine term of Eq. (20), $\sin(k_r X_c)$, reflects the well-known fact that when a patch is centered about a nodal line of a mode, the contributions of that mode and all harmonic multiples are zero. For this situation to occur, the following must be true: $(r\pi X_c)/L_x = m\pi$ for $m=1,2,3,...$, or $X_c/L_x = m/r$ for $m,r=1,2,3,...$, which will result in a perfect pole/zero cancelation. For example, when $X_c/L_x=\frac{1}{2}$ (patch is centered on the beam), then all even modes will cancel from the response. Canceling modes from the response of the structure through spatial signal processing can be beneficial since the static contribution from out-of-bandwidth modes will be reduced. The second term of Eq. (20), $[\sin(k_r(w_x/2))]/(k_r)$, is a sinc function which is the result obtained when a Fourier integral transform is performed on a rectangular window.[23] From Eq. (12), the integral transform uses the kernel $\sin(k_r x)$, which converts the spatial dependence of the moment forcing into coefficients corresponding to the generalized forcing. The sinc function is the frequency response of the distributed transducer in terms of structural wave number and rolls off proportional to $1/k_r$. This behavior is well known in discrete time signal processing where it also represents the frequency response of a finite impulse response low-pass filter.[23] Note that both the sine term and the sinc function would partially constitute the modal residues if a partial fraction expansion were performed on Eq. (18).

The response of the piezostructure can be manipulated through the choice of the aperture width, W_x, in two ways. First, if $(W_x/2L_x)=(m/r)$ for $m,r=1,2,3,...$, then again a group of modes can be canceled from the response where the fundamental mode has a characteristic wavelength equal to W_x. Second, W_x controls the bandwidth of the transducer coupling with the structure which, for simplicity, will be taken at the first zero crossing of the sinc function, or one-quarter of the period: $(k_r W_x)/2=\pi/2$ or $W_x/L_x=l/r$. So for the limiting case of $W_x=L_x$, the cutoff frequency is at $r=1$, or the first mode. For the more typical case of $W_x=L_x/10$, the cutoff frequency will be at $r=10$ or the tenth mode of the piezostructure and from the previous constraint; every 20th mode will cancel from the response. As noted in Ref. 18, better roll-off from the spatial filtering can be achieved by using a different window shape such as a Bartlett window (triangular). Better roll-off decreases the contributions of out-of-bandwidth modes at the expense of increasing the cutoff frequency[23] which is consistent with previous work using shaped apertures.[18]

The effects of lowering the cutoff frequency of the structural filter by making the transducer wider is depicted by Figs. 1 and 2. Figure 1 represents a typical situation with a beam having an attached PZT with width $W_x=L_x/8$, and center location $X_c=L_x/8$. There are four parts to Fig. 1 where part 1(a) depicts the magnitude of the frequency response function (FRF), 1(b) shows the pole and zero locations of the structural model, 1(c) shows the normalized low-pass frequency response of the PZT aperture $[\sin(k_r W_x/2)]/k_r$, and 1(d) displays $\sin(k_r X_c)$. Uniform spacing of points in Fig. 1(c) and (d) is not observed since bending waves are dispersive. First, note from Fig. 1(a) that the first mode response at 31 Hz is undetectable, which is unfortunate since this mode is often targeted heavily by the control system. Recall that the values plotted at each mode in Fig. 1(c) and (d) are factors from the modal residues, which illustrates the low response of the first mode.

In contrast, Fig. 2(a)–(d) also show the FRF magnitude

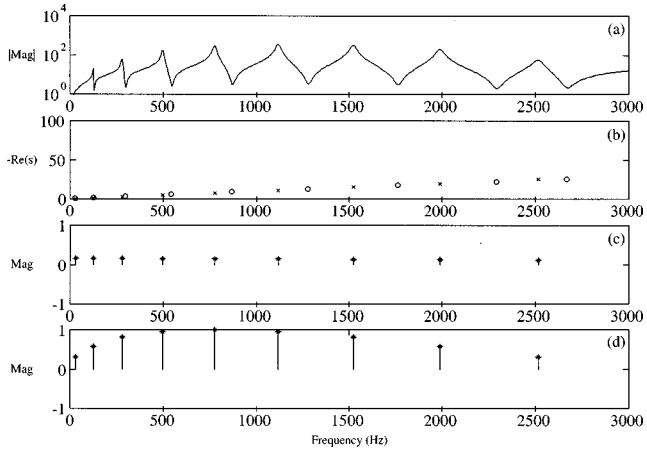


FIG. 1. FRF magnitude (a), pole and zero locations (b), low-pass frequency response $[\sin(k_r W_x/2)]/k_r$ (c), and $\sin(k_r X_c)$ (d) for $W_x = L_x/8$ and $X_c = L_x/8$.

[2(a)], pole/zero plot [2(b)], $[\sin(k_r W_x/2)]/k_r$ response [2(c)], and $\sin(k_r X_c)$ [2(d)], except now the PZT is much wider ($W_x = 0.9 \cdot L_x$, $X_c = 0.45 \cdot L_x$). The response of the first mode at 31 Hz is a couple of orders of magnitude higher, and consequently the FRF begins rolling off in magnitude at a lower frequency. This effect is much more pronounced for a two-dimensional structure since the modal density is higher. Also note that a near pole/zero cancelation occurs at the ninth mode at 2515 Hz due to the choice of the width, which was 1% shy of $4/9 \cdot L_x$.

2. Simply supported plate piezostructure

Similar behavior is observed for a two-dimensional piezostructure where the quantities given by Eq. (12) become $\chi \in \Re^2$ over the intervals $(0 \leqslant x \leqslant L_x)$ and $(0 \leqslant y \leqslant L_y)$,

$$L_u = \left[-z \frac{-\partial^2}{\partial x^2} - z \frac{-\partial^2}{\partial y^2} - 2z \frac{-\partial^2}{\partial x\, \partial y} \right]$$

for pure bending, $\Psi_r(x) = \sin(k_r x)$, $k_r = (r\pi/L_x)$ for $r = 1,2,3,...$, and $\Psi_s(x) = \sin(k_s x)$, $k_s = (s\pi/L_x)$ for $s = 1,2,3,...$ and

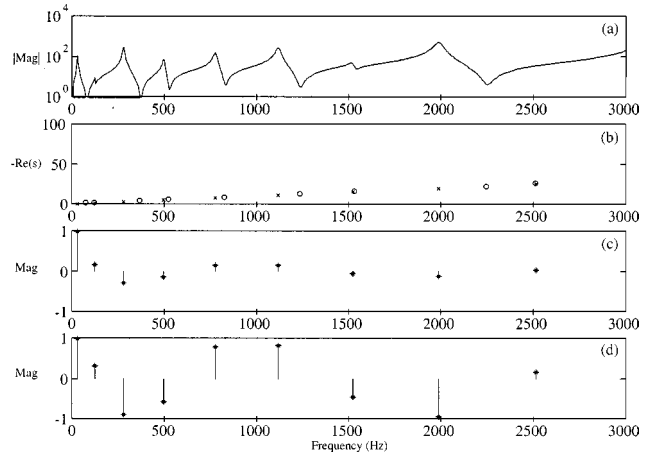


FIG. 2. FRF magnitude (a), pole and zero locations (b), low-pass frequency response $[\sin(k_r W_x/2)]/k_r$ (c), and $\sin(k_r X_c)$ (d) for $W_x = 0.9L_x$ and $X_c = 0.45L_x$.

$$\mathbf{\Theta}_{rs} = 16\mathbf{\Theta}_0 (k_r^2 + k_s^2) \left(\frac{\cos(k_r x_2) - \cos(k_r x_1)}{k_r} \right) \times \left(\frac{\cos(k_s y_2) - \cos(k_s y_1)}{k_s} \right), \tag{21}$$

$$\mathbf{\Theta}_{rs} = 16\mathbf{\Theta}_0 (k_r^2 + k_s^2) \left(\sin(k_r X_c) \cdot \frac{\sin\left(k_r \frac{W_x}{2} \right)}{k_r} \right) \times \left(\sin(k_s Y_c) \cdot \frac{\sin\left(k_s \frac{W_y}{2} \right)}{k_s} \right), \tag{22}$$

where x,y are the spatial coordinates of the plate, L_x, L_y is the width and height of the plate, respectively, $\{x_1, x_2\}, \{y_1, y_2\}$ define the boundaries of the piezoceramic, $X_c = (x_1 + x_2)/2$, $Y_c = (y_1 + y_2)/2$ is the center of the PZT patch along the x and y coordinates, respectively, and ($W_x = x_2 - x_1$), ($W_y = y_2 - y_1$) are the x- and y-dimension widths of the PZT.

Equation (22) is very similar to Eq. (20) but includes terms for both spatial coordinates, x and y. Again, the spatial differentiation causes an upward trend in the response versus frequency which is equal to $(k_r^2 + k_s^2)$. In addition, the rectangular aperture of the PZT again serves to low-pass filter the structural response. Each spatial coordinate is band-limited through selection of W_x and W_y in Eq. (22). However, keep in mind that the structural wave number, k, depends upon both k_r and k_s as $k^2 = k_r^2 + k_s^2$ and thus the aperture of the plate should cover at least 50% of the plate area to couple to all of the lower-frequency modes well. Analogous to the one-dimensional case, groups of modes, (r,s), with integral multiples in either or both spatial coordinates can be canceled from the plate response by selecting: $(X_c/L_x) = (m/r)$ for $m,r = 1,2,3,...$ and $(W_x/2L_x) = (m/r)$ for $m,r = 1,2,3,...$ for modes along the x coordinate or width of the plate and $(Y_c/L_y) = (n/s)$ for $n,s = 1,2,3,...$, $(W_y/2L_y) = (n/s)$ for $n,s = 1,2,3,...$ for modes along the height of the plate.

Performing a partial fraction expansion on Eq. (18) and considering Eqs. (6)–(12), (17), and (22) yields:

$$\mathbf{H}(s) = \sum_{r=1}^{\infty} \sum_{s=1}^{\infty} \frac{\mathbf{\Theta}_{rs}^2}{s^2 + 2\zeta_{rs}\omega_{rs} s + \omega_{rs}^2}, \tag{23}$$

where ζ_{rs} is the modal damping (viscous model), and $\omega_{rs}^2 = (k_s^2 + k_r^2)^2 (\kappa_p/\rho_p)$ are the squares of the structural resonance frequencies. Taking the static response ($s \to 0$) of Eq. (23) and expanding produces:

$$\mathbf{H}(0) = \frac{16\rho_p \mathbf{\Theta}_0^2}{\kappa_p} \sum_{r=1}^{\infty} \sum_{s=1}^{\infty} \left(\frac{1}{k_r^2 k_s^2} \right) \sin(k_r X_c)^2 \sin(k_s Y_c)^2 \times \sin\left(k_r \frac{W_x}{2} \right)^2 \sin\left(k_s \frac{W_y}{2} \right)^2. \tag{24}$$

From Eq. (24), if the structural sampling period or aperture is small ($\{(W_x/L_x), (W_y/L_y)\} \ll 1$) and/or the piezoceramic patch is located near a structural boundary ($(X_c/L_x), (Y_c/L_y) \ll 1$), then the respective $\sin(\cdot)^2$ terms in

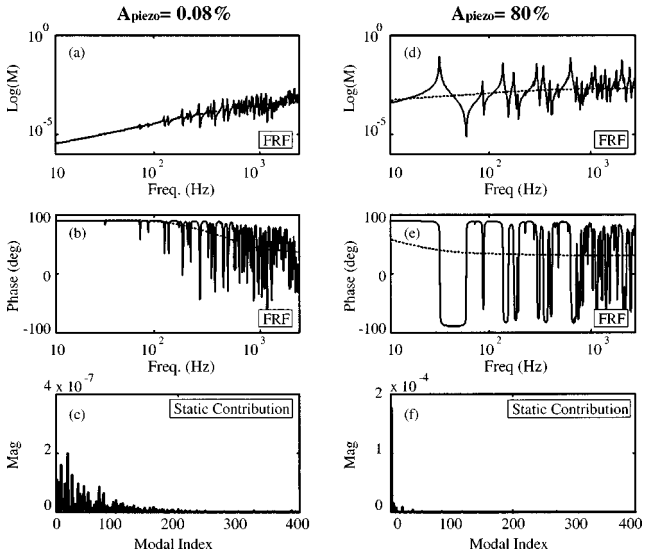


FIG. 3. Resulting FRF magnitudes [(a), (d)], FRF phases [(b), (c)], and static modal contributions [(c), (f)], for two different transducer aperture sizes.

Eq. (24) will initially experience quadratic growth, offsetting the $1/(k_r^2 k_s^2)$ roll-off. In contrast, large structural apertures experience much faster convergence of the static modal contributions given by Eq. (24) due to $1/(k_r^2 k_s^2)$ term. Because of the de-emphasis of higher modes in the response, the structural compliance is more uniform with frequency when using large apertures.

Figure 3(a)–(f) demonstrate the previous points where Fig. 3(a)–(c) correspond to a small aperture PZT (0.8% coverage) and Fig. 3(d)–(f) correspond to the large-aperture PZT (80% coverage). In Fig. 3(a) and (d), the magnitudes of FRF for the two respective transducers are shown. Figure 3(a) is characterized by increasing compliance with frequency while Fig. 3(d) has a relatively flat trend with frequency. These trends versus frequency are more easily observed from the dereverberated plant response, which is shown by the dashed lines in Fig. 3(a) and (d). The dereverberated transfer function is simply the direct field of the vibrational response and hence is a function of only the local dynamics.[24] Many techniques exist to compute or approximate the dereverberated plant response.[24] This particular estimate was obtained by low-pass filtering the complex frequency response of the structure. From the dereverberated responses in Fig. 3(a) and (d), the corner frequencies of $1/0.8k$ and $1/0.08k$ for the structural low-pass filters are more apparent. The corresponding phase angle plots for the FRF are shown in Fig. 3(b) and (e) and also include estimates for the dereverberated system shown by dashed lines. Finally, the static contributions for the first 400 successive modes are presented in Fig. 3(c) and (f). Note that the static response of the structure in Fig. 3(c) is still being mildly influenced out to the 400th mode. In contrast, very little static contribution is seen past the 50th mode for the large-aperture transducer presented in Fig. 3(f).

II. NUMERICAL SIMULATION

An analytical state-space model of the piezostructure, Eqs. (6)–(9), was created in order to predict controller performance. The steel plate measured $0.599\times0.513\times0.002$ m $(w\times h\times t)$ and had simple supports. This plate aspect ratio of 0.86 was chosen to prevent repeated eigenvalues. Constitutive properties for the lead zirconate titanate (PZT) piezoceramics were consistent with type PSI-5A-S3 manufactured by Piezo Systems, Inc. of Cambridge, MA. For all control cases presented, the disturbance source was taken to be a point force input (shaker) arbitrarily located at normalized coordinates of {0.350, 0.096}. The properties of the model and transducer locations correspond to an experimental testbed belonging to the Adaptive Systems and Structures Laboratory at Duke University which is often used for companion experimental studies.[20,21]

Direct rate feedback (DRFB) control is a very simple, robust control implementation which is the focus of this paper. Details of DRFB control design, including with radiation filters, can be found in the literature.[3,20,21,25–28] In addition to DRFB, two other control design methods are considered. Linear quadratic regulator (LQR) designs[3,20,21,28,29] serve as the best-case performance metrics since the availability of full information about the plant is unrealistically assumed. In addition, it will be shown that for plants having nonuniform compliance, and hence poor DRFB control performance, dynamic compensation such as the linear quadratic Gaussian (LQG)[20,21,29,30] design can be used to obtain much improved control performance over DRFB.

Implementing active structural acoustic control (ASAC) required that radiation filters be designed and appended to the state-space model of the piezostructure, which is well described in the literature.[3,20,21,27,28,31,32] The generalized plant model, $\mathbf{G}(s)$, used for control system synthesis and closed loop analysis includes the first 40 structural modes and first 10 radiation filters (i.e., radiation filters are appended to the first 10 structural modes of the model).

Numerical analysis will demonstrate the effect of spatial signal processing on DRFB control performance. First the effect of spatial aperture size on the open loop compliance of the simply supported plate piezostructure will be investigated. Next, DRFB ASAC will be implemented on the plate to demonstrate that uniform compliance results from increasing the spatial aperture, thus enhancing DRFB control performance.[28,33] DRFB active structural acoustic control using point forces and point sensors will not be included in this section, since it was already shown to be a viable approach for ASAC in previous work.[28]

III. THEORETICAL PERFORMANCE

A. Open loop results

In Fig. 3(a), (b), (d), and (e) the open loop responses for PZTs covering 0.8% of the structure and 80% of the structure were given and discussed. Additional points regarding the control implementation will be made from these figures at this time. First, note that the phase presented in Fig. 3(b) and (e) alternates between $\pm90°$, as is typical of colocated structural transfer functions. In addition, note that the the increasing compliance shown by Fig. 3(a) tends to move the zeros of the transfer function much closer to their companion

TABLE I. Normalized center locations of arbitrarily located PZT actuators on simply supported plate experimental/analytical test bed.

Patch	*x*-coordinate	*y*-coordinate
PZT_1	0.10	0.77
PZT_2	0.82	0.41
PZT_3	0.64	0.27
PZT_4	0.23	0.20

poles, which is known to adversely affect single-input, single-output (SISO) DRFB control performance. The first mode at 32 Hz is barely visible in Fig. 3(a), due to a near pole/zero cancelation. In contrast, Fig. 3(d) exibits much more uniform compliance with frequency and associated pole/zero spacing. Finally, when comparing the magnitudes at the first plate resonance (32 Hz) in Fig. 3(a) and (d), a 70-dB increase in coupling is noted for the larger aperture. Therefore the voltage requirements of the large, thin PZT for the same structural response at the frequency of the first mode are much reduced. This observation is very important since the large PZT was 100 times thinner in order to prevent excessively mass loading the structure. Consequently, the depoling voltage of the PZT will be 100 times less, or about 1 V for the PZT material considered. However, the increase of 70 dB in response at the first mode corresponds to a linear mechanical gain of 56.2 in actuation and sensing achieved by increasing the aperture of the PZT. Therefore, although we decreased our maximum actuation voltage by a factor of 100, we achieved a mechanical advantage of 56.2 through simple spatial signal processing.

B. Closed loop results

Three distinct closed loop control cases are presented: two direct rate feedback (DRFB) examples and one LQG example. In the first DRFB example, MIMO control is performed using an array of four small, normal-sized PZT sensoriactuators which cover a total of 3.2% of the area of the structure. Due to the relatively small combined area of the PZT apertures, good DRFB performance is not expected, despite the fact that we have a multi-channel implementation. The four PZTs are arbitrarily located as indicated in Table I. Second, SISO DRFB control is performed with a single, thinner PZT covering 80% of the face of the plate piezostructure. Since the aperture of the this transducer covers more than 50% of the plate area, good DRFB control performance is expected. The third and final control implementation is a MIMO LQG design using the same array of four sensoriactuators used in the first DRFB implementation which had a combined aperture that covers 3.2% of the plate area. For each of the three control cases presented, the performance of a linear quadratic regulator (LQR) design is also presented. Please note that LQR control could not be implemented in practice since full system information is required. However, it serves as a best-case performance metric when analyzing the other control designs.

Figure 4(a)–(c) are bar graphs which correspond to the performance of the three respective control configurations that were described above. Each bar in Fig. 4(a)–(c) represents the reduction in acoustic power in decibels, and is presented for the first eight plate resonance frequencies. These reductions range from 0 dB to 37 dB. If the results had been presented for the full frequency spectrums, minor control spillover would be observed at some off-resonance frequencies. From Fig. 4(a), it is observed that the MIMO DRFB case using a combined aperture that covers 3.2% of the plate area results in poor control performance, particularly at the resonance frequency of the first or (1,1) mode where a reduction of 3 dB is observed. The (1,1) mode is the most efficient radiator and consequently requires the most control performance for ASAC. In contrast, the LQR performance for this actuator configurations is quite good, displaying a reduction of 26 dB in acoustic power at the (1,1) mode frequency. As a result of the good LQR performance, it can be surmised that much better performance is possible through spatial (increased aperture size—DRFB) or temporal (dynamic compensation—LQG) signal processing. The next two bar graphs in Fig. 4(b) and (c) demonstrate these two methods of improvement.

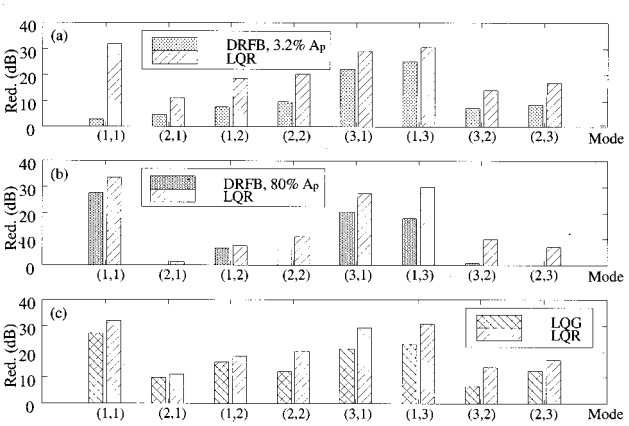


FIG. 4. Bar graph comparing LQR control reductions at resonance with (a) DRFB control with four actuators covering 3.2% of plate area, (b) DRFB control with a single actuator covering 80% of the plate, and (c) LQG control with same actuators used in part (a).

Recall that increasing the spatial aperture to cover at least 50% of the plate surface results in more uniform structural compliance, as noted in Fig. 3(d). When using the large-aperture PZT, much more damping can be added to the structure through DRFB as seen in Fig. 4(b). Much better performance is observed at the (1,1) mode (28 dB of reduction). Further, the DRFB control performance at the (1,1) mode is now comparable to the LQR control reduction of 34 dB. When comparing the LQR results of Fig. 4(a) and (b), note that coupling into some of the higher modes [e.g., (2,1), (1,2), (1,3)] has diminished, primarily due to lowered controllability/observability resulting from using a single control transducer versus an array of four. Finally, note that the control effort penalties used for this control case are 1000 times higher as a result of the significant increase in coupling to the (1,1) mode.

Finally, Fig. 4(c) shows the results of using dynamic compensation with the MIMO patch configuration which covers 3.2% of the plate area. Again, a large reduction of 27 dB is observed at the frequency of the primary radiator as

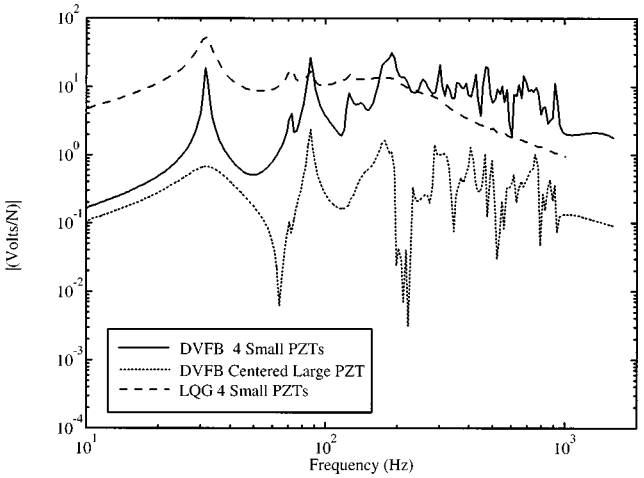


FIG. 5. Comparison of normalized control effort for small-aperture DRFB, large-aperture DRFB, and LQG controllers.

compared to a reduction of 32 dB from the DRFB control. This final case demonstrates the temporal signal processing can be used to obtain good control results when the compliance of the system is not uniform and it is impracticable to use a large-aperture strain transducer. However, the price is increased computational complexity and the loss of stability guarantees.[29,34]

C. Normalized control effort

Control effort can be normalized with respect to the disturbance input force, allowing the control effort of different types of control systems to be directly compared. Figure 5 presents the rms control voltage per rms Newton of disturbance force versus frequency for the three control configurations presented in the previous sections. From Parseval's Theorem, the total normalized control signal energy could be obtained by integrating the curves in Fig. 5 with respect to frequency. The solid line in Fig. 5 represents the DRFB relative control signal energy produced by the four small arbitrarily located PZTs used as DRFB control actuators. The peaks in the curve occur at the location of the (odd,odd) modes such as the (1,1) at 32 Hz, (1,3) at 127 Hz, etc. Next, the normalized control signal energy required for the second control configuration, the large-aperture strain-based control transducer in conjunction with a DRFB control law, is represented by the dotted line in Fig. 5. The required control voltage is nearly three orders of magnitude smaller when compared to the voltage requirements of the small-aperture patches given by the solid line in Fig. 5. This result is consistent with the increased low-frequency coupling noted in the open loop results. Last, the control voltage per unit disturbance force for the LQG control using an LQG control law (configuration 4) is plotted as the dashed line in Fig. 5. The peak value is approximately 42 V/N, which is comparable to the 30 V/N required by the DRFB control result using the same four patches shown in light solid line in Fig. 5. Keep in mind that greatly improved low-frequency control performance is achieved with a 40% increase in peak control voltage. The LQG control energy varies less with frequency when compared to DRFB control since the dynamics of the inherent Kalman–Bucy filter tends to shape the control signal in addition to the plant dynamics.

Note that since the large PZT can couple more efficiently into the lower modes and less efficiently into the higher modes as desired, the control effort and spillover are greatly reduced. Also, note that the maximum rms control voltage is approximately 2 V/N, which is twice the depoling voltage of the 2-μm-thick PZT. If the rms disturbance force were as high as a Newton, the thickness of the PZT could be doubled to 4 μm to eliminate the risk of depolarization occurring. This result is important since it demonstrates that DRFB control using very large-aperture PZTs has feasible voltage requirements and little increase in mass loading since the patches can be made much thinner. If a large spatial aperture ($\geqslant$50% of the area of the structure) is used, adequate control authority can be achieved within the range of operation. Safety is also enhanced since high voltage is no longer required to actuate the PZT transducers. In addition, high-voltage power converters would not be required to operate the controllers using a standard 24-V aircraft electrical system. The disadvantage is that no technique for manufacturing or depositing a 4-μm PZT on to a structure is known by the authors at this time. It is also questionable how damage resistant a thin ceramic transducer would be. These problems pose excellent and exciting research topics for the future.

IV. CONCLUSIONS

The thrust of this research was to take advantage of the implicit physics of the control problem by performing signal processing in the spatial domain in order to reduce the overall control system complexity by reducing temporal signal processing requirements. First, the equations for a piezostructure model were developed for a simply supported beam and plate. It was observed that compliance increases with frequency since the distributed strain transducer spatially differentiates the piezostructure response. Next, it was shown that the uniformity of the compliance can be slightly improved by canceling groups of modes through careful selection of the center location and dimensions of the piezoceramic element. More significant is the effect of the structural low-pass filtering, where the cutoff frequency (in structural wave number) was shown to be inversely proportional to size of the piezoceramic patch (relative to the structural dimensions). In the limiting case where the patch is the same size as the structure, the first mode becomes the cutoff frequency and consequently dominates the structural response. This is beneficial since the first mode is often inherently emphasized by the control cost.

Next, radiation filters were appended to the plate piezostructure model and the effects of aperture size were demonstrated by presenting open and closed loop results. First, direct rate feedback (DRFB) control was implemented using an array of 4 PZT sensoriactuators that cover a total of 3.2% of the plate area. Despite the multivariable implementation, performance was very poor [3 dB on the primary radiator—(1,1) mode] due to the nonuniform compliance of the structural response. Second, DRFB was implemented on a single

patch covering 80% of the plate area, which resulted in a 28-dB reduction in the primary radiator. The increase in performance was also accompanied by a decrease in required control effort by a factor of 1000. Third, a linear quadratic Gaussian (LQG) design was implemented using the same array of 4 PZTs as the first case, which resulted in just less than 30-dB reduction in the primary radiator as well. However, LQG is a dynamic compensation scheme with a much higher computational burden and more robust stability concerns than that associated with DRFB. An additional benefit of the large-aperture sensoriactuators was significantly reduced control voltage due to the increase in coupling to the low-frequency modes targeted by active control systems. Therefore, large strain transducers can theoretically be made much thinner to preclude potential mass loading effects.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the Structural Acoustics Branch at NASA Langley Research Center for funding this research under Grant No. NAG-1-1570. The authors also acknowledge David Cox of the Guidance and Controls Branch at NASA Langley Research Center for numerous insightful comments. J. S. V. conducted this research at Duke University as a Graduate Research Assistant.

[1] E. F. Crawley and J. de Luis, ''Use of piezoelectric actuators as elements of intelligent structures,'' AIAA J. **25**, 1373–1385 (1987).
[2] N. W. Hagood, W. H. Chung, and A. von Flotow, ''Modeling of piezoelectric actuator dynamics for active structural control,'' J. Intell. Mater. Syst. Struct. **1**, 327–354 (1990).
[3] R. L. Clark, G. P. Gibbs, and W. R. Saunders, *Adaptive Structures, Dynamics and Control* (Wiley, New York, 1998).
[4] F. L. Fanson and J. C. Chen, ''Structural Control by the Use of Piezoelectric Active Members,'' Proceedings of 1st NASA/DOD CSI Technology Conference, Norfolk, VA, 1986, pp. 809–829.
[5] C. R. Fuller, C. A. Rogers, and H. H. Robertshaw, ''Active structural-acoustic control with smart structures,'' SPIE: Conf. 1170 on Fibre Optic Smart Structures and Skins II, 1989.
[6] R. J. Benhabib, R. P. Iwens, and R. L. Jackson, ''Stability of large space structure control systems using positivity concepts,'' J. Guid. Control. Dyn. **4**, 487–494 (1981).
[7] S. E. Burke and J. E. Hubbard, ''Distributed Transducers, Collocation, and Smart Structural Control,'' SPIE's Technical Symposium on Optical Engineering and Photonics in Aerospace Sensing, Orlando, FL, Vol. 1307-31, April 16–20 1990.
[8] J. S. Vipperman and R. L. Clark, ''Implementation of an adaptive piezoelectric sensoriactuator,'' AIAA J. **34**, 2102–2109 (1996).
[9] R. L. Clark, J. S. Vipperman, and D. G. Cole, ''Adaptive piezoelectric sensoriactuator,'' U.S. Patent Number 5,578,761, November 1996.
[10] R. L. Clark and C. R. Fuller, ''Optimal placement of piezoelectric actuators and polyvinylidene fluoride (PVDF) error sensors in active structural acoustic control approaches,'' J. Acoust. Soc. Am. **92**, 1521–1533 (1992).
[11] G. C. Smith and R. L. Clark, ''The influence of frequency-shaped cost functionals on the structural acoustic control performance of static, output feedback controllers,'' J. Acoust. Soc. Am. **104**, 2236–2244 (1998).
[12] C. K. Lee and F. C. Moon, ''Modal sensors/actuators,'' J. Appl. Mech. **57**, 434–441 (1990).
[13] R. L. Clark, R. A. Burdisso, and C. R. Fuller, ''Design approaches for shaping polyvinylidene fluoride sensors in active structure acoustic control (ASAC),'' J. Intell. Mater. Syst. Struct. **4**, 354–365 (1993).
[14] S. E. Burke and J. E. Hubbard, ''Spatial filtering concepts in distributed parameter control,'' J. Dyn. Syst., Meas., Control **112**, 565–573 (1990).
[15] J. M. Sullivan, J. E. Hubbard, and S. E. Burke, ''Modeling approach for two-dimensional distributed transducers of arbitrary spatial distribution,'' J. Acoust. Soc. Am. **99**, 2965–2974 (1996).
[16] S. D. Sommerfelldt and B. L. Scott, ''Wave number sensors for active control of acoustic radiation,'' Proc. N. Am. Conf. on Smart Struct. and Mtls. **1917** (1993).
[17] C. Guigou, A. Berry, and F. Charette, ''Active control of plate volume velocity using shaped PVDF sensor,'' Proc. of the 1994 ASME Winter Annual Meeting, Chicago, IL, 1994, pp. 247–255.
[18] A. J. McCain, ''Shaped actuators and sensors for local control of intelligent structures,'' Master's thesis, M.I.T., Cambridge, MA, June 1995.
[19] C. C. Won, ''Piezoelectric transformer,'' J. Guid. Control. Dyn. **18**, 96–101 (1995).
[20] J. S. Vipperman, ''Adaptive piezoelectric sensoriactuators for active structural acoustic control,'' Ph.D. thesis, Duke University, Durham, NC (1997).
[21] J. S. Vipperman and R. L. Clark, ''Multivariable feedback active structural acoustic control using piezoelectric sensoriactuators,'' J. Acoust. Soc. Am. (accepted for publication).
[22] L. Meirovitch, *Analytical Methods in Vibrations* (Macmillan, New York, 1967).
[23] A. V. Oppenheim and R. W. Schafer, *Discrete Time Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1989).
[24] D. G. MacMartin and S. R. Hall, ''Broadband control of flexible structures using statistical enginery analysis concepts,'' J. Guid. Control. Dyn. **17**, 361–369 (1994).
[25] M. J. Balas, ''Direct velocity feedback control of large space structures,'' J. Guid. Control. Dyn. **2**, 252–253 (1979).
[26] W. S. Levine and M. Athans, ''On the determination of the optimal constant output feedback gains for linear multivariable systems,'' IEEE Trans. Autom. Control. **AC-15**, 44–48 (1970).
[27] W. T. Baumann, W. R. Saunders, and H. H. Robertshaw, ''Suppression of acoustic radiation from impulsively excited structures,'' J. Acoust. Soc. Am. **90**, 3202–3208 (1991).
[28] R. L. Clark and D. E. Cox, ''Multi-variable structural acoustic control with static compensation,'' J. Acoust. Soc. Am. **102**, 2747–2756 (1997).
[29] P. Dorato, C. Abdallah, and V. Cerone, *Linear-Quadratic Control: An Introduction* (Prentice-Hall, Englewood Cliffs, NJ, 1995).
[30] J. C. Doyle, K. Glover, P. P. Khargonekar, and B. A. Francis, ''State-space solutions to standard $\mathcal{H}_2$ /$\mathcal{H}_\infty$ control problems,'' IEEE Trans. Autom. Control. **34**, 831–847 (1989).
[31] N. S. Lomas and S. I. Hayek, ''Vibration and acoustic radiation of elastically supported rectangular plates,'' J. Sound Vib. **1**, 1–25 (1977).
[32] K. A. Cunefare, ''The minimum multimodal rediation efficiency of baffled finite beams,'' J. Acoust. Soc. Am. **90**, 2521–2529 (1991).
[33] F. M. Fleming and E. F. Crawley, ''The zeroes of controlled structures: Sensor/actuator attributes and structural modeling,'' AIAA-91-0984-CP, pp. 1806–1816 (1991).
[34] R. Lozano-Leal and S. M. Joshi, ''On the design of dissipative LQG-type controller,'' Proceedings of the Conference on Decision and Control, 1988, pp. 1645–1646.

A curved piezo-structure model: Implications on active structural acoustic control

James K. Henry[a] and Robert L. Clark[b]
Department of Mechanical Engineering and Materials Science, Duke University, Durham, North Carolina 27708-0302

(Received 27 July 1998; accepted for publication 19 May 1999)

Current research in Active Structural Acoustic Control (ASAC) relies heavily upon accurately capturing the application physics associated with the structure being controlled. The application of ASAC to aircraft interior noise requires a greater understanding of the dynamics of the curved panels which compose the skin of an aircraft fuselage. This paper presents a model of a simply supported curved panel with attached piezoelectric transducers. The model is validated by comparison to previous work. Further, experimental results for a simply supported curved panel test structure are presented in support of the model. The curvature is shown to affect substantially the dynamics of the panel, the integration of transducers, and the bandwidth required for structural acoustic control. © *1999 Acoustical Society of America.* [S0001-4966(99)01009-7]

PACS numbers: 43.40.Vn [PJR]

INTRODUCTION

Previous investigations of ASAC for aircraft interior noise applications have concentrated upon flat plates with attached piezoelectric transducers.[1–8] While these investigations showed that ASAC using piezoelectric transducers provides an effective method of achieving reduction in sound radiation into a cavity or a half-space, the structural models used must be refined in order to capture accurately the physics of an actual aircraft panel. Capturing the application physics ensures realistic predictions of closed-loop performance. Specifically, the structural model must account for the effects of aircraft panel curvature on the structural dynamics in order to evaluate transducer performance. A state variable model of a simply supported curved panel with attached piezoelectric transducers, based upon an approach set forth by Hagood,[9] is presented herein.

It should be noted that actual aircraft skin panels are not simply supported curved panels. Sections of aircraft skin are comprised of multi-layered, stringer-stiffened curved panels, presenting complex modeling issues. Still, modeling these curved panels is an evolutionary process and will serve to provide useful information concerning the coupling of induced strain actuators as well as alternative transducers for control of sound radiation.

Research has been conducted on curved panels for many years. Timoshenko, Love, Flugge and others developed the classical theories of shells and discussed the dynamics of curved panels.[10–12] Later, Banks and Tzou researched the additional dynamics introduced by attaching piezoelectric transducers to curved panels.[13–15] Tzou used lamination theory to develop the equations of motion of the curved panel system, while Banks used classical methods. However, neither development provides for simple state variable formulation. The extended Hamilton's principle used by Hagood yielded the first general model of an elastic structure with attached piezoelectric actuators that led directly to state variable formulation.[9] This approach is used to develop a model of a curved piezo-structure.

Analytical results are presented in order to validate the state-space model of the curved piezo-structure. These results show that the curved piezo-structure model demonstrates the same dynamics as a flat piezo-structure when the radius of curvature approaches infinity. Further, the results show that the curved panel model agrees, for any radius of curvature, with previous inquiries into the effects of curvature on natural frequency.[16,17] Also, the structural model presented is shown to agree with a finite element model of the curved panel. Finally, experimental results from a simply supported curved panel test structure are presented. The dynamics of the curved panel test structure show strong agreement with the dynamics predicted by the model, particularly at low frequency.

The model is then used to draw conclusions about the effects of curvature on ASAC system design and transducer integration. Curvature is shown to have a significant effect on the bandwidth of ASAC systems. Further, it raises significant issues for transducer placement and transducer pairing.

I. MODEL DEVELOPMENT

The model development is based upon the generalized modeling of piezoelectric actuator dynamics presented by Hagood *et al.*[9] First, the strain-displacement relations for a curved panel are established. Then, the strain-displacement relations are used to determine the mass and stiffness matrices of both the curved panel and the attached piezoelectric transducers, as well as the electromechanical coupling matrix between the two. Finally, the model is converted to state-space form.

A. Strain-displacement relations

The first step in the model development is to establish the strain-displacement relations for the curved panel. The generalized shell strain-displacement relations, simplified by

[a]Graduate Research Assistant, Student Member ASA.
[b]Associate Professor, Member ASA.

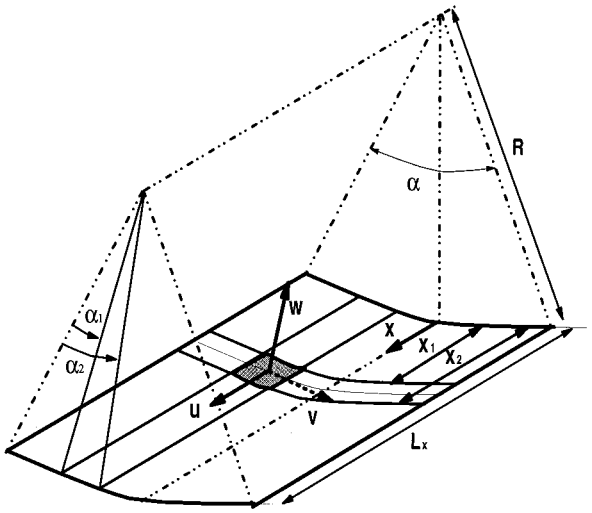


FIG. 1. Coordinate system of curved panel with attached piezoelectric transducer (shaded).

the assumptions of Timoshenko[10] and Love,[11] can be written in terms of the curved panel coordinate system shown in Fig. 1 as

$$\{\boldsymbol{\epsilon}\}=[\mathbf{L_u}]\{\mathbf{u}\}, \tag{1}$$

where the strain vector $\{\boldsymbol{\epsilon}\}$ and the displacement vector $\{\mathbf{u}\}$ are

$$\{\boldsymbol{\epsilon}\}=\{\epsilon_{xx} \quad \epsilon_{\phi\phi} \quad \epsilon_{x\phi}\}^T, \tag{2}$$

$$\{\mathbf{u}\}=\{u \quad v \quad w\}^T, \tag{3}$$

and the linear differential operator $[\mathbf{L_u}]$ for the curved panel is

$$[\mathbf{L_u}]=\begin{bmatrix} \dfrac{\partial}{\partial x} & 0 & 0 \\ 0 & \dfrac{1}{R}\dfrac{\partial}{\partial\phi} & \dfrac{1}{R} \\ \dfrac{1}{R}\dfrac{\partial}{\partial\phi} & \dfrac{\partial}{\partial x} & 0 \end{bmatrix}. \tag{4}$$

The displacements, $\mathbf{u}$, can be expressed in terms of the mid-plane displacements, $\mathbf{u_m}$:

$$\begin{Bmatrix} u \\ v \\ w \end{Bmatrix}=\begin{Bmatrix} u_m - z\dfrac{\partial w_m}{\partial x} \\ v_m + z\left(\dfrac{v_m}{R}-\dfrac{1}{R}\dfrac{\partial w_m}{\partial\phi}\right) \\ w_m \end{Bmatrix}. \tag{5}$$

Further, the strain vector can be expressed in terms of the mid-plane displacements:

$$\{\boldsymbol{\epsilon}\}=[\mathbf{L_{u_m}}]\{\mathbf{u_m}\}, \tag{6}$$

where the linear differential operator $[\mathbf{L_{u_m}}]$ is

$$[\mathbf{L_{u_m}}]=\begin{bmatrix} \dfrac{\partial}{\partial x} & 0 & -z\dfrac{\partial^2}{\partial x^2} \\ 0 & \left(\dfrac{1}{R}+\dfrac{z}{R^2}\right)\dfrac{\partial}{\partial\phi} & \dfrac{1}{R}-\dfrac{z}{R^2}\dfrac{\partial^2}{\partial\phi^2} \\ \dfrac{1}{R}\dfrac{\partial}{\partial\phi} & \left(1+\dfrac{z}{R}\right)\dfrac{\partial}{\partial x} & \dfrac{-2z}{R}\dfrac{\partial^2}{\partial x\partial\phi} \end{bmatrix}. \tag{7}$$

Derivation of these strain-displacement relations is detailed by Timoshenko.[10]

B. Curved piezo-structure dynamics

The strain-displacement relations are used in the electroelastic model to define the equations of motion of the curved panel with attached piezoelectric transducers. The field-potential relations for the piezoelectric material, also needed to define the equations of motion, are given by

$$\mathbf{E}=[\mathbf{L}_{\boldsymbol{\phi}^\mathbf{e}}]\{\boldsymbol{\phi}^\mathbf{e}\}, \tag{8}$$

where $[\mathbf{L}_{\boldsymbol{\phi}^\mathbf{e}}]$ is the electric differential operator,

$$[\mathbf{L}_{\boldsymbol{\phi}^\mathbf{e}}]=\begin{Bmatrix} 0 \\ 0 \\ -\dfrac{\partial}{\partial z} \end{Bmatrix} \tag{9}$$

for the coordinate system shown in Fig. 1.

Having defined the constitutive relations for the curved panel, a Rayleigh–Ritz procedure can be applied to determine the equations of motion. The displacement of the curved panel can be expressed in terms of generalized coordinates:

$$\mathbf{u_m}(x,\phi,t)=\sum_{k=1}^{K} \boldsymbol{\Psi}^\mathbf{d}{}_k(x,\phi)\mathbf{q}_k(t), \tag{10}$$

where the displacement mode shapes $\boldsymbol{\Psi}^\mathbf{d}{}_k$ include two longitudinal modes and one flexural mode satisfying shear-diaphragm boundary conditions, such that

$$\begin{aligned} &w=M_x=N_x=v=0 \quad \text{at } x=0,L_x, \\ &w=M_\phi=N_\phi=u=0 \quad \text{at } \phi=0,\alpha. \end{aligned} \tag{11}$$

Further explanation of shear-diaphragm boundary conditions can be found in Ref. 18. Simple mode shape functions satisfying the boundary conditions are selected:

$$\boldsymbol{\Psi}^{\mathbf{d}}{}_k(x,\phi)=\begin{bmatrix} \sin\dfrac{n\pi\phi}{\alpha}\cos\dfrac{p\pi x}{L_x} & 0 & 0 \\ 0 & \cos\dfrac{n\pi\phi}{\alpha}\sin\dfrac{p\pi x}{L_x} & 0 \\ 0 & 0 & \sin\dfrac{n\pi\phi}{\alpha}\sin\dfrac{p\pi x}{L_x} \end{bmatrix}. \tag{12}$$

Note that the assumption of three mode shapes for each modal index results in three generalized coordinates associated with each mode. Similar to the displacement, the electric potential of the piezoelectric transducer can also be expressed in generalized coordinates:

$$\boldsymbol{\phi}^{\mathbf{e}}(z,t)=\sum_{k=1}^{K}\boldsymbol{\Psi}^{\mathbf{e}}{}_k(z)\mathbf{v}_k(t), \tag{13}$$

where the electric potential mode shapes describe a linearly varying voltage profile across the thickness of the piezoelectric transducer[19] such that

$$\boldsymbol{\Psi}^{\mathbf{e}}{}_k(z)=\begin{bmatrix} z-\dfrac{h_s}{2} \\ \hline h_p \end{bmatrix}. \tag{14}$$

Application of the variational approach of the Rayleigh–Ritz method results in the following equations of motion for the curved panel system with no structural damping:

$$[\mathbf{M_s}+\mathbf{M_p}]\{\ddot{\mathbf{q}}\}+[\mathbf{K_s}+\mathbf{K_p}]\{\mathbf{q}\}=[\boldsymbol{\Theta}]^T\{\mathbf{v}\}+\mathbf{B_f}\{\mathbf{f}\} \tag{15}$$

and in the sensor equation for the attached piezoelectric transducer:

$$\{\boldsymbol{\Theta}\}\{\dot{\mathbf{q}}\}+[\mathbf{C_p}]\{\dot{\mathbf{v}}\}=\{\mathbf{i}\}. \tag{16}$$

The mass matrices for the structure, $\mathbf{M_s}$, and the piezoelectric transducer, $\mathbf{M_p}$ are defined:

$$\mathbf{M_s}=\int_{V_s}\boldsymbol{\Psi}^{\mathbf{d}T}{}_k\rho_s\boldsymbol{\Psi}^{\mathbf{d}}{}_k dV_s, \tag{17}$$

$$\mathbf{M_p}=\int_{V_p}\boldsymbol{\Psi}^{\mathbf{d}T}{}_k\rho_p\boldsymbol{\Psi}^{\mathbf{d}}{}_k dV_p. \tag{18}$$

The stiffness matrices for the structure, $\mathbf{K_s}$, and the piezoelectric transducer, $\mathbf{K_p}$ are defined:

$$\mathbf{K_s}=\int_{V_s}[\mathbf{L_{u_m}}\boldsymbol{\Psi}^{\mathbf{d}}{}_k]^T\mathbf{c_s}[\mathbf{L_{u_m}}\boldsymbol{\Psi}^{\mathbf{d}}{}_k]dV_s, \tag{19}$$

$$\mathbf{K_p}=\int_{V_p}[\mathbf{R_S}\mathbf{L_{u_m}}\boldsymbol{\Psi}^{\mathbf{d}}{}_k]^T\mathbf{c}^{\mathbf{E}}[\mathbf{R_S}\mathbf{L_{u_m}}\boldsymbol{\Psi}^{\mathbf{d}}{}_k]dV_p, \tag{20}$$

where $\mathbf{c_s}$ is the structural material stiffness matrix:

$$\mathbf{c_s}=\begin{bmatrix} \dfrac{E_s}{1-\nu_s^2} & \dfrac{E_s\nu_s}{1-\nu_s^2} & 0 \\ \dfrac{E_s\nu_s}{1-\nu_s^2} & \dfrac{E_s}{1-\nu_s^2} & 0 \\ 0 & 0 & \dfrac{E_s}{2(1+\nu_s)} \end{bmatrix}. \tag{21}$$

$\mathbf{R_S}$ is the strain rotation matrix as defined by Cook:[20]

$$\mathbf{R_S}=\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \tag{22}$$

and $\mathbf{c}^{\mathbf{E}}$ is the piezoelectric material stiffness matrix:

$$\mathbf{c}^{\mathbf{E}}=\begin{bmatrix} \dfrac{E_p}{1-\nu_p^2} & \dfrac{E_p\nu_p}{1-\nu_p^2} & 0 \\ \dfrac{E_p\nu_p}{1-\nu_p^2} & \dfrac{E_p}{1-\nu_p^2} & 0 \\ 0 & 0 & \dfrac{E_p}{2(1+\nu_p)} \end{bmatrix}. \tag{23}$$

The electromechanical coupling matrix, $\boldsymbol{\Theta}$, is given by:

$$\boldsymbol{\Theta}=\int_{V_p}[\mathbf{R_S}\mathbf{L_{u_m}}\boldsymbol{\Psi}^{\mathbf{d}}{}_k]^T\mathbf{e}^T[\mathbf{R_E}\mathbf{L}_{\boldsymbol{\phi}^{\mathbf{e}}}\boldsymbol{\Psi}^{\mathbf{e}}{}_k]dV_p, \tag{24}$$

where $\mathbf{R_E}$ is the electric field rotation matrix for the coordinate system shown in Fig. 1:

$$\mathbf{R_E}=\begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \tag{25}$$

and $\mathbf{e}$ is a matrix of piezoelectric material constants given by:

$$\mathbf{e}=\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \dfrac{d_{31}E_p}{1-\nu_p} & \dfrac{d_{31}E_p}{1-\nu_p} & 0 \end{bmatrix}. \tag{26}$$

The capacitance matrix, $\mathbf{C_p}$, is:

$$\mathbf{C_p}=\int_{V_p}[\mathbf{R_E}\mathbf{L}_{\boldsymbol{\phi}^{\mathbf{e}}}\boldsymbol{\Psi}^{\mathbf{e}}{}_k]^T\boldsymbol{\varepsilon}^{\mathbf{S}}[\mathbf{R_E}\mathbf{L}_{\boldsymbol{\phi}^{\mathbf{e}}}\boldsymbol{\Psi}^{\mathbf{e}}{}_k]dV_p, \tag{27}$$

where $\boldsymbol{\varepsilon}^{\mathbf{S}}$ is a matrix of dielectric constants measured at constant strain:

$$\boldsymbol{\varepsilon}^{\mathbf{S}}=\begin{bmatrix}\varepsilon_1^S & 0 & 0\\ 0 & \varepsilon_1^S & 0\\ 0 & 0 & \varepsilon_1^S\end{bmatrix}. \tag{28}$$

For a more detailed derivation of the previous terms, refer to Refs. 9 and 19.

C. State variable formulation

The equations of motion of the curved panel and attached piezoelectric transducers can be cast in state variable form:

$$\dot{\mathbf{x}}=\mathbf{A}\mathbf{x}+\mathbf{B}\mathbf{u}, \tag{29}$$

$$\mathbf{y}=\mathbf{C}\mathbf{x}+\mathbf{D}\mathbf{u}, \tag{30}$$

where

$$\mathbf{A}=\begin{bmatrix}\mathbf{0} & \mathbf{I}\\ -[\mathbf{M_s}+\mathbf{M_p}]^{-1}[\mathbf{K_s}+\mathbf{K_p}] & \mathbf{0}\end{bmatrix}, \tag{31}$$

$$\mathbf{x}^T=[\mathbf{q}^T \quad \dot{\mathbf{q}}^T], \tag{32}$$

$$\mathbf{B}=\begin{bmatrix}\mathbf{0} & \mathbf{0}\\ [\mathbf{M_s}+\mathbf{M_p}]^{-1}\boldsymbol{\Theta} & [\mathbf{M_s}+\mathbf{M_p}]^{-1}\mathbf{B_f}\end{bmatrix}, \tag{33}$$

$$\mathbf{u}^T=[\mathbf{v}^T \quad \mathbf{f}^T], \tag{34}$$

$$\mathbf{C}=[\mathbf{0} \quad \boldsymbol{\Theta}^T], \tag{35}$$

$$\mathbf{D}=[\mathbf{Cp} \quad \mathbf{0}]. \tag{36}$$

Note that the state variable form above observes the rate of change of the piezoelectric charge, which is proportional to strain rate.

II. RESULTS

The state variable model of the curved panel with attached piezoelectric sensoriactuators can be validated by comparison to previous work. The state variable model can be used to generate the frequency response function (FRF) for the curved panel for various input/output paths. Further, the model can be used to obtain the natural frequencies of the curved panel for a given radius of curvature which can be compared to previous investigations into the natural frequencies of curved panels, such as that by Szechenyi.[16] Also, the dynamics of the curved panel can be compared with those of a plate by setting the radius of curvature to infinity and comparing the FRFs. Further, the dynamics predicted by the model can be compared to those measured from a curved panel test structure. Once the model has been validated, discussion can ensue on the applicability of piezoelectric transducers in the control of sound radiation from curved panels.

A. Model validation

The state variable model of the curved panel was implemented in MATLAB. The following parameters were selected to match the experimental test rig and are held fixed throughout the investigation: $L_x=0.632$ m, $h_s=1.3$ mm, $h_p=0.2$ mm, $x_1=0.250L_x$, $x_2=0.337L_x$, $\alpha_1=0.166\alpha$, $\alpha_2=0.265\alpha$, $E_s=69.3\times10^9$ N/m^2, $\rho_s=2680$ kg/m^3, $\nu_s=0.33$, $E_p=5.0\times10^{10}$ N/m^2, $\rho_p=7650$ kg/m^3, $\nu_p=0.3$, $d_{31}=-180\times10^{-12}$ m/V. The radius of curvature of the panel, R, is varied between 2 m and 1000 m. The sweep angle, α, varies with the radius of curvature such that $R\alpha=0.556$ m. Thus the actual structure size does not change with the variation of the radius of curvature. The structural material parameters are for 5052 aluminum and the piezoelectric parameters are for PSI-5A-S3.[8] In addition, $K=20$, resulting in the inclusion of 20 curved panel modes. Note that 60 curved panel mode shapes exist since 2 longitudinal mode shapes and 1 flexural mode shape are assumed for each mode included in the model. Piezoelectric actuator pairs, with one patch located on each side of the panel, are assumed to be wired with opposing polarities, doubling the applied moment.

TABLE I. Comparison of natural frequencies calculated from model and Eq. (37).

Method	Natural frequencies (Hz)					
Eq. (37)	81.2	103.3	159.6	169.6	177.4	190.3
$eig(M^{-1}K)$	81.1	103.2	159.5	169.5	176.8	190.1

The first step in model validation is to compare the natural frequencies of the curved panel structure, obtained from the eigenvalues of $M^{-1}K$ for a given radius of curvature, to those obtained from a formula for the natural frequencies of a simply supported curved panel:[16,17]

$$\omega_{cp}^2=\frac{E_s h_s^2}{12\rho_s(1-\nu_s^2)}\left[\left(\frac{p\pi}{L_x}\right)^2+\left(\frac{n\pi}{R\alpha}\right)^2\right]^2 + \frac{E_s}{\rho_s(1-\nu_s^2)R^2}\left[\frac{\left(\frac{p\pi}{L_x}\right)^4(1-\nu_s^2)}{\left[\left(\frac{p\pi}{L_x}\right)^2+\left(\frac{n\pi}{R\alpha}\right)^2\right]^2}\right]. \tag{37}$$

Table I compares the natural frequencies obtained from the eigenvalues of $M^{-1}K$ to those obtained from Eq. (37). A radius of curvature of 2 m is assumed. As Table I shows, the natural frequencies found from the curved panel model correspond with those calculated from Eq. (37).

The curved piezo-structure model can be validated through comparison with the flat piezo-structure model developed by Frampton *et al.*,[7] by allowing the radius of curvature of the curved panel to approach infinity. Figures 2–4 show the reduction of the curved piezo-structure model to the flat piezo-structure model as the radius of curvature approaches infinity. Figure 2, with a radius of curvature of 20 m, shows substantial curvature effects, particularly at the lower frequencies. As the radius of curvature is increased to 50 m, as shown in Fig. 3, the effects of the curvature decrease and the curved piezo-structure model begins to approach the flat piezo-structure model. When the radius of curvature is increased to 1000 m, as shown in Fig. 4, the curved piezo-structure model and the flat piezo-structure model coincide for all practical purposes.

Finally, the curved piezo-structure model was compared with a finite element model. In the finite element model, the simply supported boundary conditions were modeled using shear diaphragm assumptions.[18] Since the finite element

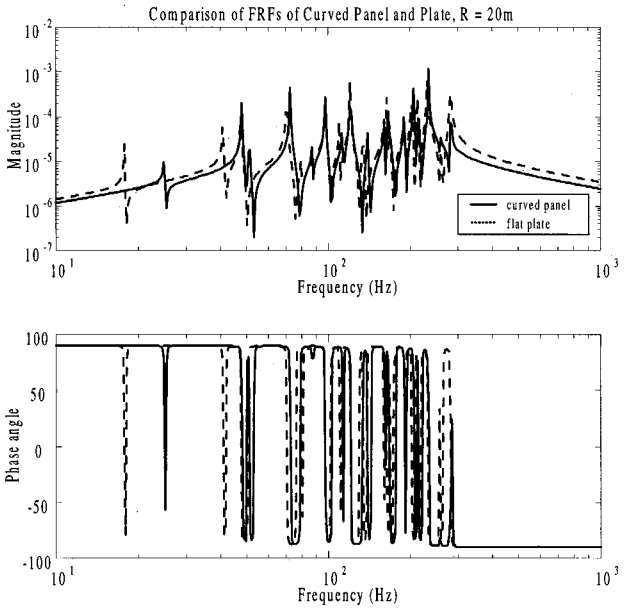


FIG. 2. FRF of curved panel model compared with FRF of flat plate model, $R=20$ m.

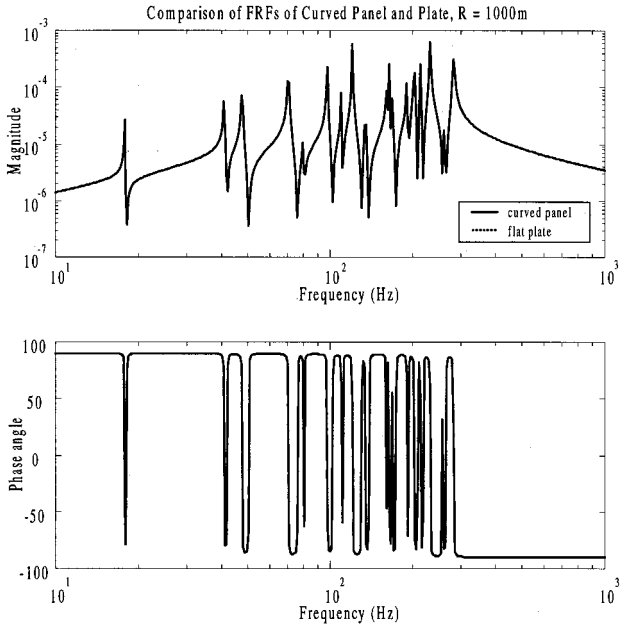


FIG. 4. FRF of curved panel model compared with FRF of flat plate model, $R=1000$ m.

model did not allow simple integration of the piezoelectric transducers, only the structural models were compared using a transfer function between a point displacement sensor located at $x=0.312L_x$, $\phi=0.712\alpha$ and a point force actuator located at $x=0.646L_x$, $\phi=0.379\alpha$. Panel dimensions and properties were consistent with those described previously, and the radius of curvature was 2.1 m. As Fig. 5 shows, there is perfect agreement between the structural model and the finite element model.

B. Experimental verification

A simply supported curved panel test structure was constructed using a frame with a radius of curvature of 2 m as shown in Fig. 6. A thin flat aluminum plate was clamped into the frame using clamping bars with curvature matching the frame. The simply supported boundary conditions were created by milling a groove in the flat plate at the clamped boundary as suggested in Ref. 21. The plate dimensions and material properties were consistent with those presented in Sec. II A. The dimensions were chosen to simplify the test rig construction and are not representative of actual fuselage panels. Piezoceramic patches, dimensioned 5.5 cm $\times$5.5 cm$\times$0.2 mm, were attached at corresponding locations on each side of the panel using M-Bond 200 strain gauge glue and excited with random noise. A VPI Laser Vibrometer was used to measure point velocities at various points on

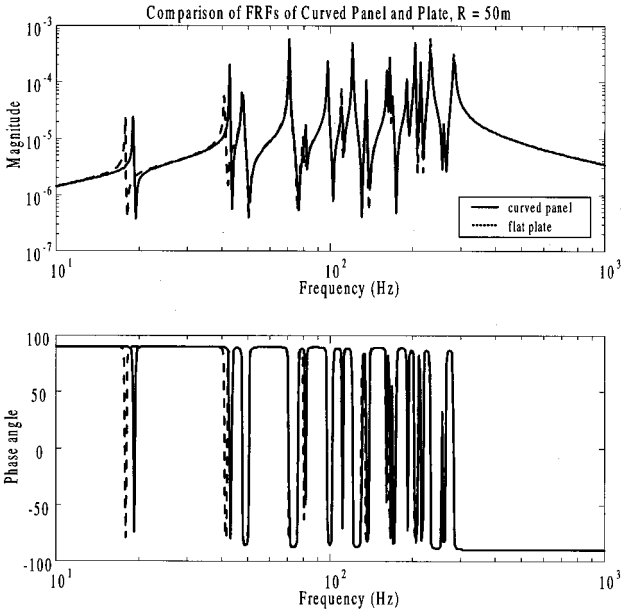


FIG. 3. FRF of curved panel model compared with FRF of flat plate model, $R=50$ m.

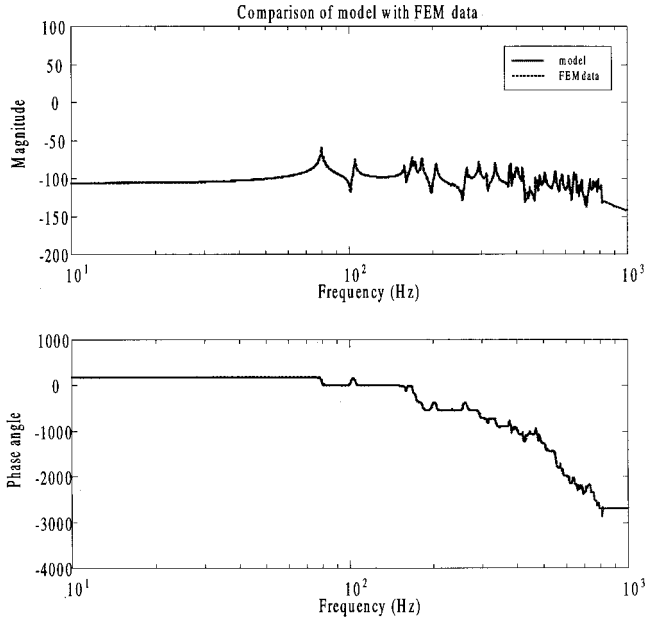


FIG. 5. Comparison of analytical model and finite element model, $R=2.1$ m.

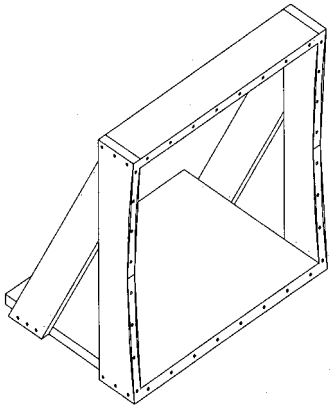


FIG. 6. Curved panel test rig frame.

the panel, and a digital spectrum analyzer was used to measure the transfer function between the piezoelectric patch inputs and the point velocity output. The model was altered to predict transfer functions between the piezoelectric patch input and the point velocity output, and the results were compared with the experimental data.

Figure 7 shows the transfer function between a piezoelectric patch input centered at $x=0.294L_x$, $\phi=0.215\alpha$ and a point velocity output at $x=0.615L_x$, $\phi=0.629\alpha$. Below 200 Hz, the model accurately captures the panel dynamics. Above 200 Hz, the model approximates the response, but error is present. This error is assumed to be the result of boundary condition inconsistencies between the model and the physical structure, but is as yet unresolved. However, the current model will suffice to investigate the effects of curvature on transducer design and integration for ASAC.

C. Effects of curvature

Figure 8 shows the FRFs of the plate and the curved panel for the first six modes of the flat plate, the (1,1), (1,2),

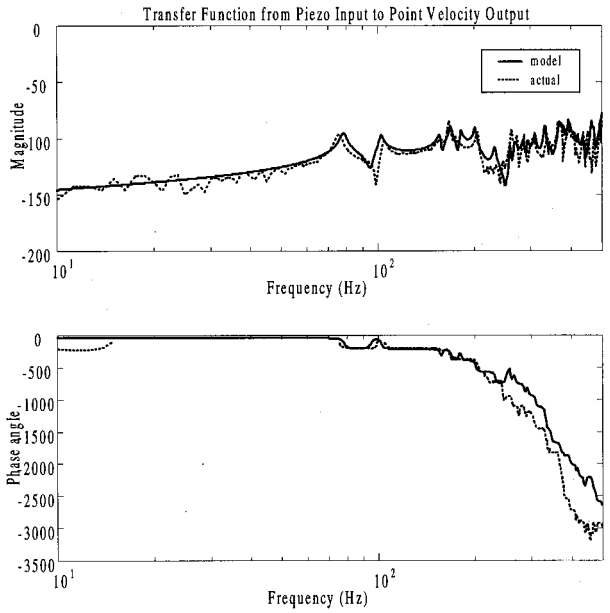


FIG. 7. Comparison of experimental data and analytical model.

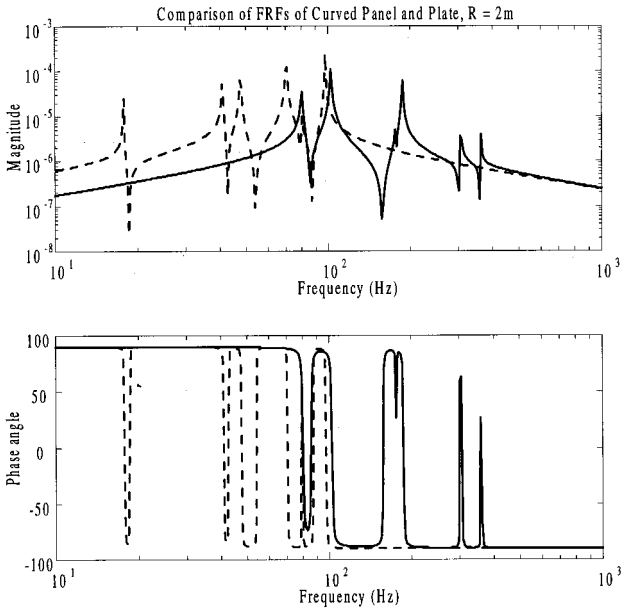


FIG. 8. FRF of curved panel model compared with FRF of flat plate model, $R=2$ m.

(2,1), (2,2), (1,3) and (3,1) modes. The effects of the curvature are easily seen in the increased natural frequencies of the modes shown. Curvature increases the natural frequencies of a panel by increasing the stiffness of the panel. What cannot be seen in Fig. 8 is that the curvature affects certain modes more than others. Table II shows the natural frequencies associated with the six modes included in the model for different radii of curvature. As the radius of curvature decreases, the (1,1) mode natural frequency increases substantially, as do the (2,1), (2,2), and (3,1) modes natural frequencies. However, the (1,2) mode natural frequency only slightly increases and the (1,3) mode natural frequency is barely affected by the curvature changes. As a result, the dominant radiation mode, (1,1), is no longer the lowest frequency mode of the curved panel when the radius of curvature becomes small enough, such as when $R=5$ m or $R=2$ m. In fact, when $R=2$ m, the (1,2) mode and the (1,3) mode both have a lower natural frequency than the (1,1) mode. This phenomenon can be explained in terms of wave theory. Waves traveling primarily circumferentially along the panel, in the ϕ-direction, such as those associated with the (1,2) and (1,3) modes, are traveling with the curvature and are barely affected by it. Thus, the natural frequencies associated with these modes are relatively unaffected. Waves traveling

TABLE II. Natural frequencies (Hz) of panel for varying radius of curvature.

	Panel radius of curvature (m)						
Mode	∞	100	50	20	10	5	2
1,1	18	18	19	25	40	73	177
2,1	41	42	42	51	74	129	308
1,2	48	48	48	48	50	55	81
2,2	71	71	72	73	80	100	190
3,1	80	80	81	87	107	163	363
1,3	98	98	98	98	98	99	103

TABLE III. Dimensions of panels with varying aspect ratio.

Aspect ratio	L_x (m)	$R\alpha$ (m)
2	0.632	0.316
1.5	0.632	0.421
1.14	0.632	0.556
1	0.632	0.632
0.66	0.421	0.632
0.5	0.316	0.632

against the curvature of the panel, in the x-direction, such as those associated with the (1,1), (2,1), (2,2) and (3,1) modes, encounter increased panel stiffness, due to the curvature, resulting in increased natural frequencies.

Consideration must also be given to the significance of curvature in panels of other aspect ratios. Table III presents the dimensions of panels with varying aspect ratio in the axial (L_x) and circumferential ($R\alpha$) directions. Table IV shows the effects of varying aspect ratio on the significance of curvature to natural frequency. The percent increase in natural frequency between a flat panel, $R=\infty$, and a curved panel with $R=2$ m is calculated from:

$$\%\text{increase} = \frac{\omega_{R=2} - \omega_{R=\infty}}{\omega_{R=\infty}}. \tag{38}$$

The data presented indicate that the increase in natural frequency is more significant in panels with aspect ratio closer to 1. However, significant curvature effects are still evident in each case.

D. Impact on active structural acoustic control

The results of curvature raise a number of issues for ASAC. The curvature has substantially increased the natural frequency of the (1,1) mode, which is the most efficient sound radiator.[22] This increase in natural frequency moves the radiation of the (1,1) mode into a more audible frequency band. Furthermore, the curvature shifts the neutral axis of the curved panel toward the inner surface as shown in Fig. 9. Piezoelectric transducer pairs function as actuators by inducing opposite surface strains along a moment arm the distance of the separation of the patch from the neutral axis. In a flat piezo-structure, the neutral axis is the mid-plane of the plate. Thus the moment arm is the same for each transducer in the pair and the surface placement on either the inner or the outer surface is irrelevant to transducer performance. However, since curvature shifts the neutral axis toward the inner surface, transducer performance is dependent upon surface placement in a curved piezo-structure. Figure 10 compares the performance of a single piezoelectric transducer placed on the inner surface with the performance of one attached to the outer surface. The predicted transfer functions are derived assuming that the transducer is both the sensor and the actuator and is arbitrarily placed on the panel. The transducer attached to the outer surface more strongly couples to the flexural modes of panel than the transducer on the inner surface, as can be seen by the difference in the response at the lower frequencies. The response seen at higher frequencies, 750 Hz and up, is the result of the excitation of the longitudinal modes of the panel by the single transducer. Figure 11 shows the effects of relative polarity of transducer pairs on performance. If a transducer pair is wired with the same polarity, the individual transducers operate in-phase and mainly excite longitudinal modes. However, as a result of the stronger performance of the outer surface transducer, some flexural excitation results. When the transducer pair is wired with opposing polarity, the transducers act primarily as a moment actuator, exciting mostly the flexural modes. However, excitation of the longitudinal modes is still evident as is indicated prevalently in the phase plots of Fig. 11. The excitation of longitudinal modes warrants attention in ASAC de-

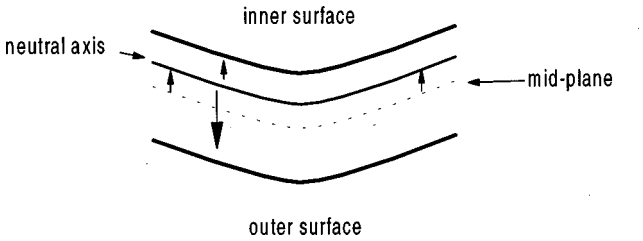


FIG. 9. Shift of neutral axis as a result of curvature.

TABLE IV. Effects of varying aspect ratio on the significance of curvature. Values given represent % increase in natural frequency from $R=\infty$ to $R=2$ m.

Mode No.	Aspect ratio					
	2	1.5	1.14	1	0.66	0.5
1,1	130%	399%	873%	1159%	985%	727%
2,1	239%	439%	641%	727%	377%	204%
1,2	1.5%	12.6%	69%	130%	215%	239%
3,1	194%	282%	352%	377%	154%	69%
1,3	0.3%	0.5%	5%	13%	36%	59%

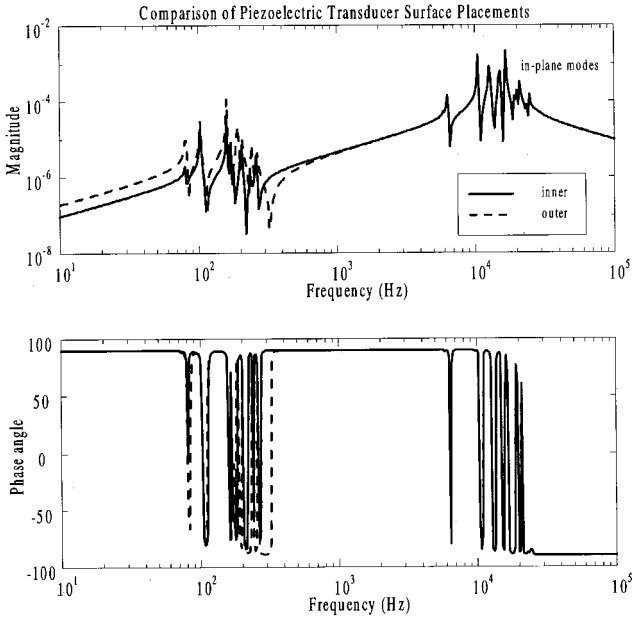


FIG. 10. Comparison of piezoelectric transducer surface placements.

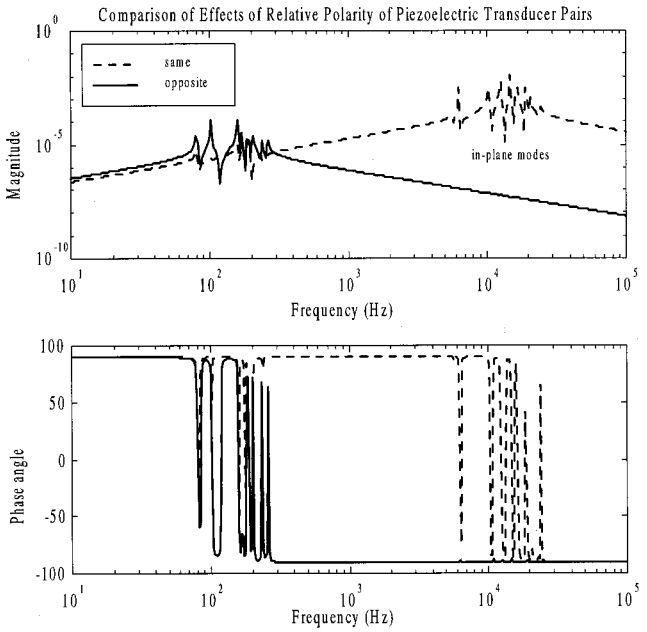


FIG. 11. Comparison of effects of relative polarity of piezoelectric transducer pairs.

sign since these higher frequency modes can produce instability if the control system does not have the proper ''roll-off'' in frequency response.

III. CONCLUSIONS

A state variable model of a simply supported curved piezo-structure was presented. This model allows the extension of ASAC research on flat plates to the curved panels which compose the skin of modern aircraft. The state variable model was validated with previous analytical work. Furthermore, the model was experimentally verified. Curvature was shown to have substantial effects on the structural dynamics of a panel, resulting in design considerations for ASAC system bandwidth and transducer integration. Increased panel stiffness as a result of curvature was shown to affect certain panel modes more than others, resulting in modal reordering. The modal reordering affects the bandwidth of structural control aimed at reducing sound radiation, particularly since the dominant radiation mode is no longer the lowest in frequency. Also, curvature shifts the neutral axis toward the inner surface of the panel, making the relative polarity of transducers within a pair important to control system design. Wiring the transducers with opposing polarity does not prevent the excitation of longitudinal modes, as in the case of a flat piezo-structure. Furthermore, wiring the transducers with the same polarity results in the excitation of flexural modes. Future work will incorporate pressure loading into the structural model, as well as investigate the coupling of a curved panel to the acoustics of a cylindrical enclosure.

ACKNOWLEDGMENTS

This work was supported by NASA Langley Research Center under Grant No. NCC 1 250 and the direction of Dr. Richard J. Silcox. The authors gratefully acknowledge the assistance of Travis Turner of NASA Langley Structural Acoustics Branch in generating the finite element model results. The authors also extend gratitude to Duke University undergraduate J.D. Kemp for his help in designing the test rig.

[1] Yi Gu, R. L. Clark, C. R. Fuller, and A. C. Zander, ''Experiments on active control of plate vibration using piezoelectric actuators and polyvinylidene fluoride modal sensors,'' Trans. ASME, J. Vib. Acoust. **116**, 303–308 (1994).
[2] R. L. Clark and C. R. Fuller, ''Optimal placement of piezoelectric actuators and polyvinylidene fluoride (pvdf) error sensors in active structural acoustic control approaches,'' J. Acoust. Soc. Am. **92**, 1521–1533 (1992).
[3] R. L. Clark and C. R. Fuller, ''Active structural acoustic control with adaptive structures and wave number considerations,'' J. Intell. Mater. Syst. Struct. **3**, 296–315 (1992).
[4] R. L. Clark and C. R. Fuller, ''Experiments on active control of structurally radiated sound using multiple piezoceramic actuators,'' J. Acoust. Soc. Am. **91**, 3313–3320 (1992).
[5] R. L. Clark and C. R. Fuller, ''Control of sound radiation with adaptive structures,'' J. Intell. Mater. Syst. Struct. **2**, 431–452 (1991).
[6] J. S. Vipperman and R. L. Clark, ''Multivariable feedback active structural acoustic control with adaptive piezoelectric sensoriactuators,'' J. Acoust. Soc. Am. **105**, 219–225 (1999).
[7] K. D. Frampton, R. L. Clark, and E. H. Dowell, ''Active control of panel flutter with piezoelectric transducers,'' J. Aircr. **33**, 768–774 (1996).
[8] J. S. Vipperman, ''Adaptive Piezoelectric Sensoriactuators for Active Structural Acoustic Control,'' Ph.D. thesis, Duke University, 1996.
[9] N. W. Hagood, W. H. Chung, and A. von Flotow, ''Modeling of piezoelectric actuator dynamics for active structural control,'' J. Intell. Mater. Syst. Struct. **1**, 327–354 (1990).
[10] S. Timoshenko, *Theory of Plates and Shells* (McGraw-Hill, New York, 1959).
[11] A. E. H. Love, *A Treatise on the Mathematical Theory of Elasticity* (Dover, New York, 1944).
[12] Wilhelm Flugge, *Stresses in Shells* (Springer-Verlag, New York, 1966).
[13] H. Banks, R. Smith, and Y. Wang, *Smart Material Structures: Modeling, Estimation and Control* (Wiley, New York, 1996).
[14] H. S. Tzou and Y. Bou, ''Parametric study of segmented transducers laminated on cylindrical shells, part 1: Sensor patches,'' J. Sound Vib. **197**, 207–224 (1996).
[15] H. S. Tzou and Y. Bou, ''Parametric study of segmented transducers laminated on cylindrical shells, part 2: Actuator patches,'' J. Sound Vib. **197**, 225–249 (1996).
[16] E. Szechenyi, ''Approximate methods for the determination of the natural frequencies of stiffened and curved plates,'' J. Sound Vib. **14**, 401–418 (1971).
[17] R. D. Blevins, *Formulas for Natural Frequency and Mode Shape* (Robert E. Krieger, Malabar, FL, 1984).
[18] A. Leissa, *Vibration of Shells* (Acoustical Society of America, Woodbury, NY, 1993).
[19] R. L. Clark, G. P. Gibbs, and W. R. Saunders, *Adaptive Structures, Dynamics and Control* (Wiley, New York, 1998).
[20] R. D. Cook, *Concepts and Applications of Finite Element Analysis*, 2nd ed. (Wiley, New York, 1971).
[21] W. Hoppman and J. Greenspon, ''An experimental device for obtaining elastic rotational constraints on the boundary of a plate,'' in *Proceedings of the 2nd National Congress on Applied Mechanics* (American Society of Mechanical Engineers, New York, 1954), pp. 187–191.
[22] C. E. Wallace, ''Radiation resistance of a rectangular panel,'' J. Acoust. Soc. Am. **51**, 946–952 (1972).

Field study of the annoyance of low-frequency runway sideline noise

Sanford Fidell,[a] Laura Silvati, Karl Pearsons, Stephen Lind, and Richard Howe
BBN Technologies, A Unit of GTE Internetworking, 21128 Vanowen Street, Canoga Park, California 91303

(Received 11 May 1998; revised 25 February 1999; accepted 24 May 1999)

Noise from aircraft ground operations often reaches residences in the vicinity of airports via grazing incidence paths that attenuate high-frequency noise more than air-to-ground propagation paths, thus increasing the relative low-frequency content of such noise with respect to overflight noise. Outdoor A-weighted noise measurements may not appropriately reflect low-frequency noise levels that can induce potentially annoying secondary emissions inside residences near runways. Contours of low-frequency noise levels were estimated in a residential area adjacent to a busy runway from multi-site measurements of composite maximum spectra of runway sideline noise in the one-third octave bands between 25 and 80 Hz, inclusive. Neighborhood residents were interviewed to determine the prevalence of annoyance attributable to runway sideline noise at frequencies below 100 Hz, and of its audible manifestations inside homes. Survey respondents highly annoyed by rattle and vibration were concentrated in areas with low-frequency sound levels due to aircraft operations in excess of 75 to 80 dB. [S0001-4966(99)01909-8]

PACS numbers: 43.50.Lj, 43.50.Qp, 43.50.Sr [MRS]

INTRODUCTION

Studies of community response to aircraft noise have generally focused on effects of overflights of neighborhoods near runway ends (cf. those reviewed by Finegold *et al.*, 1994 and by Fidell *et al.*, 1991), both because of the high A-weighted sound levels produced by direct flyovers, and because of the sizable residential populations exposed to aircraft approach and departure noise. Quantitative relationships between noise exposure and the prevalence of noise-induced annoyance in communities are characterized in A-weighted units which are little affected by energy at frequencies below about 100 Hz. As the proportion of quieter transport aircraft in the commercial fleet has increased in recent years, interest in noise created by aircraft ground operations at large civil airport has grown. These concerns include effects of noise produced behind departing aircraft (sometimes termed ''backblast''), engine run-ups for maintenance purposes, and runway sideline noise (taxiing, queuing, acceleration during takeoff, and thrust reverser application on landing).

Because such aircraft noise often reaches communities by overground rather than air-to-ground paths, it can contain proportionately less high-frequency energy than overflight noise, due to ''excess'' attenuation from ground effects (Piercy and Embleton, 1977; Sutherland and Daigle, 1997) beyond that attributable to atmospheric absorption. Aircraft ground operation noise is characteristically described in complaints as a dull rumbling sound with a slow onset time. Aircraft ground noise may be distinctively audible at considerable distances from airports, particularly at night, when less masked by other urban noise sources (cf. Fidell *et al.*, 1981).

Questions about the utility of representing low-frequency aircraft ground noise in A-weighted units, and about the adequacy of standard interpretive criteria for assessing community response to low-frequency noises, are becoming increasingly common. However, much of what is known about the annoyance of low-frequency noise has been derived from studies of indoor noise sources, or in nonresidential settings (Broner, 1978; Berglund *et al.*, 1996). Although Berglund *et al.* cite several studies as demonstrating greater annoyance for sounds with greater low-frequency noise than for sounds of equivalent loudness but less low-frequency energy, the circumstances of noise exposure in these studies tend to differ from those in residential areas.

One aspect of particular concern with respect of low-frequency noise effects is the annoyance of secondary emissions (rattling sounds of household paraphernalia) that may be induced inside residences. Measurements of the low-frequency noise produced by aircraft ground operations (Lind *et al.*, 1997) indicate that sufficient low-frequency energy may sometimes be produced to induce secondary emissions in nearby residences, as described by Hubbard (1982). Noise descriptors useful for predicting rattle focus on maximum (rather than average) sound levels in particular low-frequency bands (rather than frequency-weighted, broadband levels).

This field study of the annoyance of low-frequency runway sideline noise sought to determine the prevalence of annoyance associated with aircraft-induced rattle and vibration by means of direct questioning of residents of a community that has long experienced runway sideline noise.

I. METHOD

A. Simultaneous multi-point outdoor measurements

Outdoor ambient noise levels due to aircraft activity were measured at several simultaneous combinations of seven positions in El Segundo, CA, south of runway 25L at Los Angeles International Airport (LAX), as shown in Fig.

[a]Electronic mail: fidell@bbn.com

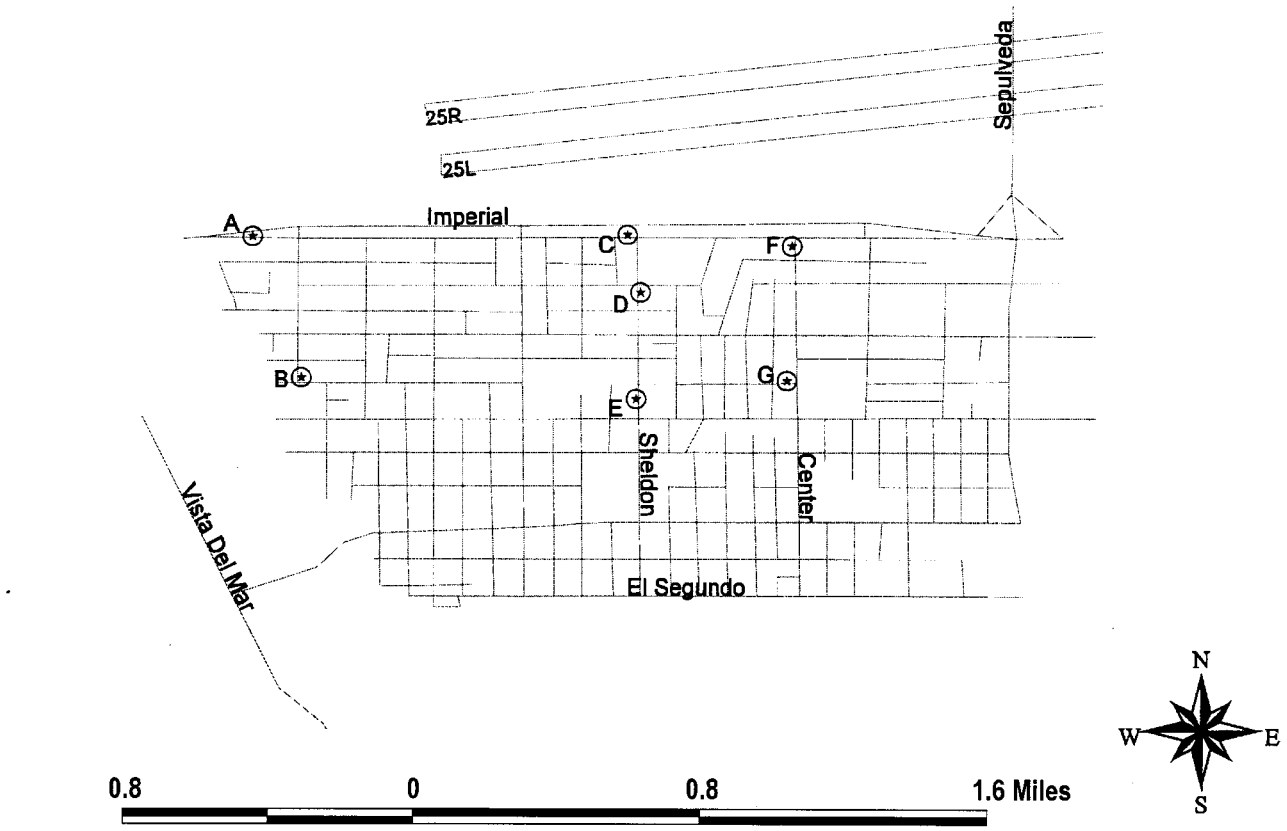


FIG. 1. Locations of monitoring positions in relationship to LAX runways 25R and 25L.

1. Microphones were tripod mounted at a height of approximately 1.7 m above the ground plane. Ground cover between the measurement positions and the runways varied from pavement to grass, as did ground surfaces in the immediate vicinity of measurement locations. Runway 25R was in use for departures and 25L for arrivals during the measurement period. Time-synchronized, wideband recordings of aircraft noise were made with digital tape recorders at six of the seven locations. At the seventh location (C), only the A-weighted level and low-frequency composite levels were measured, with a real-time spectrum analyzer and sound level meter.

B. Population of interest

The target population was adult English-speaking residents of an area of El Segundo between 1000 to 5000 ft south of the southern runway complex at LAX. ''Sampling'' was exhaustive: attempts were made to conduct an interview with one adult member of each of the 1262 eligible households with listed telephone numbers found in this area. Since Fields (1993) has shown that demographic variables such as age, sex, social status, income, education, home ownership, dwelling type, and length of residence have no documentable effect on noise-induced annoyance, interviewers were instructed to conduct the interview with any adult, verified household member. Residential construction in the interviewing area included a mixture of single-family detached dwellings and higher density, two-and three-story buildings.

C. Interviewing procedures

A structured questionnaire with thirteen closed response category items and 8–18 information requests (contingent on responses to branching items) was administered by telephone. The Appendix contains the questionnaire items. Many of these items had been posed to residents of airport neighborhoods elsewhere.

The interview was introduced as a study of neighborhood living conditions. The first three questionnaire items were preliminary questions about duration of residence and about the most and least favored aspects of neighborhood living conditions. These were intended to confirm the eligibility of respondents for interview, and to provide an opportunity for spontaneous mention of aircraft noise as the least favored aspect of neighborhood living conditions. The first explicit mention of noise occurred in item 4 (''Would you say that your neighborhood is quiet or noisy?''). The intent of this item was to solicit an overall assessment of neighborhood noisiness prior to any mention of aircraft noise in particular.

Item 5 inquired about annoyance due to street traffic noise. This item provided a context for subsequent questions regarding annoyance due to aircraft noise in the next six items:

(i) Item 6 inquired about annoyance due to aircraft noise in general.
(ii) Item 7 asked whether airplanes produced vibration and rattling sounds in respondents' homes.

(iii) Item 8 asked for a category scale rating of annoyance due to such vibration and rattle.
(iv) Item 9 sought information about frequency of notice of vibration and rattle.
(v) Item 10 asked for the identity of vibrating and rattling objects.
(vi) Item 11 sought information about actions taken to lessen vibration or rattling noises.

Respondents were constrained to reply to questions about intensity of annoyance by selecting one of the following response categories: ''not at all annoyed,'' ''slightly annoyed,'' ''moderately annoyed,'' ''very annoyed,'' and ''extremely annoyed.'' No time frame (e.g., last week, last year, etc.) was specified for these items, because maximum low-frequency aircraft ground noise levels in the interviewing area were believed to have varied little within the last several years; because it was considered counterproductive for the purposes of the present study to draw respondents' attention to particular historical periods; and because a response based on respondents' general long-term experiences was preferred to a response based on any particular recent instances of vibration or rattle. Items 12 and 13 asked about complaints concerning vibration, rattling, and aircraft noise in general.

Interviewing was conducted under central supervision by computer-assisted means. Software automatically selected telephone numbers for dialing, stored responses to questionnaire items, time-stamped interviews, and scheduled callbacks. A dozen interviewers were trained to conduct the interview in accordance with written instructions reviewed during briefing sessions prior to the start of interviewing. Nine contact attempts (an initial attempt followed by as many as eight callbacks as needed) were budgeted for each eligible household in an effort to exhaustively sample household opinion throughout the interviewing area.

II. RESULTS

A. Summary of low-frequency aircraft noise measurements

Most of the low-frequency noise events measured at each microphone position were of similar origin. The microphone positions closest to the runway threshold were more influenced by the high-power settings characteristic of the start of takeoffroll of aircraft departing on runway 25R, and less influenced by the lower noise levels created by aircraft landing on runway 25L. Microphone positions nearer to the center of the runway pair were influenced both by the high-power settings of departing aircraft accelerating past them, and by thrust reverser applications by landing aircraft. Microphone positions closer to the departure end of the runways were most influenced by noise of near-ground but airborne aircraft. All aircraft movements, including the near-ground flight path segments of arrival and departure operations, were considered aircraft noise events.

Data reduction procedures were modeled on those of Part 36 of the U.S. Federal Aviation Regulations. Half-second time series of sound levels in one-third octave bands centered at frequencies between 25 and 10 kHz were derived from the field recordings by means of a software-controlled Brüel and Kjær 2134 sound intensity analyzer. A single-event, low-frequency sound level (LFSL) descriptor was defined in preference to a cumulative or average metric to represent low-frequency aircraft noise, on the grounds that secondary emissions audible within structures are caused by instantaneous excitation, not by long-term average levels. Such a maximum band level descriptor is intentionally insensitive to noise event duration since its intended use is as a predictor of the simple occurence of rattle.

TABLE I. Average A-weighted and low-frequency aircraft noise at measurement locations.

Measurement locations	Average of maximum A-weighted aircraft noise events (dB)	Average ''low-frequency'' content of aircraft noise events[a] (dB)
A	86.0	88.2
B	76.2	72.3
C (first time period)	83.4	90.9
C (second time period)	84.1	91.7
C (third time period)	84.5	93.2
D	75.4	82.2
E	73.8	71.1
F	79.6	86.9
G	69.4	67.9

[a]Adjusted by proportion of operations conducted on runways 25R and 25L.

Maximum sound levels were identified in each of the one-third octave bands centered at 25–80 Hz in the 30 s prior to and following the (unweighted) maximum noise level of each aircraft noise event recorded in the field. These maximum sound levels were (energy) summed to construct a total maximum low-frequency sound level for each aircraft noise event at each measurement site.

Arithmetic averages of LFSL values for each aircraft overflight event at each of the seven measurement sites were calculated next. These averages of LFSL and maximum A-weighted (MXSA) aircraft noise event values are summarized in Table I. The product–moment correlation between these noise metrics for aircraft noise events ($r=0.69$) accounted for less than half of the variance in predictions of LFSL values from MXSA values by linear regression ($\mathrm{LFSL}=0.82*\mathrm{MXSA}+15.52$).

A spatial interpolation (spline) algorithm was applied to the LFSL values measured at each measurement point to generate a set of contours from which LFSL could be estimated at each street address in the interviewing area. The algorithm treated the LFSL values as pseudo-elevation information to fit a surface through the measurement points. The algorithm in effect draped a rubber sheet over the measurement area in a manner that both preserved the LFSL values observed at the measurement sites and minimized the total curvature of the resulting surface.

Figure 2 shows contours of aircraft-produced LFSL throughout the interviewing area. The LFSL values diminished by about 7 dB per 1000 ft orthogonally to the runway sideline within the interviewing area. (This site-specific empirical value is not necessarily applicable near runways elsewhere.)

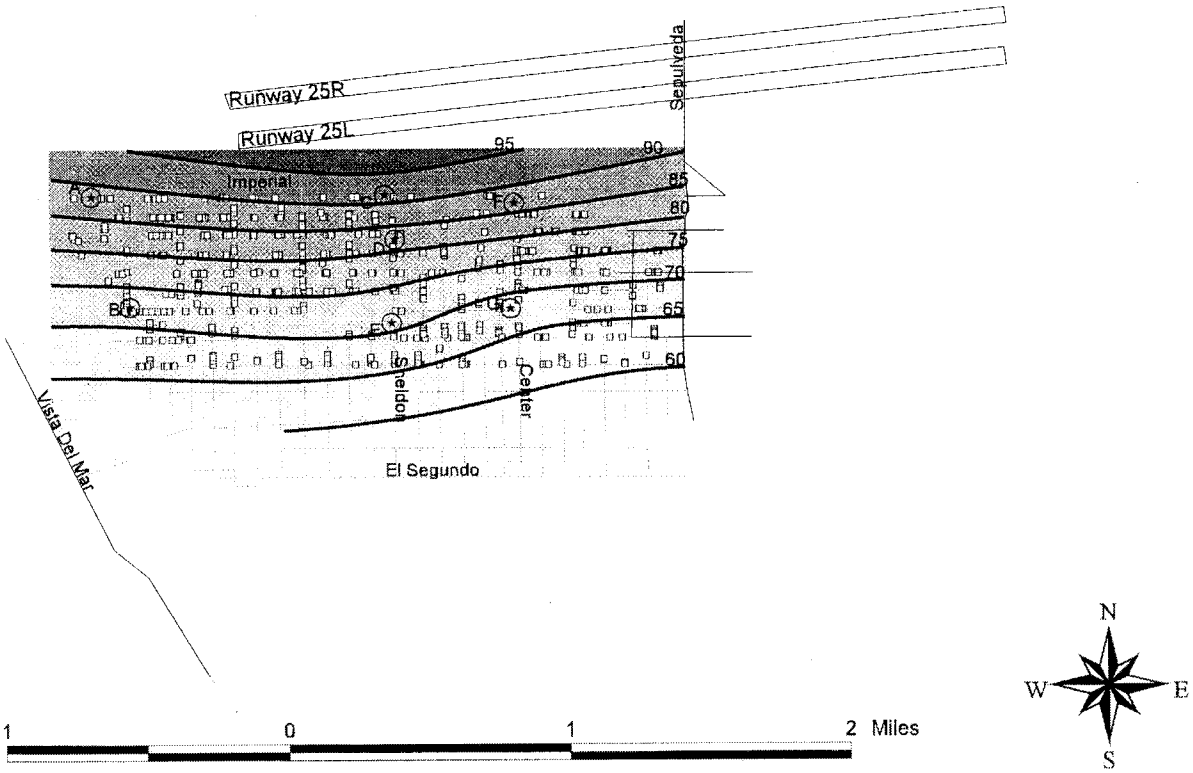


FIG. 2. Contours of low-frequency aircraft noise levels throughout the interviewing area. Shading changes occur in 5-dB intervals.

B. Summary of interviewing process

Table II documents the results of the interviewing process. In all, 644 interviews were conducted with a refusal rate of only 13%. Fifty-six percent of the respondents completing interviews were female, while 44% were male.

Figure 2 also locates households from which completed interviews were obtained with respect to the low-frequency noise contours. The density of completed interviews was fairly uniform throughout the interviewing area. Figure 3 shows the cumulative distribution of the numbers of respondents who lived in households with given low-frequency noise levels. About half of the respondents lived at addresses with low-frequency aircraft noise levels greater than 75 dB.

TABLE II. Accounting for the results of contact attempts from the sampling frame.

Disposition of telephone numbers	
Total No. in sampling frame	1665
Nonsample numbers	
Disconnect	156
Business	64
Fax/modern	54
Wrong	55
Non-English speaking	10
Number changed	64
Total Nonsample	403
Noncontact numbers	
Answering machine	64
Retired numbers (8 callbacks)	336
No answer	115
Not available	5
Total Noncontact	520
Total numbers available for interview [Total−(Nonsample+Noncontact)]	742
Number of completed interviews	644
Number of refusals	98
Completion rate [644/742]	0.87
Refusal rate [98/742]	0.13

C. Responses to questionnaire items

1. Relationships between annoyance due to aircraft noise in general and annoyance due to vibration vibration or rattle

Of the 644 respondents who completed interviews, 29%(185) described themselves as highly (‘‘very’’ or ‘‘ex-

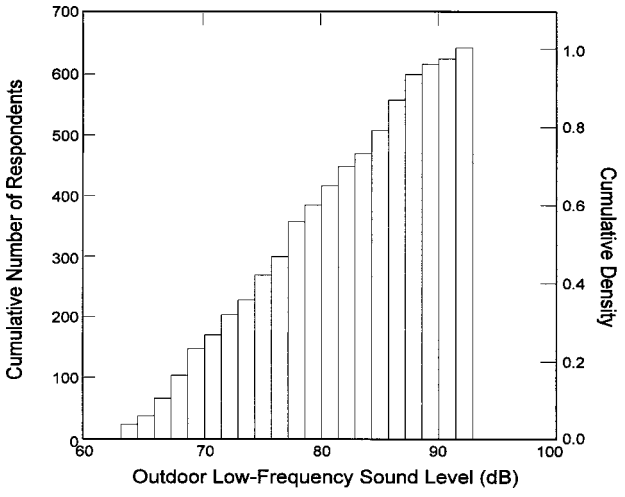


FIG. 3. Cumulative distribution of respondents by outdoor low-frequency noise levels.

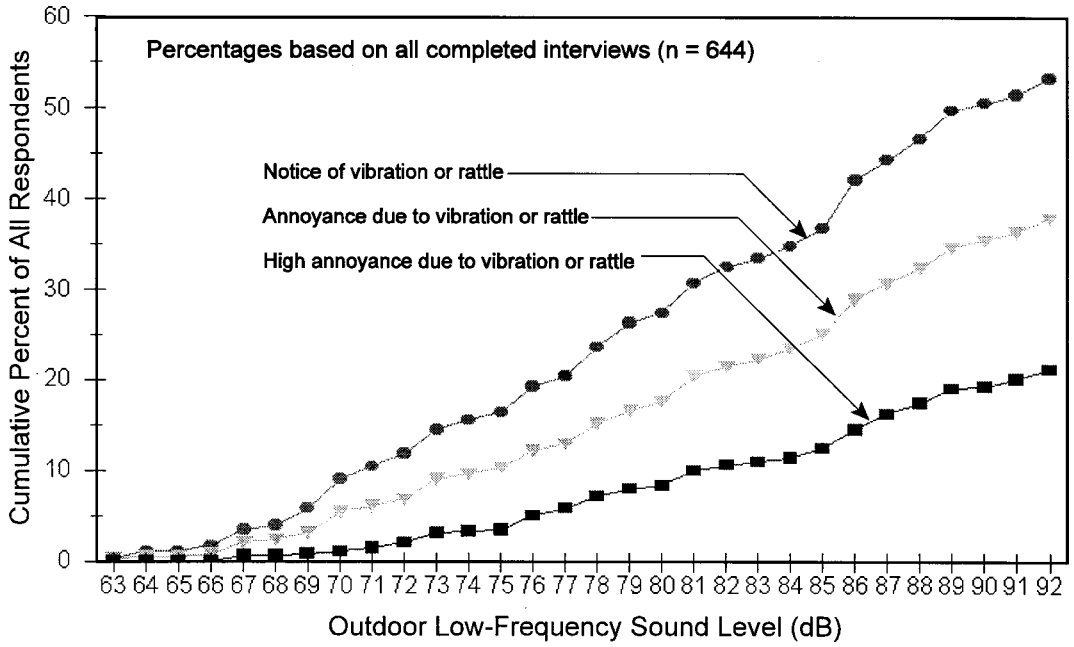


FIG. 4. Cumulative percentages of respondents noticing, annoyed by, and highly annoyed by aircraft-induced vibration or rattle, with respect to outdoor low-frequency noise level.

tremely'') annoyed by aircraft noise in general, while 21%(136) described themselves as highly annoyed by aircraft-induced vibration or rattling sounds in their homes. Of the 136 respondents who described themselves as highly annoyed by vibration or rattling sounds, 23%(25) were **not** highly annoyed by aircraft noise in general. A χ^2 contingency test indicates that this difference is unlikely to have arisen by chance alone ($\chi^2_{(df=1)}=236$, $p<0.01$). Thus, annoyance associated with secondary emissions is not completely subsumed by annoyance due to aircraft noise in general.

2. Relationship of the prevalence of notice and annoyance of vibration or rattle to low-frequency noise levels

Figure 4 compares the cumulative percentages of respondents who noticed, were annoyed in any degree, and were highly annoyed by aircraft-induced vibration or rattle in their homes. The denominator for all of the percentages shown in this figure is 644, the total number of respondents who completed interviews. Information about the distributions of notice, annoyance in any degree, and a consequential degree of annoyance is displayed in Fig. 4 in cumulative form to emphasize the orderliness and straightforward interpretability of the relationships among these variables with respect to outdoor LFSL, to compare their respective slopes, and to illustrate the lack of well-defined breakpoints.

3. Prevalence of aircraft noise annoyance associated with A-weighted cumulative noise exposure

Version 5.1 of the FAA's Integrated Noise Model (Fleming *et al.*, 1997; Olmstead *et al.*, 1995) was used to construct annual day-night average sound level (DNL) contours due to aircraft activity on the southern pair of runways at LAX. These contours were overlaid on the interviewing area so that individual residences could be associated with A-weighted aircraft noise exposure levels. The resulting relationship between the prevalence of a consequential degree of annoyance with aircraft noise and annual DNL (in 4 dB-wide intervals) is shown in Fig. 5 with respect to the dosage–response relationship preferred by FICON (1992). The proximity of the data points from the current study to the FICON curve indicates that the prevalence of annoyance with aircraft noise in general among respondents near LAX is well predicted by this relationship.

4. Relationship between complaints due to aircraft noise in general versus rattle or vibration

About 29% of the respondents who were annoyed by aircraft noise had complained to the airport about aircraft noise in general. A somewhat smaller percentage (25%) of

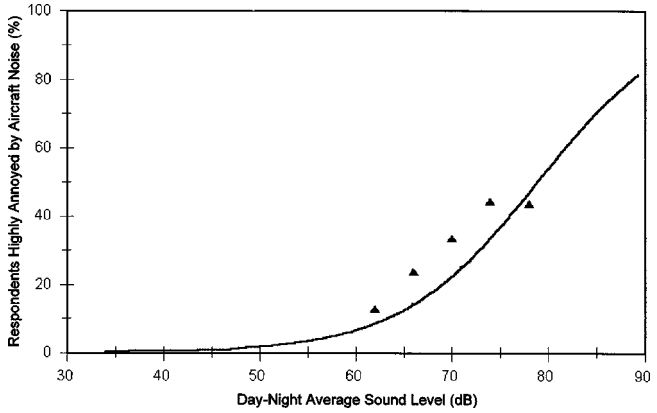


FIG. 5. Prevalence among respondents of a consequential degree of annoyance with aircraft noise.

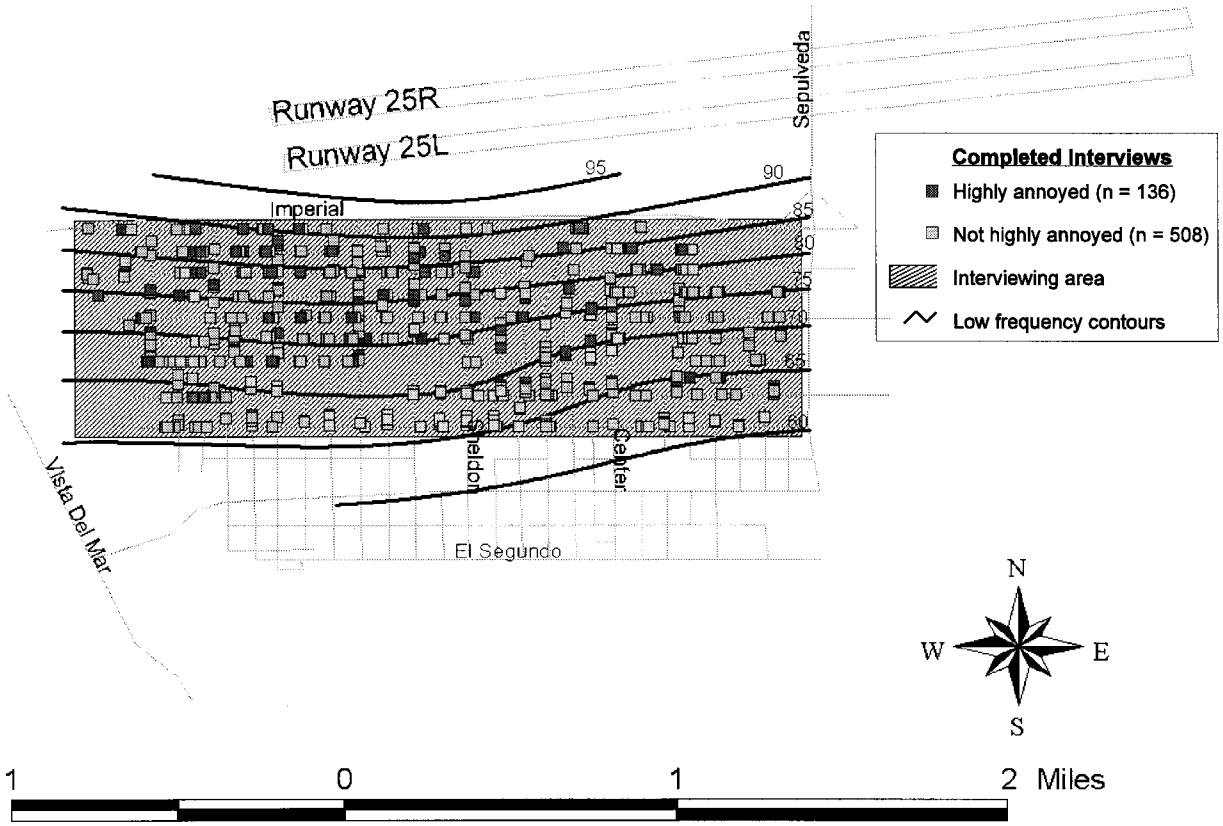


FIG. 6. Locations of households containing respondents highly annoyed and not highly annoyed by rattle or vibration due to low-frequency runway sideline noise.

the respondents who noticed vibrations or rattling sounds in their homes had complained to the airport about aircraft noise. About 30% of the respondents who noticed vibrations or rattling sounds had complained to the airport about them.

III. DISCUSSION AND CONCLUSIONS

Figure 6 shows the locations of households reporting high annoyance due to rattle or vibration as well as the lo-

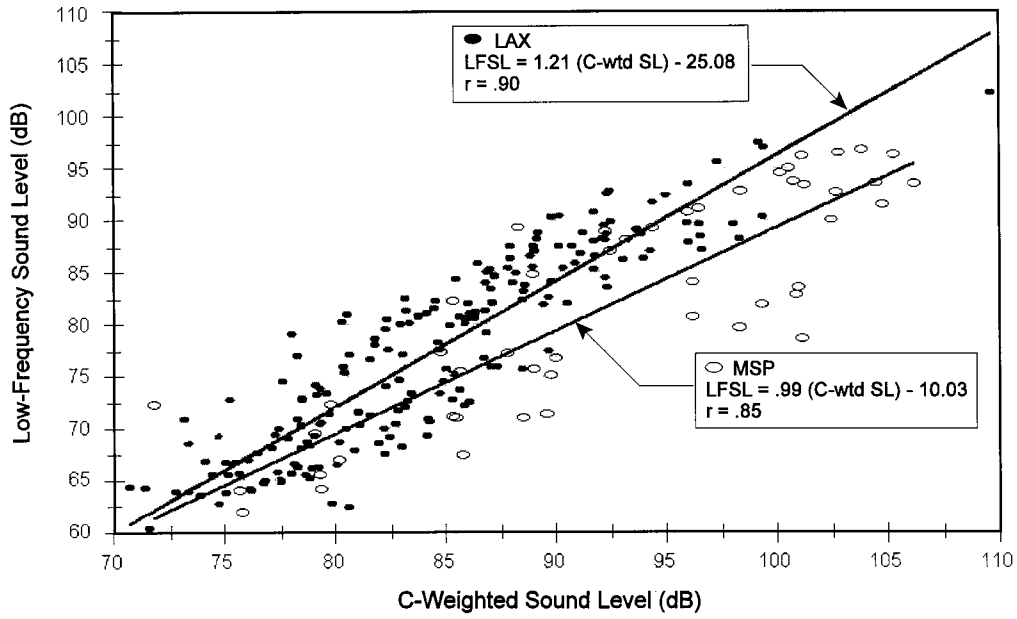


FIG. 7. Linear regressions relating low-frequency sound levels of aircraft ground noise to C-weighted levels at two airports.

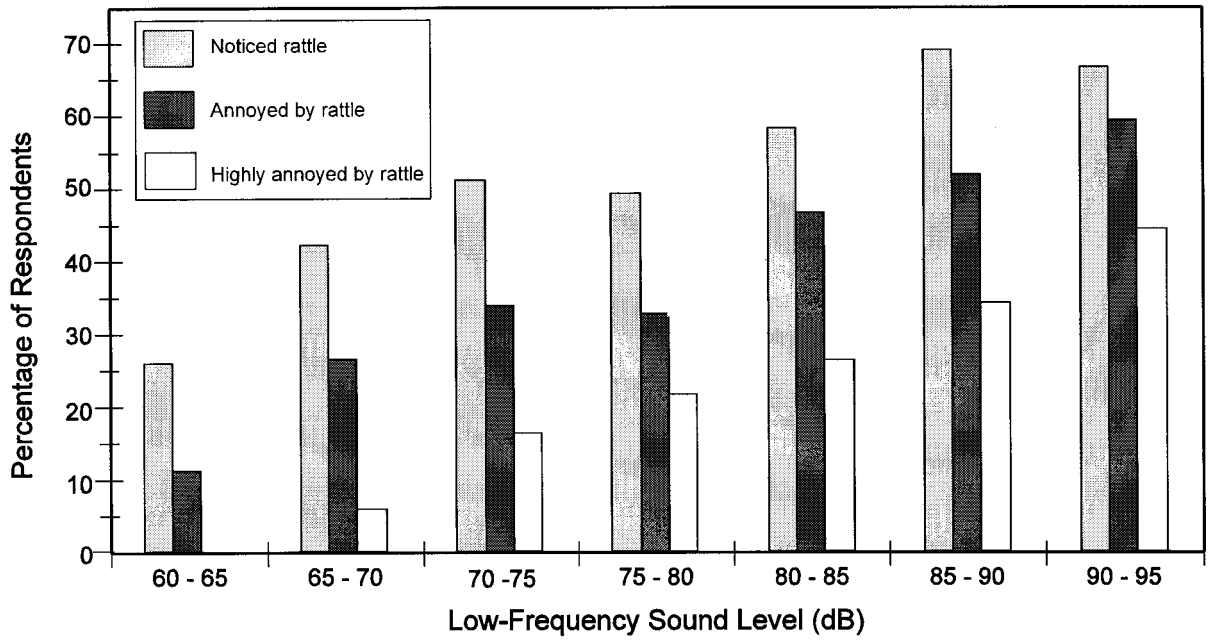


FIG. 8. Percentages of respondents who noticed rattle or vibration, were annoyed in any degree by rattle or vibration, and were highly annoyed by rattle or vibration.

cations of households not reporting high annoyance due to rattle or vibration. Simple visual inspection suggests that residents highly annoyed by low-frequency sideline noise are concentrated in areas with LFSL values in excess of 75 to 80 dB. These levels are consistent with Hubbard's (1982) estimates of low-frequency airborne sound levels capable of inducing secondary emissions in light architectural elements.

Care is required in converting a low-frequency aircraft noise level as characterized for present purposes into a C-weighted sound level due to lack of uniqueness and linearity. The uniqueness issue is that the low-frequency content of noise produced by aircraft equipped with turbofan and other engines of different power ratings may vary considerably even though they share similar C-weighted sound levels. The linearity issue is that increases in C-weighted sound levels may not yield proportionate increases in secondary emissions and, hence, annoyance due to rattle or vibration.

Notwithstanding these cautions, linear regressions displayed in Fig. 7 were performed to relate LFSL values to C-weighted levels for two sets of field observations: those described above at LAX, and those of Lind *et al.* (1997) at MSP. The slopes and intercepts of these relationship will vary for other sets of operations by different aircraft fleets at other airports.

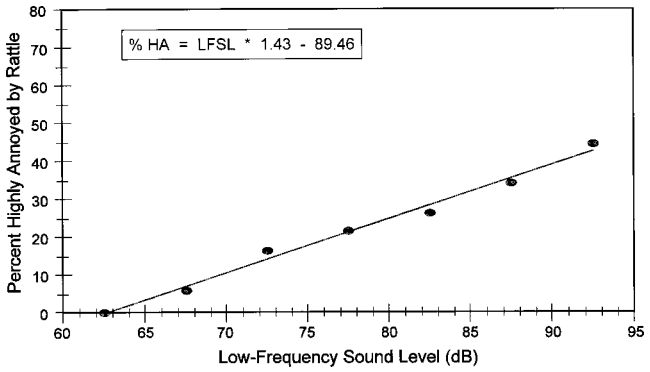


FIG. 9. Relationship between outdoor low-frequency sound levels of aircraft ground noise and the prevalence of a consequential degree of annoyance with rattle or vibration.

Figure 8 compares percentages of respondents who noticed rattle or vibration, were annoyed in any degree by rattle or vibration, and were highly annoyed by rattle or vibration, as aggregated within 5-dB intervals of LFSL. Figure 9 displays a linear regression to the findings about the prevalence of a consequential degree of annoyance within 5-dB LFSL intervals. The product moment correlation of this fit ($r = 0.99$) accounts for essentially all of the variance in the relationship between LFSL and the prevalence of annoyance with runway sideline noise. These data should not be overinterpreted as a fully generalizable dosage-response relationship, since they reflect only the reactions of residents of a single airport neighborhood.

It is nonetheless possible to interpret these initial findings in a manner similar to that adopted by FICON (1992). FICON has identified a value of $L_{dn} = 65$ dB as a threshold of residential land use compatibility. The corresponding prevalence of consequential annoyance in communities according to FICON's dosage-response relationship is 12.3%. The same prevalence of annoyance in the present data set occurs at a LFSL value slightly greater than 71 dB.

ACKNOWLEDGMENTS

The authors are grateful to the respondents for their participation in this study; to the City of Richfield, MN, for sponsoring this study; to Harvey Holden for assistance in estimating A-weighted aircraft noise exposure levels in El

Segundo; to Dr. Norman Broner for an (ill-timed) discussion of the low-frequency noise effects literature; and to Mr. Richard Horonjeff for suggesting the analysis summarized in Fig. 9.

APPENDIX: QUESTIONNAIRE ITEMS

ITEM 1. About how long have you lived at [street address]?

ITEM 2. What do you like best about living conditions in your neighborhood?

ITEM 3. What do you like least about living conditions in your neighborhood?

ITEM 4. Would you say that your neighborhood is quiet or noisy?

SKIP TO ITEM 5 if response to Item 4 was ''quiet.''

Follow up question if response to Item 4 was ''noisy'' or ''quiet, except for airplanes'':

ITEM 4A. Would you say that your neighborhood is slightly noisy, moderately noisy, very noisy, or extremely noisy?

ITEM 5. While you're at home are you bothered or annoyed by street traffic noise in your neighborhood?

SKIP TO ITEM 6 if response to Item 5 was ''no.''

Follow up question if response to Item 5 was ''yes'':

ITEM 5A. Would you say that you are slightly annoyed, moderately annoyed, very annoyed, or extremely annoyed by street traffic noise in your neighborhood?

ITEM 6. While you're at home are you bothered or annoyed by aircraft noise?

SKIP TO ITEM 7 if response to Item 6 was ''no.''

Follow up question if response to ITEM 6 was ''yes'':

ITEM 6A. Would you say that you are slightly annoyed, moderately annoyed, very annoyed, or extremely annoyed by aircraft noise while at home?

ITEM 7. Do airplanes make vibrations or rattling sounds in your home?

SKIP TO ITEM 13 if response to Item 7 was ''no.''

Ask Items 8 through 13 if response to ITEM 7 was ''yes'':

ITEM 8. Are you bothered or annoyed by these vibrations or rattling sounds in your home?

SKIP TO ITEM 9 if response to Item 8 was ''no.''

Follow up question if response to ITEM 8 was ''yes'':

ITEM 8A. Would you say that you are slightly annoyed, moderately annoyed, very annoyed, or extremely annoyed by vibrations or rattling sounds in your home?

ITEM 9. About how often do you notice vibrations or rattling sounds in your home made by airplanes?

ITEM 10. What sorts of things vibrate or rattle in your home?

ITEM 11. Have you tried to do anything in your home to reduce vibrations or rattling sounds made by airplanes?

SKIP TO ITEM 12 if response to Item 11 was ''no.''

Follow up question if response to ITEM 11 was ''yes'':

ITEM 11A. Have the vibrations or rattling sounds made by airplanes been lessened by the things you have done?

ITEM 12. Have you ever complained to the airport about vibrations or rattling sounds in your home made by airplanes?

ITEM 13. Have you ever complained to the airport about aircraft noise in general?

Berglund, B., Hassmén, P., and Job, R. F. S. (**1996**). ''Sources and effects of low-frequency noise,'' J. Acoust. Soc. Am. **99**, 2985–3002.

Broner, N. (**1978**). ''The effects of low-frequency noise on people—a review,'' J. Sound Vib. **58**(4), 483–500.

Federal Interagency Committee on Noise (FICON) (**1992**). ''Federal Agency Review of Selected Airport Noise Analysis Issues,'' Report for the Department of Defense, Washington, D.C.

Fidell, S., Barber, D. S., and Schaltz, T. J. (**1991**). ''Updating a dosage-effect relationship, for the prevalence of annoyance due to general transportation noise,'' J. Acoust. Soc. Am. **89**, 221–233.

Fidell, S., Horonjeff, R., and Green, D. M. (**1981**). ''Statistical analyses of urban noise,'' Noise Control Eng. J. **16**(2), 75–80.

Fields, J. M. (**1993**). ''Effect of personal and situational variables on noise annoyance in residential areas,'' J. Acoust. Soc. Am. **93**, 2753–2763.

Finegold, L., Harris, C. S., and von Gierke, H. E. (**1994**). ''Community annoyance and sleep disturbance: Updated criteria for assessing the impacts of general transportation noise on people,'' Noise Control Eng. J. **42**(1), 25–30.

Fleming, G., Olmstead, J., D' Aprile, J., Gerbi, P., Gluding, J., and Plante, J. (**1997**). ''Integrated Noise Model (INM) Version 5.1 Technical Manual,'' U.S. Department of Transportation, Federal Aviation Administration, Office of Environment and Energy, FAA-AEE-97-04, Washington, DC.

Hubbard, H. (**1982**). ''Noise-induced house vibrations and human perception,'' Noise Control Eng. J. **12**(2), 49–55.

Lind, S., Pearsons, K., and Fidell, S. (**1997**). ''An analysis of anticipated low-frequency aircraft noise in Richfield due to operation of a proposed north-south runway at MSP,'' BBN Report 8196.

Olmstead, J., Bryan, R., Jeng, R., Mirsky, L., Rajan, N., Fleming, G., D'Aprile, J., Gerbi, P., Rickley, E., Turner, J., Le, T., Chen, X., Plante, J., Gulding, J., Vahovich, S., and Warren, D. (**1995**). ''Integrated Noise Model (INM) Version 5.0 User's Guide,'' U.S. Department of Transportation, Federal Aviation Administration, Office of Environment and Energy, FAA-AEE-95-01, Washington, DC.

Piercy, J. E., and Embleton, T. F. W. (**1977**). ''Review of noise propagation in the atmosphere,'' J. Acoust. Soc. Am. **61**, 1403–1418.

Sutherland, L., and Daigle, G. (**1997**). ''Atmospheric Sound Propagation,'' in *Encyclopedia of Acoustics*, edited by J. Crocker (Wiley-Interscience, New York).

Steady-state air-flow and acoustic measurement of the resistivity of loose granular materials

Gino Iannace, Carmine Ianniello, Luigi Maffei, and Rosario Romano
DETEC, University of Naples Federico II piazzale Tecchio, 80-80125 Naples, Italy

(Received 23 July 1998; accepted for publication 21 May 1999)

In the course of a study for characterizing the acoustic properties of some loose granulates made of limestone chips, the authors have measured their resistivity in a very simple way by a steady-state air-flow technique. The results of these tests are compared with the corresponding values obtained by an acoustic technique. The latter relies on the assumption that at very low frequencies the real part of the normal-incidence flow impedance of a thin porous layer is very close to its steady-state air-flow resistance. The measurements results confirmed the above mentioned assumption to a certain degree. To the knowledge of the authors, no previous comparison of steady-state air-flow versus oscillating air-flow measurement of the resistivity of the considered granulates has been published yet. © *1999 Acoustical Society of America.* [S0001-4966(99)00909-1]

PACS numbers: 43.58.Bh, 43.58.Vb [SLE]

INTRODUCTION

Materials made of crushed stone, like sands and gravel, can be used as cheap sound-absorbing porous layers in the instances where their use can be accepted. The authors are engaged in a study aimed at characterizing the intrinsic acoustic behavior of these materials for the above mentioned purpose.

Since Rayleigh[1] suggested a first model, further fundamental models for describing the acoustical properties of porous media have been reported in cornerstones of the relevant scientific literature; just to name a few: Zwikker and Kosten,[2] Biot,[3–5] Morse and Ingard,[6] and Attenborough.[7] Nowadays, a large body of literature reporting the advancement of the knowledge about sound waves in porous materials is available. A review of models for the acoustical properties of air-saturated granular media covering a period up to 1993 is given, e.g., by Attenborough.[8] For the sake of simplicity, the authors consider in the first instance an equivalent dissipative-gas model like that reported in Ref. 9. It is a macroscopic phenomenological model that allows the calculation of the characteristic impedance and propagation constant of the bulk medium on the basis of the knowledge of three nonacoustical physical parameters, that is, the resistivity, the porosity, and the structure factor. This article discusses the first step of the above mentioned work, that is, the determination of reliable values for the resistivity of the considered granulates.

The measurement of the resistivity of porous materials for acoustical use is an issue that dates back more than a half century. Various measurement methods are reported in the scientific literature that are based essentially on forcing a steady-state air-flow through a sample of the porous material. The measurement of both the volume flow rate and the differential pressure across the sample yields the air-flow resistance of the sample.[2,10–12] Measurement methods of this type are reported in the international standard ISO 9053[13] and in other standards as well. By following this simple principle, the authors built an apparatus to measure the steady-state air-flow resistivity.

The ISO 9053 also reports a method for the measurement of the air-flow resistance with an oscillating flow. It is based on the measurement of the sound pressure in a small chamber having rigid walls except the one constituted by the sample to be tested. A piston, oscillating at a very low frequency (about 2 Hz) and connected to the chamber, acts as a constant and known volume-velocity generator. The test condition ensures that the volume velocity of the generator is the volume flow rate through the porous sample. As the realization of a measuring apparatus along the above mentioned lines posed some problems, the authors implemented a simpler apparatus to measure the resistivity with an alternating air-flow. It is similar to the one reported by Ingard and Dear.[14]

This note reports an account of both the measurement techniques for the determination of the resistivity of the considered granular materials and a comparison of the results.

I. TESTED MATERIALS

Limestone crushed stones were sifted through a cascade of square-mesh sieves. The available gauges allowed the selection of the material into six intervals of grain size, namely: 0.2–0.5 mm, 0.5–0.8 mm, 0.8–1.0 mm, 1.0–1.6 mm, 1.6–2.0 mm, and 2.0–4.0 mm. For these materials no further analysis was performed to determine the shapes and the size distribution of the grains belonging to each gauge interval. As dust content can influence the air-flow resistivity of these materials, they were washed and dried naturally before testing.

II. STEADY-STATE AIR-FLOW TECHNIQUE

The resistivity of a porous layer having parallel and flat faces is defined as:

$$R_1 = \frac{1}{v}\frac{\Delta p}{d} = \frac{A}{Q}\frac{\Delta p}{d}, \tag{1}$$

where Δp (Pa) is the pressure difference across the sample faces caused by a steady-state volume flow of air Q (m^3/s)

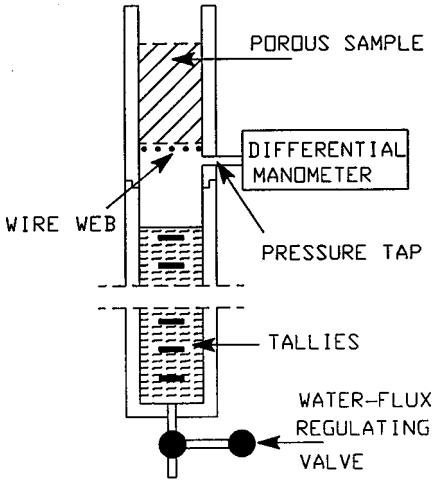


FIG. 1. Test rig for the measurement of the steady-state air-flow resistance of samples of loose granular materials.

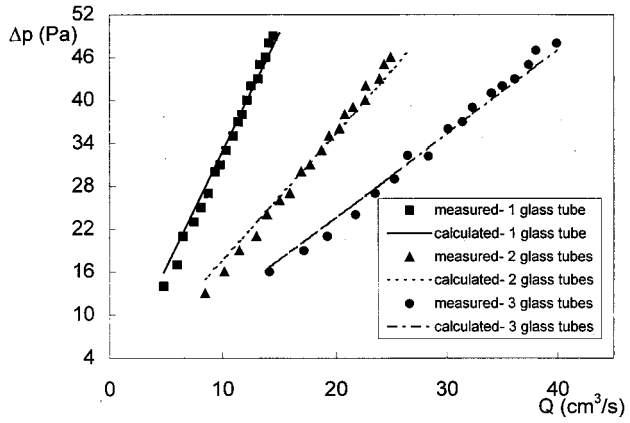


FIG. 2. Pressure drop versus volume flow rate for three samples constituted of calibrated glass tubes.

forced through the material. d (m) is the layer thickness. $v=Q/A$ (m/s) is the average speed of the air above or below the sample having a frontal area A (m^2). The equality of v, over and under the layer, implies that the pressure drop Δp be small enough.

The simple test rig sketched in Fig. 1 was built-up for the measurements according to the basic definition of R_1. It consisted of a Plexiglas tube (inner diameter=99 mm; length =110 cm) that was held vertically by a framework. A piece of the same tube 40 cm long, that was coupled coaxially above the previous one, acted as a sample holder. A wire web with an overlaid metal gauze retained the granular material when poured into the sample holder like a liquid. An electronic differential manometer, which was connected to a pressure tap under the wire web, allowed the reading of the pressure drop Δp when the water filling the lower tube flowed out through a flow regulating valve mounted at its bottom. At the same time, the speed of the air v was measured by reading on a stopwatch the time the lowering surface of the water column took for passing by a number of tallies marked at equal steps of 5 cm along the outer surface of the transparent tube. The validity of this measurement procedure was checked initially by comparing the results of measurements performed with known ''air-flow resistors'' with the theoretical prediction of their behavior. These samples were constituted of a number of calibrated glass tubes (inner diameter=2.7 mm; length=26 cm) mounted through a rubber plug sealing the upper mouth of the 40 cm long tube that acted as a sample holder in the test rig. Figure 2 shows the measured pressure drop Δp versus the measured volume flow rate Q for a sample realized with a single glass tube and for two further samples realized with two and three glass tubes in parallel as well. All tubes were nominally equal. The straight lines in Fig. 2 represent the relevant predictions calculated for each sample with the law of Poiseuille for the laminar flow in a circular duct, that is:

$$\Delta p=\frac{128\mu l Q}{\pi D^4}, \tag{2}$$

where Δp (Pa) is the pressure drop across the duct; μ is the viscosity of the fluid flowing in the duct (for air at the temperature of 20 °C, $\mu=1.8\times10^{-5}$ kg m^{-1} s^{-1}); l (m) is the duct length; D (m) is the duct inner diameter, and Q (m^3/s) is the volume flow rate in the duct. The results reported in Fig. 2 were considered as a fairly good warranty that the described measurement procedure was suitable for the intended purpose.

III. ACOUSTIC TECHNIQUE

If a simple harmonic plane wave impinges normally on a thin layer of a rigid-frame porous material, an oscillating air-flow is forced through it and a time-dependent (sound) pressure drop is observed across its faces. A thickness small with respect to the involved wavelength ensures that the particle velocity at the front face and that at the back face of the porous layer are the same, so that an air-flow through the layer can be considered. In this instance the ratio $\Delta p/v$ between the complex pressure drop and the complex particle velocity defines the flow impedance of the layer. The pressure drop and the particle velocity are generally not in-phase. The flow impedance is not solely resistive, but it has also a mass-type reactance. This causes a slowly increasing magnitude of the flow impedance with the frequency.[15] Ingard and Dear[14] reported that for sufficiently low frequencies the inertia effects of the air moving in the pores can be neglected. Therefore, the real part of the flow impedance of a rigid-frame porous layer approximates its steady-state air-flow resistance $\Delta p/v$.

Figure 3 shows a sketch of the test rig. It consists of a vertical steel tube (length=2.08 m; inner diameter=98 mm; thickness=10 mm) closed at its bottom on the floor with a steel plate 10 mm thick. Its upper mouth is connected coaxially with another piece of brass tube having the same inner diameter and a length of 21 cm. This short tube, acting as a sample holder, has the same wire work as previously described for the steady-state air-flow test rig. Finally, a brass tube (length=100 cm) is connected at its bottom with the sample holder and at its higher mouth with a loudspeaker in a wooden box. Some small holes were tapped into the tubes

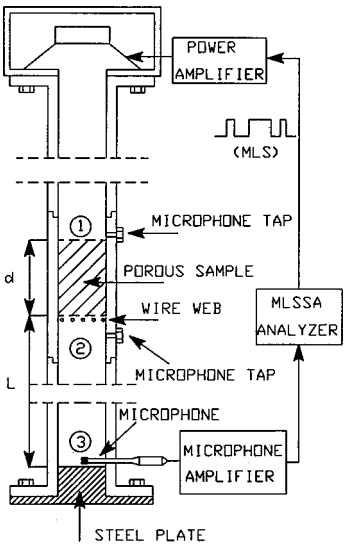


FIG. 3. Test rig for the measurement of the flow impedance of samples of loose granular materials.

for inserting a $\frac{1}{4}$-in. condenser microphone B&K type 4136 at the desired locations. Each tap was closed by a screw when not used for the above mentioned purpose.

By assuming the time-dependence $\exp(j\omega t)$ and a unit reflection coefficient of the sound pressure at the bottom of the terminating tube, a simple plane wave analysis for the lossless tubes yields:

$$Z_f = \frac{p_1 - p_2}{v} = -j\rho c \frac{p_1}{p_3 \sin(kL)} + j\rho c \cot(kL), \tag{3}$$

where Z_f is the flow impedance of the porous layer, p_1 is the complex sound pressure measured at location 1 just over the upper surface of the sample, p_2 is the complex sound pressure measured at location 2 just under the lower surface of the sample, ρc is the characteristic impedance of the free air (ρ being the density and c the sound speed), p_3 is the complex sound pressure at location 3 just over the rigid termination at the bottom of the tube, $k = \omega/c$ is the wave number in the free air at the radian frequency ω, and L is the distance between the lower face of the sample and the rigid termination at the bottom of the tube.

At the frequencies that satisfy the following condition:

$$L = (2n+1)\frac{\lambda}{4}, \quad n = 0,1,2,3,..., \tag{4}$$

that is, the length of the closed tube backing the sample is an odd integer number of quarter wavelengths, the term $\cot(kL)$ in Eq. (3) vanishes and $\sin[k(2n+1)\lambda/4]$ takes the value 1 for $n = 0,2,4,...$ and the value -1 for $n = 1,3,5,...$. Then, at the special frequencies $f_n = (2n+1)c/4L$, $n = 0,1,2,3,...$, the nondimensional flow impedance is:

$$\frac{Z_f}{\rho c} = \frac{-j}{(-1)^n}\frac{p_1}{p_3} = \frac{-j}{(-1)^n}\left[\mathrm{Re}\left(\frac{p_1}{p_3}\right) + j\,\mathrm{Im}\left(\frac{p_1}{p_3}\right)\right].$$

$$f = f_n, \quad n = 0,1,2,3,.... \tag{5}$$

The real part of the nondimensional flow impedance is:

$$\frac{R_f}{\rho c} = \frac{1}{(-1)^n}\mathrm{Im}\left(\frac{p_1}{p_3}\right),$$

$$f = f_n, \quad n = 0,1,2,3,..., \tag{6}$$

and its imaginary part is:

$$\frac{X_f}{\rho c} = \frac{-j}{(-1)^n}\mathrm{Re}\left(\frac{p_1}{p_3}\right)\nu,$$

$$f = f_n, \quad n = 0,1,2,3,.... \tag{7}$$

For f_n low enough, the pressure drop $(p_1 - p_2)$ and the particle velocity v in Eq. (3) are almost in phase and the flow impedance Z_f approximates a real value. Therefore, the imaginary part, given by Eq. (7) for $f = f_n$, can be neglected and the resistivity of the sample material can be approximated by:

$$R_1 \equiv \frac{R_f}{d} \cong \frac{\rho c}{d}\left|\frac{p_1}{p_3}\right|,$$

$$f = f_n, \quad n = 0,1,2,3,.... \tag{8}$$

Actually, measurements were performed by evaluating the frequency response of the linear system as defined by the input of the power amplifier connected with the loudspeaker and by the output of the microphone preamplifier. The frequency response was obtained by using a MLSSA analyzer.[16] This single-channel analyzer convolves the output signal of a linear system with the maximum length sequence (MLS) signal feeding its input. The convolution yields the linear system impulse response. Subsequent FFT transformation gives the required frequency response. If the parameters of the system are time invariant, it is easy to accept that any sound pressure ratio can be determined by the ratio of the corresponding frequency responses obtained with the same microphone located once at the first involved location and a second time at the other one.

The ratio $|p_1/p_3|$ in Eq. (8) was always evaluated at the lowest special frequency allowed by the 2.08-m tube backing the sample, that is, $f_0 = 41$ Hz. For $L = \lambda/4$, there is a single sound pressure minimum just behind the sample and a single maximum on the steel plate terminating the tube. Because of the dependence of the quarter-wavelength frequency on the temperature, it was preferred to check its value by measuring a frequency response at location 2 before each test run.

The frequency of the first minimum in this last-mentioned frequency response was taken as the frequency at which the ratio $|p_1/p_3|$ had to be calculated.

IV. MEASUREMENT RESULTS

The steady-state air-flow measurements were performed by maintaining the differential pressure across the sample in the range 0–50 Pa. Depending on the granulate gauge and sample thickness, the observed speed of the air-flow was in the range 0.3–26.2 mm/s. For each granulate size interval at least three samples having a different thickness in the range 3–38 cm were tested. Furthermore, each sample was renewed at least three times by pouring over again the material into the emptied sample holder. This was done in the aim of including in the results the influence of both the different

TABLE I. Resistivity data of the considered loose granular materials.

Grain-size interval (mm)	R_{1ss} (Pa s/m^2)	σ_{ss}(%)	R_{1a} (Pa s/m^2)	σ_a(%)
0.2–0.5	2.70×10^5	2.2	2.44×10^5	9.7
0.5–0.8	5.08×10^4	5.2	4.48×10^4	7.9
0.8–1.0	2.40×10^4	8.9	2.19×10^4	8.8
1.0–1.6	1.24×10^4	5.7	1.37×10^4	8.1
1.6–2.0	5.88×10^3	7.8	6.60×10^3	8.7
2.0–4.0	2.49×10^3	10.9	4.99×10^3	17.9

grain size distribution in each sample of the same granulate size interval—if any—and the different material settlement. The second column in Table I reports the grand mean of the resistivity R_{1ss} for the corresponding granulate size interval (first column). The third column shows the resulting percent standard deviation σ_{ss}, that is, 100 times the standard deviation divided by the mean. It represents a measure of the above mentioned sample-dependent variability and measurement random errors as well.

The results for the same materials obtained by the acoustic technique are reported in the last two columns of Table I. R_{1a} is the average resistivity deriving from tests carried out on samples having thicknesses of 8 cm and 16 cm for each granulate size interval, except the 0.2–0.5 mm material. In this case a thickness of 3 cm and 4 cm was used in order to obtain an acceptable signal-to-noise ratio at the backface of this highly resistive material. In these tests each sample was renewed four times; σ_a is the relevant percent standard deviation.

Analysis of the data showed that the means of the resistivity obtained by the two different measurement methods are adequately compatible for each granulate size interval, except the 2.0–4.0 mm. The uncertainties were estimated according to ISO.[17] Type B uncertainties were evaluated on the basis of the linear error propagation law. As regards the steady-state air-flow technique, the expanded relative uncertainties (coverage factor=2, that is, 95% confidence interval) were found between 12% and 23%. The corresponding range for the acoustic technique spanned from 7% to 15%. By assuming a negligible correlation degree between the reported measurement methods, a compatibility index I was calculated for each grain size interval as follows:

$$I=\frac{(R_{1ss}-R_{1a})}{\sqrt{(U_{R1ss}^2+U_{R1a}^2)}}, \tag{9}$$

where R_{1ss} and R_{1a} are the average resistivities for the same grain size interval. U_{R1ss} and U_{R1a} are the relevant expanded uncertainties. The values of the compatibility index spanned from 0.49 to 0.83, except the value 2.1 for the granulate size 2.0–4.0 mm.

V. CONCLUSION

Analysis of data obtained by the considered techniques for the measurement of the resistivity of loose granular materials disclosed a good compatibility, that is, the measurement methods are compatible in the range of the performed measurements, except for the material having the largest grain size. The method based on the steady-state air-flow, although simpler, appears to be more time consuming and yielding more uncertainties with respect to the acoustic method.

The granular materials selected for the investigation reported herein have resistivity values in a range that is quite similar to that spanned by the resistivity of fibrous materials, like glass fiber and rock wool, that are used to implement sound-absorbing layers.[12,18]

[1] J. W. Strutt (Lord Rayleigh), *Theory of Sound*, 2nd ed. (1945 re-issue) (Dover, New York, 1877), Vol. II, pp. 319–333.
[2] C. Zwikker and C. W. Kosten, *Sound Absorbing Materials* (Elsevier, Amsterdam, 1949).
[3] M. A. Biot, ''Theory of propagation of elastic waves in a fluid-saturated porous solid. I. Low frequency range,'' J. Acoust. Soc. Am. **28**, 168–178 (1956).
[4] M. A. Biot, ''Theory of propagation of elastic waves in a fluid-saturated porous solid. II. Higher frequency range,'' J. Acoust. Soc. Am. **28**, 179–191 (1956).
[5] M. A. Biot, ''Generalized theory of acoustic propagation in porous dissipative media,'' J. Acoust. Soc. Am. **34**, 1254–1264 (1962).
[6] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).
[7] K. Attenborough, ''Acoustical characteristics of porous materials,'' Phys. Rep. **82**, 179–227 (1982).
[8] K. Attenborough, ''Models for the acoustical properties of air-saturated granular media,'' Acta Acust. **1**, 213–226 (1993).
[9] M. C. Bèrengier, M. R. Stinson, G. A. Daigle, and J. F. Hamet, ''Porous road pavements: Acoustical characterization and propagation effects,'' J. Acoust. Soc. Am. **101**, 155–162 (1997).
[10] L. L. Beranek, *Acoustic Measurements* (Wiley, New York, 1959), pp. 844–854.
[11] D. Huszty, A. Illenyi, and Gy. Vass, ''Equipment for measuring the flow-resistance of porous and fibrous materials,'' Appl. Acoust. **5**, 1–14 (1972).
[12] D. A. Bies and C. H. Hansen, ''Flow resistance information for acoustical design,'' Appl. Acoust. **13**, 357–391 (1980).
[13] ISO 9053, ''Acoustics: Materials for acoustical applications-determination of airflow resistance'' (1988).
[14] K. U. Ingard and T. A. Dear, ''Measurement of acoustic flow resistance,'' J. Sound Vib. **103**, 567–572 (1985).
[15] U. Ingard, ''Notes on sound absorption technology,'' Chap. 10, Noise Control Foundation, P.O. Box 2469 Arlington Branch, Poughkeepsie, NY 12603 (1994).
[16] D. D. Rife, MLSSA Reference Manual Version 10.0W (1997).
[17] ISO GUM, ''Guide to the expression of uncertainty in measurement'' (1995).
[18] D. A. Bies, ''Acoustical properties of porous materials,'' in *Noise and Vibration Control*, edited by L. L. Beranek (McGraw-Hill, New York, 1971).

Relationships between DPOAE and TEOAE amplitude and phase characteristics

Richard D. Knight and David T. Kemp
Auditory Biophysics Group, ILO, University College London, RNTNE Hospital, 330/332 Gray's Inn Road, London WC1X 8EE, United Kingdom

(Received 30 December 1998; accepted for publication 3 June 1999)

Most published data comparing the amplitudes of transient evoked otoacoustic emissions (TEOAEs) and bi-tonally evoked otoacoustic emissions (DPOAEs) indicate a low level of correlation, raising the question to what extent do the two responses share the same relationship with hearing function. However, DPOAE intensities are sensitive to stimulus parameters and comparisons with TEOAE have mostly been made with the specific range of DPOAE parameters found to optimize DP output level. To determine if other DPOAE stimulus parameter domains give closer correspondence between TEOAE and DPOAE characteristics, $2f_1-f_2$ and $2f_2-f_1$ DPOAE intensity and phase measurements were made across a sample 1/2-octave frequency range centered on 2 kHz in nine normally hearing subjects using a wide range of stimulus parameter configurations. The DP fine structure was resolved by detailed measurements and the mean DP levels were compared to those in the corresponding frequency range from TEOAE measurements obtained with the same probe fitting. The closest relationships between TEOAE and DPOAE amplitude were obtained with the smallest DPOAE stimulus frequency ratios and with lower DPOAE stimulus levels. For these conditions, the DPOAE intensity for individuals in this subject group could be predicted from TEOAE with a standard deviation of 1 dB, similar to the test–retest difference. The $2f_1-f_2$ DPOAE phase versus frequency gradients for fixed f_2/f_1 corresponded closely with phase gradients in TEOAE data when a small DP primary stimulus frequency ratio was used. They differed markedly at wider frequency ratios. In contrast, $2f_2-f_1$ DP phase agreement with TEOAE was good for all stimulus parameters, where measurable. These data suggest that the detailed mechanism of TEOAE and all DPOAEs is very similar when close stimulus tones are used to stimulate DP's. Significant divergence exists with the $2f_1-f_2$ DP with the wider stimulus ratios typically employed for clinical testing. The reasons are discussed. [S0001-4966(99)03309-3]

PACS numbers: 43.64.Jb, 43.64.Kc, 43.64.Ri [BLM]

INTRODUCTION

Sound emission from the ear is well established as a sign of healthy cochlear status and evoked otoacoustic emission (OAE) testing is widely used both for screening and as a diagnostic indicator. It is generally accepted that otoacoustic emissions originate in the physiological response of the cochlea's sensory mechanism to mechanical stimulation and that outer hair cells play a major part in this. (For a general review, see Robinette and Glattke, 1997.) The exact mechanism for the escape of sound energy from the cochlea has not been established in detail, but the creation of OAEs appears to be intimately associated with the cochlear traveling wave from which OAEs inherit a substantial part of their characteristic latency.

The OAEs are in general evoked by any acoustic stimulus, but the most appropriate method and the relative difficulty of response extraction varies with the type of stimulus given. The two techniques most commonly used to study OAEs and which are most often used in clinical practice are the transient evoked OAE method (TEOAE) and the distortion product OAE method (DPOAE). These techniques are superficially very different so that it is reasonable to expect that the OAEs recovered by the techniques will also differ in their relation to hearing function.

With TEOAE recording, a wideband click stimulus is repeatedly presented to the closed ear canal. The cochlea is therefore observed repeatedly ''returning'' to its rest state following brief stimulation. Synchronous averaging is applied to recover low-level delayed ear canal acoustic signals which include the OAE. The acoustic response component attributable to cochlear activity is separable from that of the ear canal and the middle ear by signal processing techniques sensitive to the time delay and nonlinearity of the OAE (Kemp *et al.*, 1986). The linear components of the cochlear response are sacrificed in this process.

With the DPOAE method, stimulation can be continuous, consisting of two simultaneously applied tones with frequencies f_1 and f_2, with f_2/f_1 usually set to between 1.3:1 and 1.2:1 for maximum output. Nonlinearity in the cochlear mechanism means that OAE components are produced not only at the stimulus frequencies f_1 and f_2, but also at intermodulation frequencies given by $f_2+N(f_2-f_1)$, where N may be a small positive or negative integer. Only those frequency components not coincident with f_1 and f_2 are extractable by spectral analysis and of these $2f_1-f_2$ and to a much lesser extent $2f_2-f_1$ are most commonly selected for examination. While mechanical nonlinearity may be expected to cause $2f_1-f_2$ and $2f_2-f_1$ distortion products at

any point in the f_2 traveling wave envelope, it is thought that the $2f_1-f_2$ DPOAE arises primarily from the region of the f_2 peak on the basilar membrane, whereas the higher frequency $2f_2-f_1$ DPOAE is emitted from a more basal region near the DP frequency place (Martin *et al.*, 1987; Kemp, 1998; Martin *et al.*, 1998).

Irrespective of differences in the signal recovery method, DPOAE production contrasts markedly with TEOAE production because (a) with DPOAE, the portion of the cochlea stimulated is restricted to the excitation pattern of the two close tones f_1 and f_2 whereas a click applies all frequencies to the cochlea simultaneously, (b) because TEOAEs and DPOAEs are different subsets of the total OAE response present, (c) because the DPOAE response is produced in the ''steady stimulation'' state whereas TEOAE is recorded after stimulation has ceased, and (d) TEOAE is obtained across frequencies simultaneously, whereas DPOAE is obtained sequentially. Point (c) is moderated by the fact that the combined DPOAE stimulus of f_1 and f_2 is pulsatile with a period of f_2-f_1.

The OAEs are accepted as an audiological tool, however it is not self-evident which of these two convenient means of evoking and extracting an OAE yields a signal most closely linked to the functional status of the sensory mechanism.

Also, it is not clear that the intensity of OAEs, which is only a byproduct of sensory activity, is the measure most likely to be closely related to function. Furthermore, DPOAE intensity depends not only on the stimulus intensity, but it also varies very substantially with the particular stimulus frequency and amplitude ratios used. It is sometimes assumed that the frequency-specific nature of DPOAE stimulation will logically lead to DPOAEs being more closely associated with function, but a close relationship to auditory threshold has yet to be found (e.g., Gorga *et al.*, 1993; Moulin *et al.*, 1994; Gorga *et al.*, 1997; Dorn *et al.*, 1998).

A relationship between TEOAE and $2f_1-f_2$ DPOAE amplitude has been shown (Moulin *et al.*, 1993; Probst and Harris, 1993; Smurzinsky and Kim, 1992; Smurzinsky *et al.*, 1993), but the results only indicate weak correlations. However, these studies for the most part only covered the small range of DP stimulus parameters which are in clinical use. For example, f_2/f_1 ratios only ranged from 1.17–1.23. They also used little or no averaging across the OAE fine structure, which is prominent in both DPOAE and TEOAE. There has been no comparable study between the $2f_2-f_1$ DP and TEOAE.

The TEOAE and DPOAE arise from such different spatial and temporal excitation patterns of stimulation within the cochlea that relative differences between the levels of each is to be expected between individuals. Interference between elemental contributions distributed along the cochlea must affect the ear canal signal greatly and this may vary between individuals. One would therefore expect at least to have to average OAE activity over a range of stimulus parameters in order to arrive at a reliable indicator of cochlear status, but this is not common practice. Also, the very different stimuli used for TE and DP OAE measurements must, to some extent, exercise different aspects of the sensory mechanism's input–output characteristic and this may also differ from ear to ear.

These considerations raise a question: which set of OAE stimulation and measurement parameters is least influenced by any intra- and intersubject differences which do not alter hearing sensitivity but may be significant to OAEs? Such parameters would be most appropriate for the purpose of obtaining an index of cochlear status.

As noted above, most comparative work on OAEs has used limited ranges of parameters because of the interest in clinical and screening applications. For DPOAE the stimulus amplitude and frequency ratios giving maximum DP output levels have been preferred in order to optimize testing time and test specificity (e.g., Harris *et al.*, 1989; Gaskill and Brown, 1990; Hauser and Probst, 1991) and to maximize differences between hearing and hearing impaired ears (e.g., Stover *et al.*, 1996; Moulin *et al.*, 1994). Reductions in the absolute stimulus levels which give increased sensitivity to cochlear disorder are traded against increased testing time within the environment of a clinic. As a result, levels exceeding 60 dB SPL are commonly used to achieve adequate specificity and speed. With TEOAE, sensitivity is generally maintained even with the highest click levels compatible with minimal transducer artifact levels (around 90 dB SPL peak) and these higher levels optimize the specificity (Prieve *et al.*, 1993).

Central to this paper is the consideration that stimulus parameters typically used for DP and TEOAE in audiological investigation may not necessarily be optimal for describing the functional differences between two cochleae. The influence of spurious factors may or may not be minimized by the adoption of stimulus conditions yielding maximum DPOAE intensity. They may also not be optimal for deciding how closely the two OAE response modes (TEOAE and DPOAE) are physiologically related.

There are more fundamental reasons for expecting a divergence between even averaged TEOAE and DPOAE intensity levels and auditory threshold. First, the otoacoustic reemission mechanism itself is a by-product of cochlear function and is not directly part of the hearing process, and, second, the vital process of transduction at the inner hair cell which leads to audition does not participate in OAE production. The relationship between OAEs and hearing threshold is beyond the scope of this paper.

The OAEs arise from functionally important elements of the cochlea—almost certainly the outer hair cells. This paper is concerned with the relationship between two modes of OAE detection: TEOAE and DPOAE. Clearly there are many unexplained sources of variance in OAEs, for example, those between OAE detection modes, between parameter configurations, and between individuals which may not all be related to hearing function. If, by the selection of appropriate stimulation parameters, very closely correlated OAE characteristics were obtainable by two different and independent OAE measurement techniques across a subject population, it is likely that those particular OAE detection modes are equally influenced by the intervening variables. More specifically, if for any stimulus configuration the DPOAE and TEOAE characteristics are found to agree

across a population, then it becomes less likely that, for that stimulus configuration, TEOAE and DPOAE generation relates to very different aspects of cochlear function.

The purpose of the present study is therefore to establish the relationship between TEOAE and DPOAE characteristics over a broader range of stimulus parameters than is presently represented in the literature and from the results to draw inferences of the functional relationships between the two modes of OAE observation. To this end we have included measurements of OAE phase gradients with frequency and also measurements of the less studied $2f_2-f_1$ DPOAE.

I. METHOD

A DPOAE and TEOAE study was designed to enable a detailed investigation of the relationships between the two OAE forms without demanding an excessive amount of time from each subject. As the phase and general waveform of TEOAE are not critically dependent on stimulus level, a single stimulus amplitude was used for TEOAE measurements. A narrow frequency spacing of DPOAE measurements was desired to allow averaging across the OAE fine structure and derivation of phase gradients. Therefore, to remain within the time constraints the f_2 frequency of the DPOAE measurements was limited to approximately a 1/2 octave range centered on 2 kHz, within which a range of stimulus frequency ratios and stimulus levels were explored.

A. Subjects

Nine left ears from subjects ranging in age from 22–41 years (mean 31.4 years) were given the OAE test battery (eight female and one male). All ears, when tested via standard audiometry, demonstrated normal hearing threshold sensitivity ($\leqslant$20 dB HL) from 125 Hz to 8 kHz with the exception of one ear, which gave a threshold of 25 dB HL at 4 kHz. All ears complied with normal tympanometry patterns.

B. OAE recordings

1. TEOAE

For TEOAE, the stimulus was presented at 20.5-ms intervals using the ILO 88 DPI in the nonlinear mode (Kemp *et al.*, 1986; Kemp *et al.*, 1990). Each measurement was derived from 260 nonlinear subaverages using a 2.5–20.5-ms time window with a 2.5-ms risetime. Use of a 512-point fast Fourier transform (FFT) yielded spectral harmonic values at 48.8-Hz intervals.

In order to calibrate the probe fitting in the ear prior to recording, the click stimulus, measured with the microphone contained within the probe, was assessed to confirm an absence of excessive ringing or other evidence of a poor probe fit (Kemp *et al.*, 1990) and the frequency spectrum (obtained by Fourier transform of the stimulus click) was essentially flat in the frequency range from 1 to 4 kHz. For each subject, the stimulus amplitude was adjusted to register a peak close to 0.3 Pa at the probe in the ear canal.

2. DPOAE

Short DP sweeps (i.e., DP intensity versus frequency) were constructed using 16 f_2 frequencies in the frequency range 1660–2393 Hz, which was adequate to cover an auditory filter bandwidth. The frequencies of f_2 were chosen to coincide with each harmonic of the TEOAE response in this range.

The parameters chosen for the DP sweeps were selected to cover as wide a range of conditions as possible within the constraint of a reasonable time period for the subject. The DPOAE stimulus parameters are defined by L_1 (sound pressure level of lower stimulus tone of frequency f_1) and L_2 (sound pressure level of higher stimulus tone of frequency f_2).

L_1: Levels of 65, 70, and 75 dB SPL were used, which are similar to those used by Harris *et al.* (1989). Gaskill and Brown (1990) used stimulus levels from 40 to 65 dB SPL, but these lower levels were not included in this study as the lower amplitude responses which would have resulted would have necessitated a longer averaging period for each measurement, reducing the range of conditions which could be studied within the time period.

L_1-L_2 (level difference): measurements were made with L_2 reduced below L_1 by 0, 5 and 10 dB, so including the condition $L_1=L_2$ and also the conditions with L_2 reduced below L_1 which can result in larger amplitude $2f_1-f_2$ emissions. The latter has been preferred for clinical applications, e.g., by Whitehead *et al.* (1995) and Stover *et al.* (1996).

f_2/f_1 (frequency ratio): Ratios of 1.05, 1.2, 1.27, and 1.32 were used, the latter three being clustered around the ratios 1.2–1.3 known to give maximum-amplitude DPOAEs (Harris *et al.*, 1989; Gaskill and Brown, 1990), and the first ratio investigating a much smaller frequency ratio which has been shown to have different $2f_1-f_2$ DP phase properties (Kemp, 1986) and which may produce a larger amplitude $2f_2-f_1$ distortion product response (Erminy *et al.*, 1996).

The DPOAE test sequence was programmed to run semi-automatically under the control of a macro. Both the $2f_1-f_2$ and $2f_2-f_1$ DPs were recorded simultaneously and automatically saved from within the macro. Measurements were made using the ILO 88 DPi, with software version 5.6Z. Before each DP sweep, a calibration procedure was undertaken in which three stimuli were presented cyclically: a 250-Hz sinewave and a broadband click presented to each of the two probe output transducers. The 250-Hz sine wave was used to provide an estimate of the ear canal volume by comparing the measurement from the probe microphone with data obtained from a 1 cc cavity. The responses to the broadband clicks obtained by the probe microphone were Fourier transformed and it was ensured that the two spectra obtained were approximately equal and essentially flat in the frequency range 1–4 kHz. Once this stage had been accepted, the broadband clicks were presented a further 16 times to obtain a reference spectrum of the accepted fit. This reference was used to balance and normalize the two stimuli at each of the test frequencies of the DP sweep. Normally, where the software detects wide variations in the spectrum with frequency, a standing wave effect within the

ear canal is suspected and the compensation is not performed. However, in this study, this condition was not accepted within the frequency range of interest at the calibration stage and so the level compensation always took place.

For each stimulus condition, one full frequency sweep was completed and then individual points were repeated at frequencies at which the signal-to-noise ratio was judged to be the lowest. Each DP measurement was terminated manually, usually after all $2f_1-f_2$ responses exceeded $2\times$ the standard deviation of the noise.

The DPOAE test sequence typically lasted approximately 45 min, depending on the signal-to-noise ratio of the responses. The equipment was situated in a quiet office in which the ambient noise was typically 40 dBA SPL.

C. OAE measurement procedure

Following the audiometric tests, for each subject the OAE probe was fitted with a foam tip and inserted into the ear canal and the TEOAE calibration procedure was performed. The same adult probe was used (with a different foam tip) for each subject. A TEOAE measurement was then obtained.

Following this, the complete series of DP sweeps was collected. Generally the original probe position was retained throughout the session. The probe was only reinserted on occasions when it moved during the test sequence. Where this occurred during a DP sweep, the DPOAE calibration procedure was repeated and the DP sweep measurement was rerun.

For each subject the DP measurements were obtained in the same order. Initially, with $L_1=L_2=65$, dB SPL, four DP sweeps were obtained with frequency ratios of 1.05, 1.20, 1.27, and 1.32. Here L_2 was reduced by 5 dB and the DP sweeps with the four frequency ratios were repeated. Then L_2 was reduced by a further 5 dB and the four DP sweeps were repeated again. This sequence was then repeated, beginning with $L_1=L_2=70$ dB SPL and then with $L_1=L_2=75$ dB SPL. In addition, at the start and end of the test sequence an additional DP measurement was obtained using the test parameters $L_1=L_2=70$ dB SPL and $f_2/f_1=1.2$. These two extra measurements were used to check the stability of the responses.

D. Analysis and calculations

1. Comparison of TEOAE and DPOAE amplitudes

In order to allow a straightforward comparison between TEOAE and DPOAE levels, a single figure was required which would represent the OAE level obtained across the frequency range for each OAE stimulus configuration. A different calculation method was adopted for TEOAE and DPOAE because the TEOAE measurement presents all frequencies to the cochlea simultaneously, whereas the DPOAE measurement tests each frequency sequentially. This means that for TEOAE the capacity for OAE generation is divided across frequency, whereas for DPOAE the full capacity of OAE generation is available for every frequency in the DP sweep. It is therefore appropriate to average DPOAE across frequency, but to sum TEOAE across a suitable frequency range.

For this study, the sum of the 16 TEOAE levels from 1660 to 2393 Hz was calculated and compared to the average level of each 16-point DP sweep:

$$\mathrm{DPOAE}_{\mathrm{ave}}=10\log\left(\left(\sum_{i=1}^{16}10^{\mathrm{DPOAE}/10}\right)\Big/16\right),$$

$$\mathrm{TEOAE}_{\mathrm{sum}}=10\log\left(\sum_{i=1}^{16}10^{\mathrm{TEOAE}/10}\right).$$

This calculation difference makes the reasonable assumption that the frequency range under investigation covers approximately one auditory filter and increases the numerical value for TEOAE levels relative to DPOAE by $10\log(16)$, i.e., 12 dB. The difference in calculation method has no effect on the subsequent data analysis of relationships between TEOAE and DPOAE.

2. Interrelations between TEOAE and DPOAE

For each configuration of DPOAE stimulus parameters, a scattergraph was generated in which the frequency-averaged DPOAE level was plotted against the band-limited frequency averaged TEOAE level of each subject.

The $2f_1-f_2$ DPOAE was plotted against the TEOAE frequency band corresponding to the f_2 frequency range, whereas the $2f_2-f_1$ DPOAE was plotted against the TEOAE band corresponding to the $2f_2-f_1$ DP frequency.

For each set of DPOAE stimulus parameters, a linear regression line was calculated from the paired DPOAE and TEOAE level data and the standard deviation of the data from the regression line was calculated. This measure was used as the main descriptor of the relationship between the measurements as it indicates the uncertainty associated with using the regression line to predict the DPOAE level from the TEOAE measurement.

The standard deviation calculated in this way will be referred to as the prediction standard deviation (prediction s.d.).

In order to investigate the influence of the across frequency averaging on the closeness of the relationship between TEOAE and DPOAE level, prediction s.d. was also calculated using single point TEOAE and DPOAE data without cross-frequency averaging for two DPOAE stimulus configurations, with $L_1=L_2=70$ dB SPL and $f_2/f_1=1.05$ and 1.32. To ensure that the optimum frequency alignment of the TEOAE and DPOAE fine structure was achieved, the TEOAE data was shifted by successive frequency steps and a series of prediction s.d. values were calculated. The lowest calculated value for prediction s.d. was then compared to the corresponding figure for the averaged data. The difference indicated the extent to which the apparent relationship between TEOAE and DPOAE was weakened by the fine structure mismatch of the OAE's when unaveraged data was compared.

The gradient (ΔDP/ΔTE) of the best fit line was also obtained as an indicator of the relationship between TEOAE

and DPOAE as a gradient close to 1 means that intersubject factors affecting OAE level effect both OAE forms in equal proportion, implying similarities between the underlying mechanisms of TEOAE and DPOAE.

3. Phase and phase gradients

The phase of the OAE response relative to the stimulus is readily measurable and may help to characterize the origin of an emission. The phase versus frequency gradient associated with $2f_1-f_2$ DPOAE with stimulus sweeps of a constant frequency ratio has previously been shown to be fairly flat when f_2/f_1 exceeds approximately 1.1, but to develop a much steeper gradient at smaller frequency ratios (Kemp, 1986; Knight and Kemp, 1999). This may indicate a difference in the means by which the $2f_1-f_2$ DPOAE is emitted at small and large f_2/f_1 ratios, perhaps the wave-fixed and place-fixed behavior first described by Kemp and Brown (1983). There is limited data in the literature regarding the phase behavior of the $2f_2-f_1$ DP. Wable *et al.* (1996) used an f_2 sweep test and found that the $2f_2-f_1$ DP group delay was less than that of the $2f_1-f_2$ DP if the emission frequencies were matched, and Moulin and Kemp (1996) found no difference in $2f_2-f_1$ DP group delay obtained from f_1 or f_2 sweeps.

The TEOAE measurements have associated latencies of 5–15 ms (depending on frequency) measured both in the delay seen in the time domain (Kemp, 1978) and derived from the phase gradient (Wilson, 1980).

The phase data obtained from the OAE recordings were only defined within the range ± 180 degrees. The data sequences were ''unwrapped'' using the following procedure to yield the phase gradient. First, to deal with instances in which the data jumped up from -180 to $+180$ degrees, 360 degrees was subtracted if the phase of a data point differed from adjacent data points by more than $2\times$ the typical variance for a data series plus the maximum variation due to noise at each point (based on $2\times$s.d. above the noise mean). Second, to correct downward jumps from $+180$ to -180 degrees, 360 degrees was added if the phase gradient fell below the overall line gradient by more than 360 degrees minus the maximum noise and $2\times$ the typical variance.

These criteria were chosen as they are logical and also appear to select the most appropriate phase correction without being triggered excessively by data points with a poor signal-to-noise ratio.

Subsequently, the phase gradient (relative to the change in emission frequency) was used to provide a calculation of the group delay. The full data series from each DP gram was utilized provided that all points exceeded the mean noise by at least $1\times$s.d. Where this was not the case, the data sequence was shortened to include the longest continuous sequence of acceptable data. For a gradient to be recorded, a minimum of three consecutive data points were required.

The phase gradient of the $2f_1-f_2$ DP was compared to the TEOAE phase gradient across the frequency range of f_2, whereas the $2f_2-f_1$ DP was compared to the TEOAE phase across the frequency range corresponding to the emission frequency.

II. RESULTS

A. DPOAE levels

The $2f_1-f_2$ distortion product exceeded the mean noise by $1\times$ the s.d. across the measured frequency range in all ears for all stimulus conditions with the exception of a few isolated points. The $2f_2-f_1$ distortion product was also measurable in all ears, but some parameter configurations gave low-amplitude responses containing a large number of points within the DP sweep which failed to meet the S/N criteria. Figure 1 shows all DPOAE data obtained with $L_1=70$ dB SPL from one subject. The levels of each DP to each stimulus condition, averaged across the frequency range and across the nine subjects, are shown in the graph in Fig. 2.

The overall trends seen in the amplitudes of the $2f_1-f_2$ distortion product responses with stimulus frequency ratio and stimulus level are similar to those reported by Harris *et al.* (1989).

The lowest amplitude $2f_1-f_2$ DP responses were obtained when $f_2/f_1=1.05$, whereas the frequency ratio to give the maximum amplitude response increased from 1.2 to 1.32 as L_1 increased and L_1-L_2 increased. The level difference resulting in the largest response level also increased from 5 to 10 dB as the frequency ratio increased.

The highest growth rate with increasing stimulus level seen in the $2f_1-f_2$ responses was 1.4 dB/dB and occurred when $f_2/f_1=1.32$ and $L_1=L_2$, while the lowest growth rate was 0.1 dB/dB and occurred when $f_2/f_1=1.05$ and $L_1=L_2+10$ dB.

In contrast to the $2f_1-f_2$ DP, the highest level $2f_2-f_1$ DP responses were obtained when $f_2/f_1=1.05$ and $L_1=L_2$. Reducing L_2 below L_1 resulted in reductions in $2f_2-f_1$ DP response, the greatest being 1.4 dB per 1-dB change in L_2 when $f_2/f_1=1.05$ and $L_1=75$ dB SPL.

Increasing the frequency ratio above 1.05 also reduced the level of the $2f_2-f_1$ response, especially at low stimulus levels.

The greatest growth rates in $2f_2-f_1$ DPOAE with increasing stimulus level occurred when $f_2/f_1=1.32$, although the response level generally remained lower than corresponding measurements with lower frequency ratios. As with the $2f_1-f_2$ DP, there was little growth in the $2f_2-f_1$ DP when $f_2/f_1=1.05$, especially with $L_2<L_1$, when virtually no response growth occurred.

The $2f_1-f_2$ DP level was always greater than the $2f_2-f_1$ DP level except when $L_1=L_2$ and $f_2/f_1=1.05$, in which case the $2f_2-f_1$ DP exceeded the $2f_1-f_2$ DP by 2–2.5 dB. The $2f_1-f_2$ DP, however, was up to 20 dB greater than the $2f_2-f_1$ DP with large stimulus frequency ratios and when L_2 was reduced below L_1.

B. Comparison of TEOAE and DPOAE levels

The TEOAE responses exceeded the noise within the frequency region under investigation in all ears tested, with the exception of a few isolated points in some subjects. The sum of the TEOAE level within the frequency range varied from -2.0 to 12.6 dB SPL between subjects, with the mean being 6.2 dB SPL. The mean was close to the mean, for example, of the averaged DPOAE sweeps with $L_1=L_2=65$

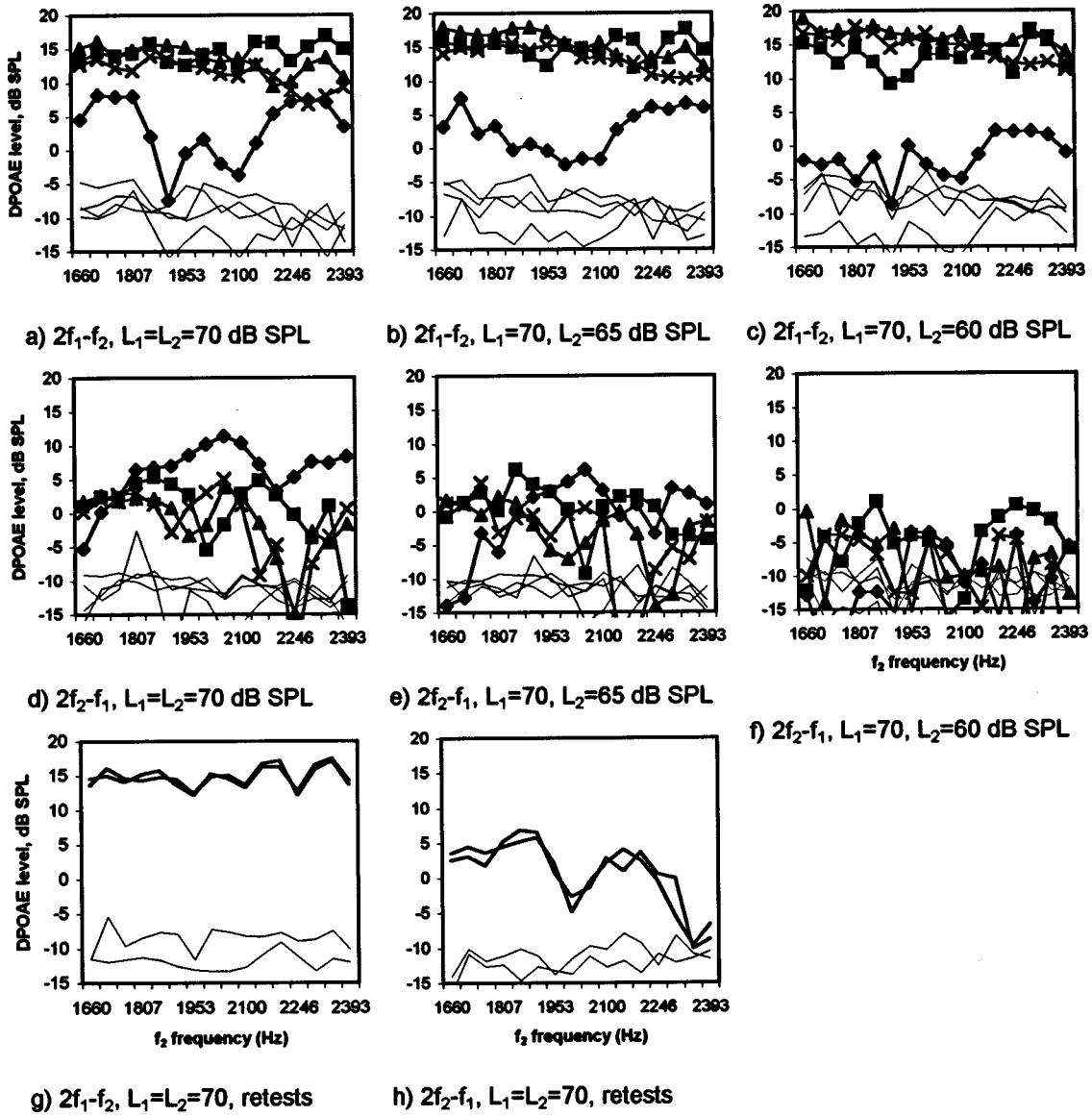


FIG. 1. DPOAE amplitude data for subject *D* with $L_1=70$ dB SPL. ♦ $=f_2/f_1=1.05$, ■ $=f_2/f_1=1.2$, ▲ $=f_2/f_1=1.27$, × $=f_2/f_1=1.32$. (a)–(c) shows the $2f_1-f_2$ DP, (d)–(f) the corresponding $2f_2-f_1$ DP, and (g) and (h) repeat tests with $L_1=L_2=70$ dB SPL and $f_2/f_1=1.2$ from the beginning and end of the DPOAE test sequence. Thin lines show the mean noise estimates from the adjacent ten spectral frequency bands to the DP. The standard deviation associated with each noise estimate was typically 2–3 dB.

dB SPL and $f_2/f_1=1.2$ or 1.27. For each subject, the standard error in the mean TEOAE level was calculated from the set of individual TEOAE data points (this measure is likely to reflect the amount of fine structure, although it could also indicate an underlying smooth trend across the frequency range under test). The average of these was 1.3 dB, comparable with standard errors of all the $2f_2-f_1$ DPOAE measurements and $2f_1-f_2$ DPOAE measurements with $f_2/f_1=1.05$, which were typically in the range 1.2–1.6 dB. The mean $2f_1-f_2$ DPOAE measurements with larger f_2/f_1 had smaller standard errors, mostly in the range 0.5–0.9 dB.

C. Relationships between TEOAE and DPOAE amplitude

For each DP parameter configuration, a scattergram was generated on which the TEOAE total level was plotted against the average DPOAE response for each subject. Examples of the scattergrams are shown in Fig. 3 with best-fit lines calculated by linear regression.

The standard deviation of the DPOAE level predicted from the TEOAE level with respect to the actual DPOAE level (''prediction s.d.'') was calculated for each of the 72 data sets. The calculated values are shown in Table I(a) and (b).

Considering the $2f_1-f_2$ DP, the general trend was for the closest relationships to TEOAE to be with smaller frequency ratios. Paired *t* tests from whole columns of data in Table I(a) indicate that the prediction s.d. values with stimulus frequency ratios of 1.27 or 1.32 were significantly higher ($p<0.05$) than with any lower ratios.

When $f_2/f_1=1.05$ or 1.2, there was also a trend for the prediction s.d. values to be lower with lower stimulus levels. Therefore, the closest amplitude relationship between the

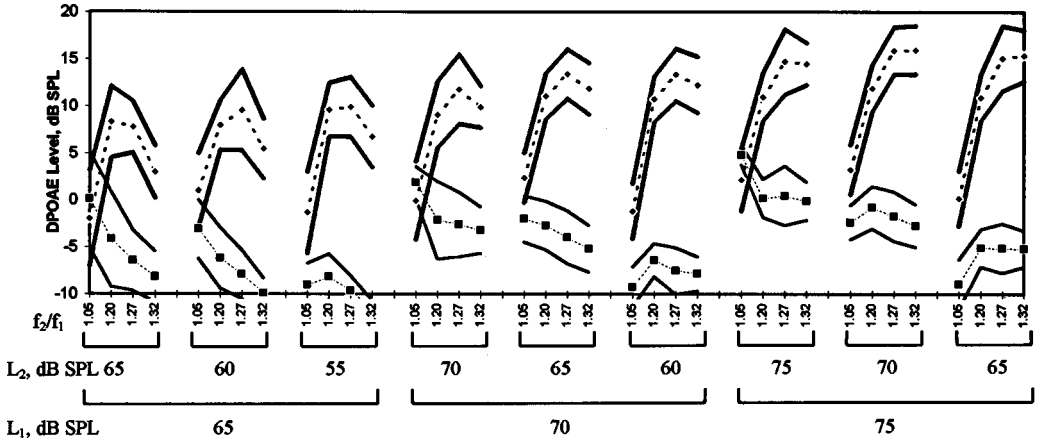


FIG. 2. Average $2f_1-f_2$ and $2f_2-f_1$ DPOAE levels obtained with each combination of stimulus parameters in the frequency range 1660–2393 Hz. Errors shown are 2×s.e. from the mean of nine ears. The average noise floor for each condition was always below −10 dB SPL. ◆ $=2f_1-f_2$ DP, ■ $=2f_2-f_1$ DP.

$2f_1-f_2$ DP and TEOAE occurred when $L_1=65$ dB SPL and $f_2/f_1=1.05$ or 1.2. The lowest single result (1.66 dB) was obtained with $L_1=65$ dB SPL, $L_2=60$ dB SPL, and $f_2/f_1=1.2$.

The $2f_2-f_1$ DP also showed an overall trend to have closest relationships to TEOAE level with lower stimulus frequency ratios. Paired *t* tests from whole columns of data in Table I(b) indicate that the prediction s.d. values obtained with $f_2/f_1=1.05$ are significantly lower ($p<0.05$) than any other frequency ratio. When $L_1=65$ dB SPL, the prediction s.d. values were also lower with L_2 reduced below L_1, although with a very small response level. However, when $L_1=70$ or 75 dB SPL, $L_1=L_2$ produced lower values of prediction s.d. Comparing rows of data from Table I(b), $L_1=L_2=75$ dB SPL gave prediction s.d. values significantly lower ($p<0.005$) than with L_2 reduced by 5 or 10 dB.

When L_1 and L_2 were equal, higher stimulus levels gave closer relationships between the $2f_2-f_1$ DP and TEOAE except when $f_2/f_1=1.05$, in which case there was no clear trend as the relationship was always close. Stimulus levels of $L_1=L_2=65$ dB SPL gave significantly higher values of prediction s.d. than $L_1=L_2=70$ or 75 dB SPL ($p<0.05$). Meanwhile, when $L_1=L_2+10$, a lower stimulus level produced lower prediction s.d. values but with very low level DPOAE's. Comparing rows of data in Table I(b), the stimulus level combination which gave the highest values for prediction s.d. was $L_1=75$ dB SPL and $L_2=65$ dB SPL. The tests indicate a high statistical significance ($p<0.05$) to the difference between this and all other stimulus level combinations except $L_1=L_2=65$ dB SPL.

The $2f_2-f_1$ prediction s.d. values close to 1 dB were obtained with $L_1=L_2=70$ dB SPL and $f_2/f_1=1.05$ and with $L_1=L_2=75$ dB SPL and $f_2/f_1=1.2$. The lowest single value (0.69 dB) was obtained with $L_1=65$ dB SPL, $L_2=55$

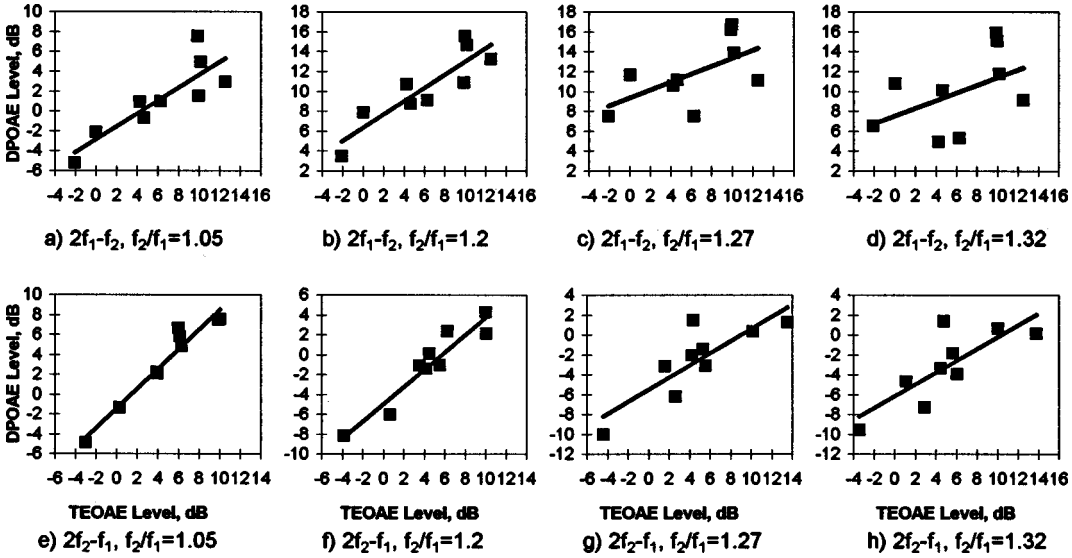


FIG. 3. Comparison of average DPOAE and total TEOAE level in the frequency range 1660–2392 Hz for nine ears, with the best-fit line calculated by linear regression. The TEOAE response is plotted against the f_2 frequency range of the $2f_1-f_2$ distortion product and the emission frequency of the $2f_2-f_1$ distortion product. The examples shown were obtained with stimulus amplitudes $L_1=L_2=70$ dB SPL and frequency ratios: (a) $2f_1-f_2$, $f_2/f_1=1.05$, (b) $2f_1-f_2$, $f_2/f_1=1.2$, (c) $2f_1-f_2$, $f_2/f_1=1.27$, (d) $2f_1-f_2$, $f_2/f_1=1.32$, (e) $2f_2-f_1$ $f_2/f_1=1.05$, (f) $2f_2-f_1$, $f_2/f_1=1.2$, (g) $2f_2-f_1$, $f_2/f_1=1.27$, and (h) $2f_2-f_1$, $f_2/f_1=1.32$. Both the $2f_1-f_2$ and the $2f_2-f_1$ DPs appear more closely related to TEOAE when the stimulus frequency ratio is smaller.

TABLE I. The relationship between DPOAE and TEOAE level indicated by standard deviation and the gradient associated with the best fit line for each configuration of DPOAE parameters. (a) and (b) Standard deviations of (a) $2f_1-f_2$ and (b) $2f_2-f_2$ DPOAE level (dB) predicted from best-fit line relationship with TEOAE (defined as prediction s.d.). Heavy shading indicates prediction s.d.$>$3.0 dB, light shading indicates 1.5$<$prediction s.d.$<$3.0 dB, no shading indicates prediction s.d. below 1.5 dB. (c) and (d) Gradients (ΔDP/ΔTE) of best fit lines between TEOAE and (c) $2f_1-f_2$ and (d) $2f_2-f_1$ DPOAE amplitudes. Heavy shading indicates gradient $<$0.6, light shading indicates 0.6$<$gradient$<$0.8, no shading indicates gradient above 0.8. (e) and (f) Standard errors of the gradients between TEOAE and (e) $2f_1-f_2$ and (f) $2f_2-f_1$ DPOAE levels.

Stim(dB SPL)		1.05	1.2	1.27	1.32
L1	L2	2f1-f2	2f1-f2	2f1-f2	2f1-f2
65	65	1.7	1.8	3.4	4.8
	60	1.8	1.7	3.5	5.1
	55	1.9	1.9	3.4	4.8
70	70	2.1	1.9	2.8	3.7
	65	2.0	1.9	3.3	4.0
	60	2.2	2.3	3.3	3.9
75	75	1.8	1.9	2.9	3.7
	70	2.3	2.4	3.1	3.8
	65	2.5	4.2	3.0	3.4

a) Prediction S.D., $2f_1$-f_2

Stim(dB SPL)		1.05	1.2	1.27	1.32
L1	L2	2f2-f1	2f2-f1	2f2-f1	2f2-f1
65	65	1.4	2.4	2.7	3.9
	60	1.2	2.3	2.1	2.7
	55	0.7	1.9	1.4	2.0
70	70	1.1	1.3	2.2	2.4
	65	1.2	2.1	2.6	1.9
	60	1.6	2.7	2.3	1.7
75	75	1.3	1.0	1.6	1.6
	70	2.1	1.9	2.4	2.8
	65	2.3	2.9	3.0	3.0

b) Prediction S.D., $2f_2$-f_1

Stim(dB SPL)		1.05	1.2	1.27	1.32
L1	L2	2f1-f2	2f1-f2	2f1-f2	2f1-f2
65	65	0.82	0.76	0.56	0.78
	60	0.72	0.81	0.70	0.90
	55	0.78	0.88	0.77	0.88
70	70	0.65	0.67	0.40	0.39
	65	0.63	0.71	0.54	0.48
	60	0.60	0.71	0.65	0.69
75	75	0.68	0.63	0.40	0.40
	70	0.63	0.62	0.51	0.51
	65	0.57	0.67	0.60	0.62

c) Gradient, $2f_1$-f_2

Stim(dB SPL)		1.05	1.2	1.27	1.32
L1	L2	2f2-f1	2f2-f1	2f2-f1	2f2-f1
65	65	1.25	1.04	0.84	0.94
	60	1.12	0.66	0.37	0.64
	55	0.85	0.38	0.06	0.48
70	70	0.99	0.91	0.84	0.81
	65	0.89	0.89	0.75	0.69
	60	0.52	0.60	0.50	0.10
75	75	0.67	0.76	0.68	0.62
	70	0.66	0.84	0.69	0.77
	65	0.51	0.68	0.28	0.54

d) Gradient, $2f_2$-f_1

Stim(dB SPL)		1.05	1.2	1.27	1.32
L1	L2	2f1-f2	2f1-f2	2f1-f2	2f1-f2
65	65	0.12	0.13	0.24	0.34
	60	0.13	0.12	0.25	0.36
	55	0.14	0.13	0.24	0.34
70	70	0.15	0.13	0.20	0.26
	65	0.15	0.13	0.23	0.28
	60	0.16	0.17	0.24	0.28
75	75	0.13	0.14	0.20	0.26
	70	0.16	0.17	0.22	0.27
	65	0.18	0.30	0.21	0.24

e) Gradient standard error, $2f_1$-f_2

Stim(dB SPL)		1.05	1.2	1.27	1.32
L1	L2	2f2-f1	2f2-f1	2f2-f1	2f2-f1
65	65	0.12	0.20	0.19	0.28
	60	0.10	0.19	0.14	0.20
	55	0.06	0.15	0.10	0.14
70	70	0.09	0.11	0.15	0.17
	65	0.10	0.17	0.18	0.13
	60	0.13	0.22	0.16	0.12
75	75	0.11	0.08	0.11	0.11
	70	0.18	0.15	0.17	0.20
	65	0.19	0.23	0.21	0.22

f) Gradient standard error, $2f_2$-f_1

TABLE II. A comparison of prediction s.d. calculated from single frequency points and the average prediction s.d. values drawn from Table I (a) and (b). Stimulus levels are $L_1 = L_2 = 70$ dB SPL.

f2/f1	D.P.	Single Point	Average
1.05	2f1-f2	5.21	2.08
	2f2-f1	4.65	1.09
1.2	2f1-f2	4.22	1.87
	2f2-f1	6.24	1.30
1.32	2f1-f2	4.00	3.71
	2f2-f1	5.45	2.39

dB SPL, and $f_2/f_1 = 1.05$, but this is a parameter combination which may not be very useful to pursue as it gives a very low level $2f_2 - f_1$ DPOAE.

The overall magnitude of the prediction s.d. values obtained for both the $2f_1 - f_2$ and $2f_2 - f_1$ distortion products has been confirmed by an additional study in which five of the DP sweeps were investigated for a further ten subjects and prediction s.d. values of a similar magnitude were obtained.

D. Averaging across frequency and the relationship between TEOAE and DPOAE

Prediction s.d. was also calculated using single point TEOAE and DPOAE data without cross-frequency averaging for three DPOAE stimulus configurations; with $L_1 = L_2 = 70$ dB SPL and $f_2/f_1 = 1.05$, 1.2, and 1.32. The calculation was repeated with successive steps in the frequency alignment between DPOAE and TEOAE, and the minimum prediction s.d. was recorded. The results are shown in Table II.

In general, a much closer relationship was found between TEOAE and DPOAE when the data was averaged across frequency, with the exception of the $2f_1 - f_2$ DP with $f_2/f_1 = 1.32$ in which case the averaging only resulted in a small reduction in prediction s.d. This implies that when the $2f_1 - f_2$ DPOAE was stimulated by a wide frequency ratio, the poor correlation of the OAE fine structures was not the main limiting factor in the prediction s.d. calculated without averaging. With a frequency ratio of 1.05, the prediction s.d. was lowest with both the $2f_1 - f_2$ and $2f_2 - f_1$ DPs when the TEOAE frequency was matched the DP frequency. However, with a frequency ratio of 1.32, the exact frequency alignment was much less critical, suggesting that the fine structure pattern of the wide ratio DPOAE responses differed more from that of the TEOAE responses so that frequency misalignment was less significant. With $f_2/f_1 = 1.2$, the frequency alignment of the $2f_1 - f_2$ DP was of little significance (suggesting that the DPOAE fine structure was not correlated with that of TEOAE), but with the $2f_2 - f_1$ DP the prediction s.d. was lowest when the TEOAE frequency was aligned with the DP frequency.

E. Comparison of $2f_2 - f_1$ and f_2 TEOAE frequency mapping

The prediction s.d. associated with the $2f_2 - f_1$ distortion product was also calculated using the TEOAE component corresponding to the DPOAE f_2 stimulus frequency and compared to the prediction s.d. values shown in Table I(b), which were calculated using the TEOAE component corresponding to the DPOAE $2f_2 - f_1$ emission frequency.

When $f_2/f_1 = 1.05$, the prediction s.d. values were significantly lower ($p < 0.005$) when based on matching the emission frequencies of TEOAE and DPOAE, consistent with the $2f_2 - f_1$ DPOAE level being defined in the region of the emission frequency place on the basilar membrane. However when $f_2/f_1 = 1.27$ or $1.32 (p < 0.005)$, the prediction s.d. was lower when the TEOAE band was matched to the DPOAE f_2 frequency. With a frequency ratio of 1.2, the prediction s.d. values showed a nonsignificant tendency to be lower at the f_2 frequency. It is surprising that the upper sideband DPOAE with $f_2/f_1 = 1.27$ or 1.32 may be related to the f_2 region on the basilar membrane, as the DP is not expected to be able to propagate efficiently along the basilar membrane in this region. It may be that the relevant frequencies of the TEOAE come from a basilar membrane region more basal than their frequency place, and so the $2f_2 - f_1$ DPOAE also corresponds to a region more basal than f_2.

F. Amplitude best-fit line gradients

The gradients of the best fit lines can also be considered to be indicators of the similarities between the DPOAE and TEOAE levels, a gradient of 1 being indicative of a close agreement. This is distinct from a comparison of OAE growth rates, as it is an intersubject measure and each gradient is derived from only one stimulus level for DPOAE and TEOAE. Instead it is a measure of whether intersubject factors causing a difference in TEOAE amplitude result in the same proportional difference in DPOAE amplitude. (In the stimulus parameter ranges used in this study, TEOAE growth rate is normally lower than that of DPOAE.) The level best-fit line gradients are shown in Table I(c) and (d). The standard errors associated with these gradients are shown in Table I(e) and (f).

The best fit line gradients of the level relationship between $2f_1 - f_2$ DPOAE and TEOAE ranged from 0.39 to 0.90 (the gradient of the amplitude relationship is defined as ΔDPOAE/ΔTEOAE), but were generally closer to 1 at lower stimulus levels and smaller stimulus frequency ratios. At larger stimulus frequency ratios, the maximum reduction of L_2 below L_1 also produced gradients closer to 1. Paired t tests of whole rows of data from Table I(c) indicate a strong significance ($p < 0.05$) for the tendency of the gradients with stimulus levels of $L_1 = 65$ dB SPL and $L_2 = 60$ or 55 dB SPL (rows 2 and 3) to be higher than with any combination with $L_1 = 70$ or 75 dB SPL (rows 4–9).

At higher stimulus levels the gradients were higher with smaller stimulus frequency ratios, and comparing whole columns of data, the best fit line gradients with $f_2/f_1 = 1.05$ or 1.2 were significantly higher ($p < 0.05$) than with f_2/f_1

$=1.27$ and the gradients with $f_2/f_1=1.2$ were also significantly higher than with $f_2/f_1=1.32$.

All gradients over 0.8 associated with the $2f_1-f_2$ DP occurred with $L_1=65$ dB SPL but were spread across the frequency ratios and the highest single figure (0.90) occurred with $f_2/f_1=1.32$. The highest best fit line gradients with the $2f_2-f_1$ DP occurred with $L_1=L_2$ [Table I(d)]. Equilevel stimuli at 65 or 70 dB SPL (rows 1 and 4) had significantly higher best-fit line gradients than any combination with a 10 dB stimulus level difference (rows 3, 6, and 9).

Higher gradients also occurred with smaller stimulus frequency ratios, with gradients with $f_2/f_1=1.05$ or 1.2 (columns 1 and 2) being higher than with $f_2/f_1=1.27$ or 1.32 (columns 3 and 4, $p<0.005$). With $f_2/f_1=1.05$, the gradient was also higher at low stimulus levels, but at higher frequency ratios this only remained true when $L_1=L_2$.

Although the highest gradients occurred with lower stimulus levels and $f_2/f_1=1.05$, there were also some gradients over 0.9 with $L_1=L_2=65$ or 70 dB SPL at the larger stimulus frequency ratios.

G. Comparison of repeat DPOAE measurements

Verification that the response level was consistent throughout the test sequence was achieved by repeating one DP sweep at the start and end of the DPOAE test series as well as in sequence halfway through the series. The repeated test used the following stimulus parameters: $L_1=L_2=70$ dB SPL and $f_2/f_1=1.2$. For each subject the levels of DPOAE were determined for the first and last tests and the differences between the two measurement levels were calculated. The standard deviation of the differences for the $2f_1-f_2$ DPOAE was 0.9 dB, whereas for the $2f_2-f_1$ DPOAE it was 1.4 dB.

The higher variations seen in the $2f_2-f_1$ distortion product are likely to result from a greater influence of noise, as the two distortion products were measured simultaneously and the greatest retest differences for the $2f_2-f_1$ distortion product were seen in subjects with the lowest amplitude responses. There was no significant trend for the DPOAEs to increase or decrease during the test sequence, although the $2f_2-f_1$ distortion product showed a slight tendency to have a higher level in the later test (mean difference 0.4 dB).

H. Phase gradients and delays

The results for all subjects (Fig. 4) fell into two distinct categories, with the $2f_1-f_2$ DP at wide frequency ratios producing very shallow phase gradients and hence a small group delay while the $2f_1-f_2$ DP with $f_2/f_1=1.05$ and all $2f_2-f_1$ DP measurements produced steep phase gradients, similar to the TEOAE phase.

The TEOAE group delays in Fig. 4(b) are calculated in the frequency band corresponding to the DPOAE f_2 frequency range and also at the frequencies corresponding to the $2f_2-f_1$ emission frequency for each of the stimulus frequency ratios. The TEOAE group delays corresponding to the $2f_2-f_1$ DP frequency with wider stimulus frequency ratios are slightly shorter as they are derived from slightly higher frequency bands.

The $2f_2-f_1$ results with $L_1=65$, $L_2=55$, and $f_2/f_1=1.27$ and 1.32 are based on one subject because of low level responses resulting in a failure to meet the signal-to-noise ratio criteria in at least three consecutive frequencies in most subjects. All other data is based on at least five subjects.

The phase of the DPOAE and TEOAE data from one subject is shown in Fig. 5(a)–(d). The phase gradient of the $2f_1-f_2$ DP at stimulus frequency ratios of 1.2 and greater was considerably less than that of all other DPOAE and the TEOAE measurements.

The behavior of the $2f_1-f_2$ DP phase confirms the earlier findings of Kemp (1986). In the case of the $2f_2-f_1$ DP meanwhile, the steep phase gradient which is seen only at small frequency ratios in the $2f_1-f_2$ DP is retained out to ratios of at least $f_2/f_1=1.32$.

III. DISCUSSION OF RESULTS

A. DPOAE amplitudes

1. $2f_1-f_2$

The amplitudes of the $2f_1-f_2$ DPOAE measurements passed through a maximum at a stimulus frequency ratio which increased from 1.2 to 1.32 as the stimulus level increased. This finding is in agreement with results presented by Harris *et al.* (1989) and other workers.

All stimulus configurations used in the test sequence gave responses which exceeded the mean noise by at least $1\times$ the s.d. with the exception of a few isolated points. However, measurements with ratios of stimulus frequency 1.2, 1.27, and 1.32 gave larger amplitude responses than the 1.05 ratio and were therefore quicker and easier to measure.

The lower growth rates with stimulus level seen in the DPOAE measurements with a stimulus frequency ratio of 1.05 imply that the generation or emission mechanisms may be more saturated. The relatively low level of response indicates that under these stimulus conditions either the distortion product does not easily escape back to the ear canal or the response induced in the basilar membrane produces less energy at that DP frequency.

Conversely, the high growth rates seen with a frequency ratio of 1.32 (of up to 1.4 dB/dB) suggest that the DP generation or emission mechanisms for wider spaced tones are not so easily saturated leading to the higher DP level.

2. $2f_2-f_1$

The levels of the $2f_2-f_1$ DPOAEs are sensitive to changes in frequency ratio and differences between L_1 and L_2, particularly at lower stimulus levels, with the highest level responses being obtained with smaller frequency ratios and equal stimulus levels. This is the opposite of the $2f_1-f_2$ DP behavior, for which a larger frequency ratio and $L_1>L_2$ increased the OAE level. The trend shown by Erminy *et al.* (1996) for a larger $2f_2-f_1$ DP level to be obtained with $f_2/f_1=1.12$ than with $f_2/f_1=1.22$ has been shown to continue at both higher and lower ratios.

As with the $2f_1-f_2$ distortion product, the $2f_2-f_1$ DP growth rates were smallest when $f_2/f_1=1.05$, implying a

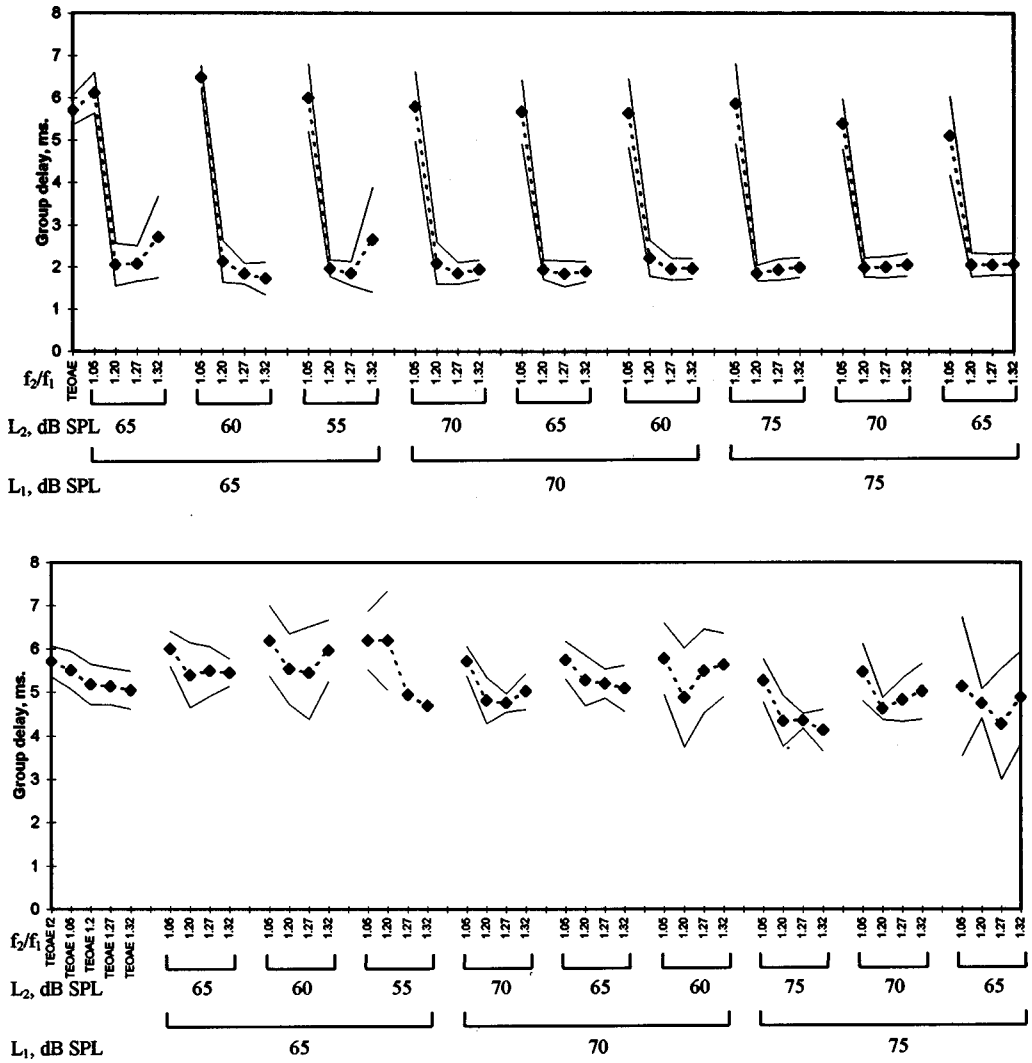


FIG. 4. Group delay, derived from the DPOAE phase gradients obtained from sweeps with constant f_2/f_1. Results are an average from nine subjects. Errors shown are 2×s.e. from the mean. The corresponding TEOAE data is shown for comparison: (a) $2f_1-f_2$ DP and (b) $2f_2-f_1$ DP. In (b), the TEOAE data is shown corresponding to the f_2 frequency of the DP measurements and also the frequency of the $2f_2-f_1$ DP at each frequency ratio. Group delay is considerably lower for the $2f_1-f_2$ DP with $f_2/f_1=1.2-1.32$ than with the $2f_1-f_2$ DP with $f_2/f_1=1.05$ or the $2f_2-f_1$ DP or TEOAE.

more saturated generation or emission mechanism, and greatest when $f_2/f_1=1.32$, implying a less saturated response.

B. Relationships between TEOAE and DPOAE amplitude

1. $2f_1-f_2$

A close relationship was found between the frequency averaged levels of the $2f_1-f_2$ DPOAE and TEOAE, particularly when the frequency ratio was small.

With larger stimulus ratios, there was a tendency for the prediction s.d. to be smaller with higher stimulus levels, a condition which coincided with lower DPOAE versus TEOAE level gradients [Table I(c)]. A smaller DPOAE versus TEOAE level gradient would reduce the values obtained for prediction s.d. and so the lower prediction s.d. values obtained with higher stimulus levels may not in this case indicate a closer relationship between DPOAE and TEOAE. The relationships found in the present study are closer than expected from literature. Calculations from data presented by Smurzinsky and Kim (1992) and Smurzinsky *et al.* (1993) produced approximate $2f_1-f_2$ prediction s.d. values at 2 kHz of 3.7 dB (preterm infants), 3.6 dB (fullterm infants), and 2.7 dB (adults) using DPOAE stimulus parameters of $L_1=L_2=65$ dB SPL, $f_2/f_1=1.18-1.23$ with $\frac{1}{4}$-octave steps between points ($\frac{1}{2}$-octave averaging in analysis). It is likely that a closer agreement was seen between TEOAE and DPOAE with adult subjects than with infants because of easier measurement conditions. A difference remains between the adult prediction s.d. (2.7 dB) and the corresponding result using a comparable stimulus frequency ratio and level from the present study (1.8 dB). This is likely to be because of the finer frequency interval within the DP sweeps of the present study and hence the greater degree of across frequency averaging and also the use of the same probe fitting for both DPOAE and TEOAE measurements in the present study resulting in the removal of possible variance arising from variations in probe fitting.

The corresponding calculated prediction s.d. associated with the work of Moulin *et al.* (1993) and Probst and Harris

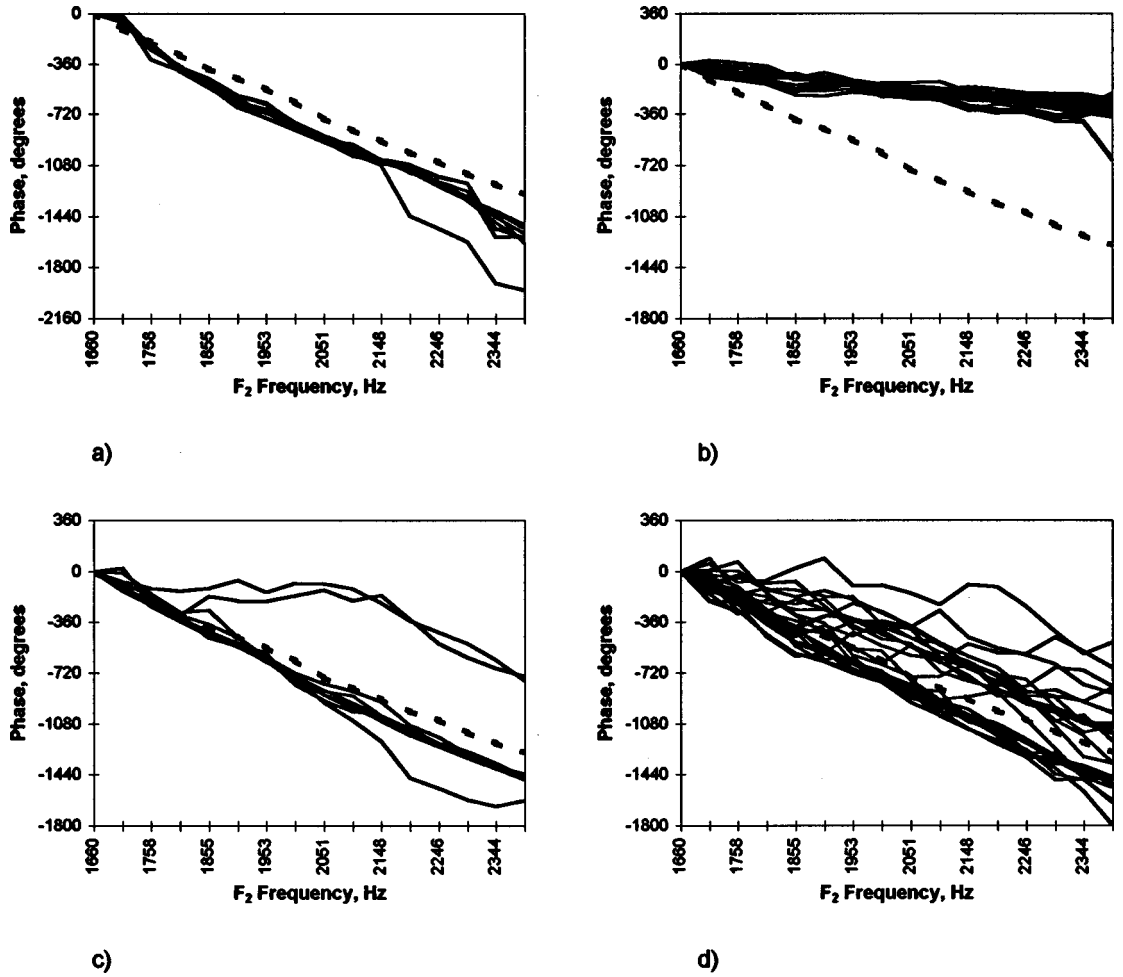


FIG. 5. The DPOAE phase (solid lines) plotted with TEOAE phase (broken lines) corresponding to the f_2 frequency of the DPOAE measurements for one subject. (a) 2 f_1-f_2 DP, $f_2/f_1=1.05$. (b) 2 f_1-f_2 DP, $f_2/f_1=1.2$, 1.27, and 1.32. (c) 2 f_2-f_1 DP, $f_2/f_1=1.05$. (d) 2 f_2-f_1 DP, $f_2/f_1=1.2$, 1.27, and 1.32. The gradient of the 2 f_1-f_2 DP phase in (b) is considerably shallower than all other DP data and the TEOAE. Scattering seen in (d) occurs in data with a poor signal-to-noise ratio.

(1993) are 4.1 and 3.9 dB, respectively, which correspond to prediction s.d. values of 1.8 and 1.7 dB obtained with comparable stimulus parameters in the present study. Moulin *et al.* used DPOAE stimulus parameters of $L_1=L_2=60$ dB SPL and $f_2/f_1=1.17$ and compared unaveraged single DPOAE points to 200-Hz averaged TEOAE bands. The larger prediction s.d. seen in the Moulin data presumably arises from the reduced averaging in both the TEOAE and DPOAE data. Probst and Harris used a frequency ratio of 1.22 and a stimulus level difference of 6 dB SPL and compared the TEOAE and DPOAE levels across the full frequency range and so the TEOAE and DPOAE mean responses may have been dominated by different frequency regions.

Therefore, it appears that there are two main reasons why closer relationships have been observed between the levels of TEOAE and DPOAE in the present study than had been previously shown:

(a) The averaging process across a frequency range removes to a large extent the effect of the OAE fine structure, and
(b) the DPOAE configurations which give the closest agreements are not those which have the largest amplitude and so have not previously been investigated so extensively.

2. Relationships between TEOAE and DPOAE amplitude, 2 f_2-f_1

As with the 2 f_1-f_2 DP, the 2 f_2-f_1 DP was also most closely related to TEOAE when evoked by smaller frequency ratios and stimulus levels, although the relationship did not deteriorate at wider DP stimulus frequency ratios to as great an extent as was the case with the 2 f_1-f_2 distortion product. Interestingly, although when $L_1=70$ or 75 dB SPL it was preferable for the stimuli to be of equal level, when $L_1=65$ dB SPL the values of prediction s.d. were lower when $L_1>L_2$. However, these have very low levels and are associated with very shallow amplitude gradient relationships and, therefore, in this case the association with TEOAE may not be as close as it may appear.

C. Comparison of 2 f_1-f_2 and 2 f_2-f_1 prediction s.d. values

A comparison of the average 2 f_1-f_2 and 2 f_2-f_1 prediction s.d.'s in Table I(a) and (b) at each frequency ratio is shown in Table III.

TABLE III. A comparison of the average prediction s.d. of the $2f_1-f_2$ and $2f_2-f_1$ distortion products for each frequency ratio from Table I(a) and (b). A positive difference indicates that the $2f_1-f_2$ prediction s.d. values were larger than the corresponding $2f_2-f_1$ values. *P* values are derived via a paired *t* test.

	Stimulus frequency ratio			
DPOAE	1.05	1.2	1.27	1.32
2f1-f2	2.04	2.21	3.18	4.12
2f2-f1	1.42	2.05	2.25	2.44
Difference	0.62	0.16	0.93	1.68
p value	<0.001	≈0.25	<0.001	<0.001

The data confirms the apparent trend seen in Table I(a) and (b) that the $2f_2-f_1$ distortion product measurements were more closely related to the TEOAE level than the $2f_1-f_2$ distortion product. The differences were statistically highly significant with frequency ratios of 1.05, 1.27, and 1.32.

D. Gradient of DPOAE/TEOAE amplitude relationship

Gradients close to 1 imply that intersubject differences affecting OAE level act in equal proportion on both DPOAE and TEOAE, thus suggesting similarities in generation and/or emission. There are various factors which could effect the amplitudes of TEOAE and DPOAE. Variation in the efficiency of sound conduction through the middle ear would maintain a 1:1 relationship between TEOAE and DPOAE and would probably affect TEOAE and DPOAE level equally. Changing the DP parameters may change the DPOAE level, but would not necessarily of itself alter the 1:1 gradient unless the change led to unequal saturation of TEOAE and DPOAE. However, if DPOAE and TEOAE were of unequal sensitivity to the condition of the outer hair cells or if they arose from different emission mechanisms, the 1:1 gradient would probably be lost. As both DPOAE and TEOAE measured with the nonlinear stimulus method are both a result of nonlinear cochlear processes, their amplitudes could be expected to behave similarly.

1. Gradient of DPOAE/TEOAE amplitude relationship, $2f_1-f_2$

The DPOAE/TEOAE amplitude gradients associated with the $2f_1-f_2$ DP results were generally closer to 1 with lower stimulus levels and when $L_1>L_2$ [Table 1(c)]. This indicates that inter subject factors affecting OAE level have a proportionate effect on DPOAE and TEOAE with these stimulus conditions, implying that the mechanisms involved in generating the two OAE types are related. At higher stimulus levels there was also a tendency for the gradients to be higher with smaller frequency ratios, in spite of the lower DP growth rates associated with these stimulus parameters. These findings generally agree with the prediction s.d. data with the exception of the level difference: Reducing L_2 tended to increase the best-fit line gradient (implying closer agreement), but also increased the prediction s.d. (implying worse agreement) when f_2/f_1 was small.

2. Gradient of DPOAE/TEOAE amplitude relationship, $2f_2-f_1$

The $2f_2-f_1$ gradients were near to 1 with low stimulus levels, small frequency ratios, and $L_1=L_2$ and were lower with other DPOAE stimulus parameters. The pattern of results is therefore consistent with the prediction s.d. data. The tendency for the gradients obtained with a low stimulus level to be high may be a result of avoidance of a saturated DP response, whereas high gradients obtained with small frequency ratios and $L_1=L_2$ are associated with higher DPOAE levels and may therefore arise from the avoidance of the influence of noise.

Comparing the average DPOAE/TEOAE amplitude gradients of the two distortion products at each frequency ratio, the gradients associated with the $2f_2-f_1$ DP were closer to 1 than those of the $2f_1-f_2$ DP when $L_1=L_2$, but the $2f_1-f_2$ DP gradients were closer to 1 when $L_1=L_2+10$.

Considering the prediction s.d. and amplitude gradients together, both the $2f_1-f_2$ DP and the $2f_2-f_1$ DP levels appear to be generally most closely related to TEOAE with smaller frequency ratios and lower stimulus levels. Additionally, the $2f_2-f_1$ DP is also more closely related to TEOAE when $L_1=L_2$.

E. Retest error

The test–retest difference for the $2f_1-f_2$ distortion product was found to be 0.9 dB, so the values of prediction s.d. would not reliably fall below this even if there is no difference at all in the mechanisms involved in the TEOAE and DPOAE. Additionally, if a similar test–retest error is present in the TEOAE measurements, the combined effect would result in a prediction s.d. floor of 1.25 dB. Several of the $2f_2-f_1$ DP test configurations have produced prediction s.d. values which are of this order, and so there may not be any significant difference between the TEOAE measurement form and some configurations of DPOAE.

Possible additional random error arises from the possibility that the fine structure ripple may not be completely eliminated by the frequency averaging over only approximately $\frac{1}{2}$ of an octave and may not match perfectly between the frequency regions adopted from the TEOAE and DPOAE responses.

F. Phase gradients and delays

Derivation of the group delay from the phase data associated with the TEOAE and DPOAE measurements (Figs. 4 and 5) has indicated that $2f_1-f_2$ measurements with $f_2/f_1=1.05$ and all $2f_2-f_1$ measurements have phase gradients close to that of the corresponding TEOAE measurements, implying that the mechanisms by which the two forms of OAE are emitted are related. Stimulus frequency otoacoustic emissions (SFOAE) are also known to have similar steep phase versus frequency gradients (Shera and

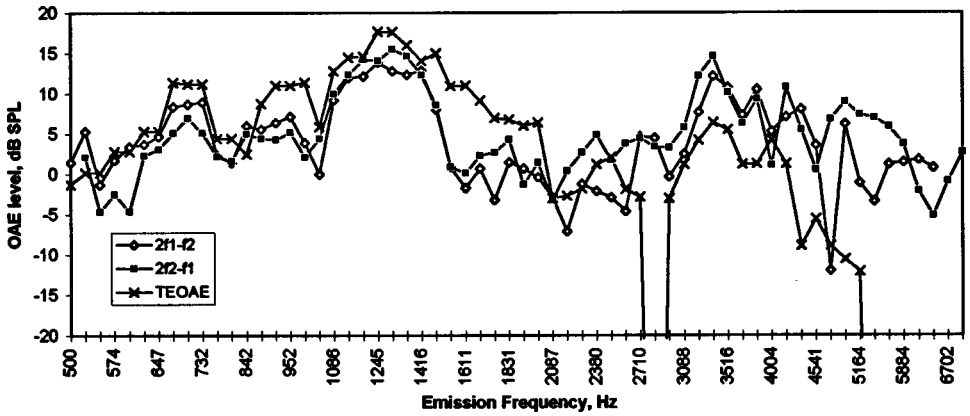


FIG. 6. Extended fine structure DP grams with corresponding TEOAE response on a logarithmic frequency scale. Each OAE is plotted against its emission frequency. The TEOAE amplitude is increased by 12 dB to correspond with the summation method used elsewhere in this study. The DPOAE stimulus parameters are $L_1 = L_2 = 70$ dB SPL $f_2/f_1 = 1.05$. Close agreement is seen in the fine structure of the three OAE forms.

Guinan, 1999). However, the phase gradients of the $2f_1 - f_2$ distortion product with larger frequency ratios were considerably reduced, suggesting that these emissions differ fundamentally from TEOAE.

G. Extension to full frequency range

In order to investigate the relationship of DPOAE to TEOAE across a full frequency range with a small DPOAE frequency ratio, an extended fine structure DP sweep was obtained with stimulus parameters $f_2/f_1 = 1.05$ and $L_1 = L_2 = 70$ dB SPL for one subject covering the frequency range 500–7000 Hz. The $2f_1 - f_2$ and $2f_2 - f_1$ DPs are shown in Fig. 6 with a corresponding standard TEOAE, with each response plotted against the frequency of the emitted response on a logarithmic scale. The level of the TEOAE response in Fig. 6 was increased by 12 dB to correspond with the summation method used in the averaging of the responses earlier in this study.

It can be seen that both of the distortion products shown were measurable across the complete frequency range, whereas as expected the TEOAE response level decreased sharply above 4 kHz as a result of the time windowing to remove low latency artifacts. The behavior of the $2f_2 - f_1$ distortion product with these parameters therefore differs from the wider frequency ratios used by Erminy *et al.* (1996), who observed a reduction in $2f_2 - f_1$ level towards higher frequencies. It can also be seen that the fine structures of the three responses show excellent agreement, indicating that with these stimulus parameters the fine structure of each OAE is determined by the emitted frequency rather than the stimulating frequency.

He and Schmiedt (1997) have shown that the fine structure of the $2f_1 - f_2$ DPOAE shifts in frequency with changes in stimulus level at frequency ratios of $f_2/f_1 = 1.11$, 1.2, and 1.33. Therefore, the finding presented here indicates that either the frequency shifting effect does not occur when $f_2/f_1 = 1.05$ or that the DPOAE parameters chosen coincidentally are subject to the same frequency shift as TEOAE, allowing the accurate relationship with TEOAE to occur.

IV. ORIGINS OF TEOAE AND DPOAE

The frequency ratio which produced the maximum level of the $2f_1 - f_2$ DPOAE was in the region of $f_2/f_1 = 1.2$–1.3 and it is largely for this reason that ratios between 1.2 and 1.3 have tended to be preferred for clinical use as the larger response results in a clearer differentiation between normal and hearing impaired ears.

In the case of the $2f_2 - f_1$ DP, the largest levels were obtained with a small stimulus frequency ratio and with $L_1 = L_2$. This pattern could be expected as a small frequency ratio brings the $2f_2 - f_1$ frequency closer to the stimulus frequencies where the transverse displacement on the basilar membrane due to the traveling wave is greater, while reducing L_2 below L_1 may reduce the travelling wave modulation depth in the $2f_2 - f_1$ frequency region and so reduce the effect of nonlinear distortion associated with the basilar membrane movement. Indeed, it is a general property of compressive nonlinearities that upper sideband distortion product generation is enhanced if L_2 exceeds L_1 by a small amount (Kemp, 1987).

The $2f_2 - f_1$ DP growth with stimulus level was faster with larger frequency ratios. This may be because the $2f_2 - f_1$ DP frequency is moved further to the base from the primary tones on the basilar membrane with larger stimulus ratios to a region where the primary tone traveling wave amplitudes grow linearly with increasing stimulus level as opposed to the compressive growth seen in the basilar membrane motion in the region of the traveling wave peak.

The amplitude relationships between TEOAE and DPOAE in some stimulus configurations have been found to be unexpectedly close despite the many obvious differences in the manner in which each emission is stimulated and extracted, supporting the view that with these stimulus conditions the same information is provided by results obtained by either OAE method and that the same underlying mechanism is responsible for both.

Both the $2f_1 - f_2$ and $2f_2 - f_1$ DP levels were more closely related to TEOAE when stimulated with a frequency ratio of $f_2/f_1 = 1.05$ than with the wider ratios of 1.2–1.32. Therefore DPOAE is more like TEOAE when stimulated with a small frequency ratio in which the emission frequency is close to that of the stimuli than when stimulated with a wide frequency ratio in which the emission frequency differs more markedly from the stimulus frequencies.

The amplitude relationship between TEOAE and DPOAE was closer with the $2f_1 - f_1$ DP than the $2f_1 - f_2$ DP. There are fundamental differences between these two DPOAEs which may indicate why the $2f_2 - f_1$ DP is more closely related to TEOAE. (1) The $2f_2 - f_1$ DPOAE genera-

tion site must be linked to the emission frequency place on the basilar membrane rather than the region of the stimulus frequencies. (2) The $2f_2-f_1$ DPOAE has only one possible emission region whereas the $2f_1-f_2$ DPOAE has two. (3) The phase behavior of the $2f_1-f_2$ and $2f_2-f_1$ DPOAEs differs.

Close agreement between the TEOAE and the $2f_2-f_1$ DP and also the low ratio $2f_1-f_2$ DP was seen in the phase data. The shallower phase gradient seen in the $2f_1-f_2$ DP at wider stimulus frequency ratios suggests that the $2f_1-f_2$ DPOAE may have a different mechanism of generation and/or emission depending on the ratio f_2/f_1.

The two distinct phase gradient categories are consistent with the two emission creating modes described as ''place fixed'' and ''wave fixed'' (e.g., Kemp, 1986; O Mahoney and Kemp, 1995; Kemp and Knight, 1999) or ''reflection'' and ''distortion'' (Shera and Guinan, 1998, 1999). The wave-fixed model, in which the emission source is locked to the traveling wave envelopes and so moves smoothly with frequency, would produce an almost flat phase gradient because of the approximate geometric frequency scaling of the cochlea. This is seen in the wide ratio $2f_1-f_2$ DP sweeps. The place fixed model, which requires a series of fixed impedance irregularities on the basilar membrane to reflect or scatter OAE energy back towards the base of the cochlea, would result in a comparatively steep phase gradient in a constant frequency ratio sweep, formed by the traveling wave envelopes of the stimulus tones passing over the reflectors as the stimulus frequencies change. This steep phase gradient behavior is consistent with TEOAE, SFOAE, the $2f_2-f_1$ DP, and the $2f_1-f_2$ DP behavior when f_2/f_1 is small.

Yates and Withnell (1998) have suggested that TEOAE may consist mostly of intermodulation distortion components, but it nevertheless has the phase characteristic of the reflection-based place-fixed model and therefore is not dominated by wide stimulus ratio, low-phase gradient, lower sideband components. If TEOAEs included a component like the $2f_1-f_2$ DP at wide ratios, this would be a low-latency component. It remains possible that low-latency components of the TEOAE exist, but are lost in the time windowing employed to extract the OAE from the stimulus and middle ear response.

The results raise the question of whether it would be valuable to change the DPOAE parameters used in clinical applications in order for the response to fall within the place fixed category. However, the issue of whether the wave fixed or place fixed category of OAE is more closely linked to auditory function remains unresolved, and so currently the only reason to change DP parameters would be to allow reliable interchangeability between DPOAE and TEOAE measurements. This would be at the expense of a loss of DPOAE intensity, resulting in a less efficient DPOAE measurement.

V. CONCLUSIONS

A closer agreement has been observed between TEOAE and DPOAE amplitude than has been previously presented.

This is thought to be because

(a) averaging over a frequency range has been used to reduce the variance which arises from the fine structure in DPOAE and TEOAE, and
(b) DPOAE parameters which are seldom used clinically have been found to produce the closest relationships.

The phase gradients of the TEOAE, $2f_2-f_1$ DPOAE, and $2f_1-f_2$ DPOAE with $f_2/f_1=1.05$ are similar and are consistent with a place-fixed reflection-based emission mechanism. Also, SFOAE is known to have a steep phase versus frequency characteristic. The phase of the $2f_1-f_2$ DP with a larger stimulus frequency ratio deviates markedly from this and instead is consistent with a wave fixed emission mechanism. It therefore seems that the DPOAE parameters chosen for clinical use result in an emission which is different from all other evoked OAEs.

ACKNOWLEDGMENTS

The authors would like to thank Peter Bray and Otodynamics Ltd. for customizing the software and each of the subjects who generously gave their time.

Dorn, P. A., Piskorski, P., Gorga, M. P., Neely, S. T., and Keefe, D. H. (**1998**). ''Predicting audiometric status from distortion product otoacoustic emissions using discriminant analysis,'' Association for Research in Otolaryngology Abstracts 594.

Erminy, M., Avan, P., and Baril, P., and Bonfils, P. (**1996**). ''Caractéristiques du produit de distorsion $2f_2-f_1$ chez l'homme normo-entendant,'' Ann. Otolaryngol. Chir. Cervicofac. **113**, 321–327.

Gaskill, S. A., and Brown, A. M. (**1990**). ''The behavior of the acoustic distortion product, $2f_1-f_2$, from the human ear and its relation to auditory sensitivity,'' J. Acoust. Soc. Am. **88**, 821–839.

Gorga, M. P., Neely, S. T., Bergman, B. M., Beaućhaine, K. L., Kaminski, J. R., Peters, J., Schulte, L., and Jesteadt, W. (**1993**). ''A comparison of transient-evoked and distortion product otoacoustic emissions in normal-hearing and hearing-impaired subjects,'' J. Acoust. Soc. Am. **94**, 2639–2648.

Gorga, M. P., Neely, S. T., Ohlrich, B., Hoover, B., Redner, J., and Peters, J. (**1997**). ''From laboratory to clinic: a large scale study of distortion product otoacoustic emissions in ears with normal hearing and ears with hearing loss,'' Ear Hear. **18**(6), 440–455.

Harris, F. P., Lonsbury-Martin, B. L., Stagner, B. B., Coats, A. C., and Martin, G. K. (**1989**). ''Acoustic distortion products in humans: systematic changes in amplitude as a function of f_2/f_1 ratio,'' J. Acoust. Soc. Am. **85**, 220–229.

Hauser, R., and Probst, R. (**1991**). ''The influence of systematic primary-tone level variation L_2-L_1 on the acoustic distortion product emission $2f_1-f_2$ in normal human ears,'' J. Acoust. Soc. Am. **89**, 280–286.

He, N., and Schmiedt, R. A. (**1997**). ''Fine structure of the $2f_1-f_2$ acoustic distortion product: effects of primary level and frequency ratios,'' J. Acoust. Soc. Am. **101**, 3554–3565.

Kemp, D. T. (**1978**). ''Stimulated acoustic emissions from within the human auditory system,'' J. Acoust. Soc. Am. **64**, 1386–1391.

Kemp, D. T. (**1986**). ''Otoacoustic emissions, travelling waves and cochlear mechanisms,'' Hearing Res. **22**, 95–104.

Kemp, D. T. (**1987**). ''Otoacoustic emission characteristics during moderate continuous stimulation-intermodulation,'' 12th International Congress of Acoustics B7-1.

Kemp, D. T. (**1998**). ''Otoacoustic emissions: distorted echoes of the cochlea's travelling wave,'' in Otoacoustic Emissions Basic Science and Clinical Applications, edited by C. I. Berlin (Singular, San Diego), Chap. 1, pp 1–59, fig. 1-12.

Kemp, D. T., and Brown, A. M. (**1983**). ''An integrated view of cochlear mechanical nonlinearities observable from the ear canal,'' in *Cochlear Mechanics*, edited by E. DeBoer and M. A. Viergever (Delft U. P., Delft, Holland).

Kemp, D. T., and Knight, R. D. (**1999**). ''Virtual DP reflector explains DPOAE wave and place fixed dichotomy,'' Association for Research in Otolaryngology Abstracts **22**, 99–100, number 396.
Kemp, D. T., Ryan, S., and Bray, P. (**1990**). ''A guide to the effective use of Otoacoustic Emissions,'' Ear Hear. **11**(2), 93–105.
Kemp, D. T., Bray, P., Alexander, L., and Brown, A. M. (**1986**). ''Acoustic emission cochleography-practical aspects,'' Scand. Audiol. Suppl. **25**, 71–95.
Knight, R. D., and Kemp, D. T. (**1999**). ''f_1, f_2 area representations of DPOAEs indicate that the clinical $2f_1$-f_2 DPOAE is an atypical form of OAE,'' Br J. Audiol. **33**, 87.
Martin, G. K., Jassir, D., Stagner, B. B., Whitehead, M. L., and Lonsbury-Martin, B. L. (**1998**). ''Locus of generation for the $2f_1-f_2$ vs $2f_2-f_1$ distortion-product otoacoustic emissions in normal-hearing humans revealed by suppression tuning, onset latencies, and amplitude correlations,'' J. Acoust. Soc. Am. **103**, 1957–1971.
Martin, G. K., Lonsbury-Martin, B. L., Probst, R., Scheinin, S. A., and Coats, A. C. (**1987**). ''Acoustic distortion products in rabbit ear canal. II. Sites of origin revealed by suppression contours and pure tone exposures,'' Hearing Res. **28**, 191–208.
Moulin, A., Bera, J. C., and Collet, L. (**1994**). ''Distortion product otoacoustic emissions and sensorineural hearing loss,'' Audiology **33**, 305–326.
Moulin, A., Collet, L., Veuillet, E., and Morgon, A. (**1993**). ''Interrelations between transiently evoked otoacoustic emissions and acoustic distortion products in normally hearing subjects,'' Hearing Res. **65**, 216–233.
Moulin, A., and Kemp, D. T. (**1996**). ''Multicomponent acoustic distortion product otoacoustic emission phase in humans. I. General characteristics,'' J. Acoust. Soc. Am. **100**, 1617–1639.
O Mahoney, C. F., and Kemp, D. T. (**1995**). ''Distortion product otoacoustic emissions,'' Arch. Otolaryngol. Head Neck Surg. **119**, 858–860.
Prieve, B. A., Gorga, M. P., Schmidt, A., Neely, S., Peters, J., Schules, L., and Jesteadt, W. (**1993**). ''Analysis of transient-evoked otoacoustic emissions in normal hearing and hearing-impaired ears.'' J. Acoust. Soc. Am. **93**, 3308–3319.
Probst, R., and Harris, R. (**1993**). ''Transiently evoked and distortion-production otoacoustic emissions,'' Arch. Otolaryngol. Head Neck Surg. **119**, 858–860.
Robinette, M. S., and Glattke, T. J. (eds.) (**1997**). *Otoacoustic Emissions: Clinical Applications* (Thiele, New York).
Shera, C. A., and Guinan, J. J. (**1998**). ''Reflection emissions and distortion products arise by fundamentally different mechanisms,'' Association for Research in Otolaryngology Abstracts 344.
Shera, C. A., and Guinan, J. J. (**1999**). ''Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs,'' J. Acoust. Soc. Am. **105**, 782–798.
Smurzinsky, J., and Kim, D. O. (**1992**). ''Distortion-product and click-evoked otoacoustic emissions of normally-hearing adults,'' Hearing Res. **58**, 227–240.
Smurzinsky, J., Jung, M. D., Lafreniere, D., Kim, D. O., Kamath, M. V., Rowe, J. C., Holman, M. C., and Leonard, G. (**1993**). ''Distortion-product and click-evoked otoacoustic emissions of preterm and full-term infants,'' Ear Hear. **14**(4), 258–274.
Stover, L., Gorga, M. P., Neely, S. T., and Montoya, D. (**1996**). ''Toward optimising the clinical utility of distortion product otoacoustic emission measurements,'' J. Acoust. Soc. Am. **100**, 956–967.
Wable, J., Collet, L., and Chéry-Croze, S. (**1996**). ''Phase delay measurements of distortion product otoacoustic emissions at $2f1-f2$ and $2f2-f1$ in human ears,'' J. Acoust. Soc. Am. **100**, 2228–2235.
Whitehead, M. L., McCoy, M. J., Lonsbury-Martin, B. L., and Martin, G. K. (**1995**). ''Dependence of distortion-product otoacoustic emissions on primary levels in normal and impaired ears. I. Effects of decreasing L_2 below L_1,'' J. Acoust. Soc. Am. **97**, 2346–2358.
Wilson, J. P. (**1980**). ''Evidence for a cochlear origin for acoustic re-emissions, threshold fine-structure and tonal tinnitus,'' Hearing Res. **2**, 233–252.
Yates, G. K., and Withnell, R. H. (**1998**). ''Intermodulation distortion in click-evoked otoacoustic emissions,'' Association for Research in Otolaryngology Abstracts 17.

Effects of aspirin on psychophysical measures of frequency selectivity, two-tone suppression, and growth of masking[a)]

Michelle L. Hicks[b)] and Sid P. Bacon
Psychoacoustics Laboratory, Department of Speech and Hearing Science, Arizona State University, P.O. Box 871908, Tempe, Arizona 85287-1908

(Received 16 December 1998; revised 18 March 1999; accepted 4 June 1999)

Three psychophysical measures of nonlinearity were evaluated before and during a course of aspirin ingestion to investigate the role of outer hair cells (OHCs) in these measures, as aspirin is thought to alter the functioning of OHCs. Six normal-hearing individuals received a moderate dose (3.9 g/day) of aspirin for four days, producing essentially flat, temporary hearing losses that ranged from 5–20 dB. The losses were about 2 dB greater for a 300-ms signal than for a 15-ms signal, indicating reduced temporal integration with aspirin. On the final three days of aspirin use, three experiments were completed; each was designed to measure one aspect of nonlinear behavior: (1) the effects of level on frequency selectivity in simultaneous masking using notched-noise maskers, (2) two-tone suppression using forward maskers at the signal frequency (f_s) and suppressor tones above f_s, and (3) growth-of-masking functions in forward masking using a masking tone below f_s. Signal frequencies of 750 and 3000 Hz were used to evaluate the effects of aspirin at relatively low- and high-frequency regions of the cochlea. In experiment 1, aspirin broadened the auditory filters and reduced the effect of level on frequency selectivity. In experiment 2, aspirin reduced or eliminated two-tone suppression. And, in experiment 3, aspirin reduced the slopes of the growth-of-masking functions. Thus, the aspirin was effective in reducing nonlinearity in all three experiments, suggesting that these measures reflect the same (or a similar) active, nonlinear mechanism, namely the compressive nonlinearity provided by the OHCs. In all experiments, aspirin tended to have larger detrimental effects on the nonlinear measures at 3000 Hz than at 750 Hz, which can be explained in terms of greater involvement of nonlinear processing at higher frequencies. Finally, these effects of aspirin were found to be similar to those observed in preliminary measurements in two subjects with mild, permanent hearing loss.
[S0001-4966(99)04509-9]

PACS numbers: 43.66.Ba, 43.66.Dc, 43.66.Sr [JWH]

INTRODUCTION

One way to gain a better understanding of the processing underlying various auditory phenomena is to alter the functioning of the cochlea in some manner to determine the vulnerability of the phenomena being studied. Specifically, by altering the functioning of outer hair cells (OHCs) it would be possible to determine, at least to some extent, their influence on nonlinear phenomena, as they are believed to be responsible for nonlinear behavior in the cochlea. Aspirin (acetylsalicylic acid) is a common ototoxic drug that, in moderate doses, has long been known to induce a temporary, mild-to-moderate, sensorineural hearing loss. Although the exact mechanism is unclear, many anatomical and physiological studies have suggested that the aspirin exerts its effects via the OHCs (Stypulkowski, 1990; Shehata *et al.*, 1991; Russell and Schauz, 1995; Kakehata and Santos-Sacchi, 1996). Evidence for this is also seen in studies of otoacoustic emissions (OAEs), which have been found to be reduced or abolished after the administration of aspirin (McFadden and Plattsmier, 1984; Long and Tubis, 1988; Wier *et al.*, 1988).

There are several advantages to using aspirin to study the nonlinear behavior of the auditory system. One of the main advantages is that the various phenomena can be studied with and without aspirin in the *same* individuals, due to the temporary nature of the hearing loss. Thus subjects can act as their own control, eliminating one source of variability in the research design. However, there is considerable individual variability in the amount of aspirin-induced hearing loss (AIHL) for subjects taking similar doses of aspirin. For moderate aspirin doses (3–5 g/day), the amount of AIHL in humans has been found to range from a few dB to as much as 40 dB (Myers *et al.*, 1965; Day *et al.*, 1989; Bonding, 1979; McFadden and Plattsmier, 1984; McFadden *et al.*, 1984a, 1984b). This variability makes it difficult, if not impossible, to predict how much, if any, hearing loss individuals will develop. Although body size is most likely a factor, the manner and rate at which each individual metabolizes the aspirin may also play a role. A second advantage to using aspirin is that the hearing loss experienced by every individual will have the same physiological origin, namely the OHCs. Furthermore, in many cases, this is also the origin of

[a)]A portion of this research was presented at the Association for Research in Otolaryngology Midwinter Meeting, St. Petersburg, FL, February 1998.
[b)]Electronic mail: michelle.hicks@asu.edu

permanent sensorineural hearing loss, although it is not clear how similar these two types of losses are.

Although the effects of aspirin on hearing sensitivity have been well documented, relatively few studies have examined suprathreshold changes with aspirin usage, and even fewer studies have examined psychophysical phenomena that have been clearly tied to OHC function. One measure that has been intimately linked with OHC function and shown to be affected by aspirin is frequency selectivity. Several investigators have reported reduced frequency selectivity in human listeners after aspirin use (Bonding, 1979; Carlyon and Butt, 1993; Beveridge and Carlyon, 1996), and at least one investigator (Bonding, 1979) has reported a monotonic relationship between the amount of AIHL (20–40 dB) and the increase in auditory filter bandwidth. However, Carlyon and Butt (1993) reported that the largest threshold shifts did not necessarily result in the largest changes in filter bandwidth, and they were able to conclude, as were Beveridge and Carlyon (1996), that even small amounts of AIHL (4–5 dB) were able to reduce frequency selectivity. These results are consistent with the broadened frequency tuning curves observed at the basilar membrane (Murugasu and Russell, 1995) and auditory nerve (Stypulkowski, 1990) after aspirin exposure, and are similar to reports of impaired frequency selectivity in individuals with permanent sensorineural hearing losses (Festen and Plomp, 1983; Glasberg and Moore, 1986; Nelson, 1991; Dubno and Schaeffer, 1992; Leek and Summers, 1993), although these studies suggest that frequency selectivity is impaired primarily for sensorineural hearing losses greater than about 30–40 dB.

A second measure suggested to be linked to OHC function and shown to be affected by aspirin is overshoot, a phenomenon whereby the threshold for a brief tonal signal masked by a long-duration broadband noise is higher near the onset than near the temporal center of the masker. McFadden and Champlin (1990) found that overshoot was reduced after aspirin ingestion, primarily due to a threshold *improvement* in the onset condition while taking aspirin, much like the reduction observed in individuals with sensorineural hearing loss (Bacon and Takahashi, 1992). Furthermore, McFadden and Champlin found that overshoot was not reduced when they simulated a hearing loss with a continuous masking noise, and suggested that overshoot was diminished only when the ''active cochlear mechanism, or cochlear amplifier'' was affected, and not simply as a result of *any* shift in threshold. This finding suggests that the presence of a continuous masking noise, often used to model sensorineural hearing loss in normal-hearing individuals, may not always be an accurate model.

Several temporal aspects of hearing, including temporal resolution and temporal integration, also have been found to be altered following moderate doses of aspirin, although neither phenomenon has been closely tied to OHC function (but see Moore, 1995). McFadden *et al.* (1984a) observed that during aspirin use, recovery from forward masking was somewhat prolonged, although only for those individuals with the most AIHL (14–17 dB), and gap-detection thresholds were poorer, although only at low stimulus levels; performance at higher levels was essentially unchanged. This poorer performance in forward-masking and gap-detection tasks has also been observed in individuals with sensorineural hearing losses (Fitzgibbons and Wightman, 1982; Festen and Plomp, 1983). However, this does not necessarily imply that these individuals or those with AIHL have impaired temporal resolution; the poorer performance in these tasks may be due to the threshold shift (permanent or aspirin induced) reducing the sensation level (SL) of the stimulus. McCabe and Dey (1965) and McFadden *et al.* (1984a) observed reduced temporal integration during aspirin use, although McFadden *et al*. reported that it was reduced only for listeners with the most AIHL (10–21 dB); it was not reduced for listeners with only about 5 dB of AIHL. Reduced temporal integration is also observed in individuals with sensorineural hearing losses (Florentine *et al.*, 1988).

Although several studies have examined the effects of aspirin on measures that generally are thought to be related to nonlinear processing, and hence OHC function (i.e., OAEs, overshoot, frequency selectivity), no one has investigated multiple measures that reflect nonlinear processing in the auditory system in the same subjects. The goal of this project was to examine how three different nonlinear phenomena are affected by a moderate dose of aspirin producing temporary hearing losses in six normal-hearing individuals. Over three consecutive days of aspirin use, three experiments were completed and compared with results when no aspirin was being taken. The three experiments are the same as those examined in normal-hearing listeners by Hicks and Bacon (1999): (1) the effects of level on frequency selectivity in simultaneous masking, (2) two-tone suppression measured using forward masking, and (3) growth-of-masking functions in forward masking. Hicks and Bacon (1999) demonstrated that for these measures, there appears to be greater involvement of nonlinear processing at 3000 Hz than at 750 Hz. Thus, signal frequencies of 750 and 3000 Hz were used to provide a sample of nonlinear behavior at relatively low- and high-frequency regions. If nonlinear processing is more prominent at high frequencies, one might predict that the effect of aspirin would be more noticeable there.

I. GENERAL METHODS

A. Subjects

Six normal-hearing subjects (four female, two male), ranging in age between 22 and 39 years, participated. All subjects completed hearing and medical screenings to determine eligibility to participate; subjects had absolute thresholds at or below 10 dB HL (ANSI, 1989) at the octave frequencies from 250–8000 Hz in the test ear and had no medical contraindications to taking aspirin. Each subject was given several articles to read on the effects of aspirin and was verbally informed of the effects of aspirin on hearing and of the possible side effects they may experience while taking the prescribed dose. Subject S1 (the first author) had previous experience in psychoacoustic tasks similar to those in the present study, and subjects S2–S6 received approximately $\frac{1}{2}$ h of practice with the forward-masking conditions prior to data collection (see Hicks and Bacon, 1999). All subjects participated in the experiments from Hicks and Ba-

con (1999) and the data presented here for the no-aspirin conditions are a subset of the data presented in that study. Subjects S2–S6 were paid for their participation.

B. Drug schedule

Following medical clearance, subjects obtained the aspirin tablets from the university pharmacy with the dosage clearly identified on the bottle. Subjects took three 325-mg tablets of aspirin four times a day (6 am/pm and 12 am/pm) for four days (12 tablets/day for a total dose of 3.9 g/day). Data were collected on the second, third, and fourth days after 8, 12, and 16 doses of aspirin, respectively. After the final dose on the fourth day, a blood sample was drawn and serum salicylate levels were measured to ensure compliance with the dosage schedule. Repeat hearing and medical screenings were completed at the conclusion of the experiments to assure that there were no lasting adverse effects from the aspirin.

C. Apparatus and stimuli

The sinusoidal signals and maskers were digitally generated and produced at a 20-kHz sampling rate. The output of the digital-to-analog converter (TDT DA1) was low-pass filtered at 8 kHz (Kemo VBF 25.01; 135 dB/oct). The notched-noise maskers, which consisted of two bands of noise placed about the signal frequency (f_s), were produced by low-pass filtering (Kemo VBF 25.01) random noise from two independent noise generators (GenRad 1381), and multiplying each low-pass noise by a different sinusoid (HP 8904A synthesizer). The resultant noise masker consisted of two bands of noise, each with a bandwidth that was twice the cutoff frequency of the low-pass filter and a center frequency that was equal to the frequency of the sinusoid. The bandwidth of each noise band was $0.4\ f_s$, and the frequencies of the sinusoids were chosen such that the bands were placed either symmetrically or asymmetrically about f_s. The spectrum level of the noise masker was calibrated using a Hewlett-Packard 3561 spectrum analyzer and controlled via manual attenuators. The level of the signal was varied adaptively via a programmable attenuator (Wilsonics PATT), and the signal and masker were added together prior to being presented monaurally through a TDH-49P headphone mounted in an MX/51 cushion. Specific stimuli and experimental conditions are described prior to each experiment.

D. Experimental procedure

Subjects were tested individually or in pairs in a double-walled, sound-attenuating booth. Thresholds were measured using an adaptive, two-interval forced-choice procedure with a decision rule that estimates 79% correct on the psychometric function (Levitt, 1971). Subjects responded by pushing one of two buttons on a response panel. Lights were used to indicate when the signal might occur and then to provide correct-response feedback. The two observation intervals were separated by 400 ms. The signal level began 10–15 dB above the estimated threshold; it was decreased following three consecutive correct responses and increased following one incorrect response. The step size was initially 4 dB, but was decreased to 2 dB following the second reversal. The first two reversals (or three, if the total number of reversals was odd) in each 50-trial block were discarded, and the signal levels at the remaining reversals were averaged to obtain the threshold for that run. A run was included only if the standard deviation of the reversals was 5 dB or less and if there were at least six reversals in the average. In the no-aspirin conditions, each threshold reported is the mean from at least three runs obtained over at least two days. If the standard deviation of that mean was greater than 3 dB, additional runs were included (up to a maximum of six) until the standard deviation was less than 3 dB. In the aspirin conditions, each threshold reported is the mean from two runs obtained on the same day. If the difference between the two runs was greater than 3 dB, a third run was included in the average. Ninety-eight percent (704/720) of all the means had standard deviations less than 3 dB.

Half of the subjects (S1–S3) completed the no-aspirin conditions first, followed by the aspirin conditions; the other half (S4–S6) completed testing in the reverse order. In order to account for any order or practice effects, as well as any potential differences in the amount of threshold shift across the three days of aspirin ingestion, the subjects completed the aspirin experiments in one of three sequences; two subjects (one from each group) completed the experiments in the same order. The numbers in parentheses in Table I indicate which experiment was completed on each day of aspirin ingestion.

II. ASPIRIN-INDUCED THRESHOLD SHIFTS FOR SHORT AND LONG SIGNALS

Before examining the effects of aspirin on the three psychophysical measures of nonlinearity, first consider the effects of aspirin on hearing sensitivity for each individual. Table I shows quiet thresholds for 15-ms (including 5-ms $\cos^2$ rise/fall times) signals at frequencies of 375, 750, 1500, and 3000 Hz obtained prior to taking aspirin, and the amount of aspirin-induced threshold shift measured at the beginning of each test session (days 2, 3, and 4 of aspirin use). Also shown are the serum salicylate levels obtained on day 4.

In general, the aspirin produced a relatively constant hearing loss across frequency, which for most subjects reached a maximum by the second or third day of aspirin use.[1] The maximum amount of hearing loss (averaged across frequencies) for each subject ranged from approximately 6 (S5 and S6) to 19 dB (S1). The amount of loss is similar to that reported by McFadden *et al.* (1984a, 1984b), Carlyon and Butt (1993), and Beveridge and Carlyon (1996), but is somewhat smaller than that observed by Myers *et al.* (1965) and Bonding (1979) for similar aspirin doses. The difference may be partially due to more precise threshold measurements in the more recent studies; however, it is likely that the considerable range in the amount of AIHL, both within and across studies, also reflects individual susceptibility to aspirin ototoxicity.

All of the subjects reported tinnitus by the second day of aspirin use, and the serum salicylate levels for all subjects were within the therapeutic range, indicating that the subjects were complying with the required drug schedule. Similar to

TABLE I. Baseline quiet thresholds (in dB SPL) for a 15-ms signal (boldface entries) and the amount of aspirin-induced thresholds shift (in dB) obtained on days 2, 3, and 4 of aspirin use. The serum salicylate levels were obtained on day 4. The number in parentheses indicates which experiment was completed on a given day of aspirin ingestion.

Subject		Signal frequency (Hz)					Salicylate level (mg/dl)
		375	750	1500	3000	Mean	
S1	Baseline	**34.2**	**19.8**	**26.4**	**17.4**	**24.5**	
	Day 2 (1)	14.0	14.0	6.3	16.9	12.8	
	Day 3 (2)	16.3	16.3	11.8	22.5	16.7	
	Day 4 (3)	17.2	18.8	14.3	26.0	19.1	36.5
S2	Baseline	**34.8**	**22.4**	**24.2**	**23.6**	**26.3**	
	Day 2 (3)	9.2	8.2	8.0	15.1	10.1	
	Day 3 (1)	16.6	15.0	12.9	20.7	16.3	
	Day 4 (2)	15.2	12.7	9.7	17.2	13.7	22.5
S3	Baseline	**31.6**	**21.4**	**24.8**	**15.9**	**23.4**	
	Day 2 (2)	7.5	6.9	6.6	6.8	6.9	
	Day 3 (3)	7.9	11.4	7.5	9.6	9.1	
	Day 4 (1)	8.9	9.4	6.3	8.0	8.1	19.3
S4	Baseline	**34.4**	**20.4**	**18.1**	**16.7**	**22.4**	
	Day 2 (1)	8.4	3.9	6.9	4.4	5.9	
	Day 3 (2)	6.0	6.0	8.0	6.5	6.6	
	Day 4 (3)	4.5	1.9	3.6	4.8	3.7	15.8
S5	Baseline	**34.2**	**20.1**	**24.5**	**18.2**	**24.2**	
	Day 2 (3)	7.0	7.9	−7.4	5.0	3.1	
	Day 3 (1)	9.0	5.7	−0.3	3.6	4.5	
	Day 4 (2)	9.5	7.7	−1.6	7.4	5.8	24.4
S6	Baseline	**31.2**	**20.6**	**15.3**	**20.4**	**21.9**	
	Day 2 (2)	4.7	4.4	4.7	5.7	4.8	
	Day 3 (3)	5.1	6.2	5.6	6.9	5.9	
	Day 4 (1)	4.3	5.8	4.2	6.2	5.1	13.8
Mean	Baseline	**33.4**	**20.8**	**22.2**	**18.7**	**23.8**	
	Day 2	8.5	7.5	4.2	9.0	7.3	
	Day 3	10.2	10.1	7.6	11.6	9.9	
	Day 4	9.9	9.4	6.1	11.6	9.2	19.2

previous reports (Myers *et al.*, 1965; Bonding, 1979; McFadden and Plattsmier, 1984; McFadden *et al.*, 1984a, 1984b; Day *et al.*, 1989), there was a positive correlation between the amount of AIHL averaged across frequency and serum salicylate levels ($r=0.86$; $p<0.05$).

It was somewhat surprising that the threshold shifts with aspirin were generally as great at low frequencies as they were at high frequencies, given our previous results (Hicks and Bacon, 1999) that suggested that the influence of the compressive nonlinearity was greater at higher frequencies. It may be, however, that the influence of the OHCs at very low levels (i.e., at or near absolute threshold) is more or less independent of frequency, whereas their influence at higher levels is greater at high than at low frequencies.

Quiet thresholds were also obtained for a 300-ms (including 10-ms $\cos^2$ rise/fall times) signal prior to aspirin use and at the beginning of the test session for experiment 1. The difference between the thresholds for the 15- and 300-ms signals can thus be used to obtain estimates of temporal integration prior to and during aspirin use. Table II shows quiet thresholds for the 15- and 300-ms signals at each signal frequency, and the difference between those thresholds (boldface entries) in the aspirin and no-aspirin conditions. In both conditions, the difference between the thresholds for the 15- and 300-ms signals became smaller as f_s increased from 375 to 3000 Hz consistent with shorter time constants of integration at higher frequencies (Plomp and Bouman, 1959). As duration increased from 15 to 300 ms, thresholds decreased by an average of about 15 and 13 dB in the no-aspirin and aspirin conditions, respectively, suggesting somewhat reduced temporal integration with aspirin use. Although the decrease in temporal integration with aspirin was small, it was fairly constant across frequency and was observed for all subjects, regardless of the amount of AIHL. A two-factor, repeated-measures, analysis of variance (ANOVA) on the amount of temporal integration revealed significant effects of both aspirin and signal frequency [$F(1,15)=25.3$; $p<0.01$, and $F(3,15)=10.1$; $p<0.001$, respectively], and no significant interaction [$F(3,15)=0.02$; $p=0.99$]. A Tukey *post hoc* analysis indicated that only nonadjacent frequencies differed from each other ($p<0.05$). A reduced rate of temporal integration with aspirin was also observed by McFadden *et al.* (1984a), although only for a signal frequency of 3500 Hz, not 500 Hz, and only for those subjects with the largest (10–21 dB) AIHL.

The reduced temporal integration with aspirin may pro-

TABLE II. Quiet thresholds (in dB SPL) for the 15- and 300-ms signals. Thresholds were obtained prior to and on the day of aspirin administration when experiment 1 was completed. The differences between the thresholds (in dB) for the 15- and 300-ms signals are shown in boldface.

		No aspirin				Aspirin			
		Signal frequency (Hz)							
Subject	Signal duration (ms)	375	750	1500	3000	375	750	1500	3000
S1	15	34.2	19.8	26.4	17.4	48.2	33.8	32.7	34.3
	300	14.7	6.7	9.1	6.6	32.9	19.9	19.3	26.8
	Difference	**19.5**	**13.1**	**17.3**	**10.8**	**15.3**	**13.9**	**13.4**	**7.5**
S2	15	34.8	22.4	24.2	23.6	51.4	37.4	37.1	44.3
	300	20.9	4.7	11.0	12.7	39.5	24.3	27.2	33.5
	Difference	**13.9**	**17.7**	**13.2**	**10.9**	**11.9**	**13.1**	**9.9**	**10.8**
S3	15	31.6	21.4	24.8	15.9	40.5	30.8	31.1	23.9
	300	14.7	4.3	12.9	5.6	23.9	14.2	22.2	15.0
	Difference	**16.9**	**17.1**	**11.9**	**10.3**	**16.6**	**16.6**	**8.9**	**8.9**
S4	15	34.4	20.4	18.1	16.7	42.8	24.3	25.0	21.1
	300	11.7	3.0	3.7	3.4	22.3	8.9	9.9	6.0
	Difference	**22.7**	**17.4**	**14.4**	**13.3**	**20.5**	**15.4**	**15.1**	**15.1**
S5	15	34.2	20.1	24.5	18.2	43.2	25.8	24.2	21.8
	300	15.4	3.3	5.0	3.9	25.3	13.9	6.2	11.6
	Difference	**18.8**	**16.8**	**19.5**	**14.3**	**17.9**	**11.9**	**18.0**	**10.2**
S6	15	31.2	20.6	15.3	20.4	35.5	26.4	19.5	26.6
	300	14.0	4.6	3.1	8.9	19.6	10.1	8.8	19.8
	Difference	**17.2**	**16.0**	**12.2**	**11.5**	**15.9**	**16.3**	**10.7**	**6.8**
Mean	15	33.4	20.8	22.2	18.7	43.6	29.8	28.3	28.7
	300	15.2	4.4	7.5	6.8	27.3	15.2	15.6	18.8
	Difference	**18.2**	**16.4**	**14.7**	**11.9**	**16.3**	**14.6**	**12.7**	**9.9**

vide support for recent arguments linking temporal integration with the peripheral compressive nonlinearity (Moore, 1991, 1995; Oxenham *et al.*, 1997). The small (about 2 dB) effect of aspirin on temporal integration could reflect the fact that the amount of compression is generally thought to be negligible at low levels. The frequency-independent effect of aspirin might, as proposed above, suggest that the influence of the OHCs at low levels is independent of frequency. A possible argument against the influence of the compressive nonlinearity on temporal integration in the present study, however, is that Oxenham *et al.* (1997) argued that the influence of compression was limited to durations shorter than those used here.

III. EXPERIMENT 1: EFFECT OF LEVEL ON FREQUENCY SELECTIVITY

A. Rationale

The purpose of this experiment was to determine how AIHL alters the effect of level on frequency selectivity. It has previously been shown that filter shapes derived from notched-noise forward-masking data (Carlyon and Butt, 1993) and psychophysical tuning curves measured using forward and simultaneous masking (Beveridge and Carlyon, 1996) become more broadly tuned after aspirin administration. However, these studies only evaluated frequency selectivity at 4000 Hz, and did not examine the effects of level on frequency tuning.

B. Stimuli and conditions

Simultaneous-masked thresholds were obtained for signal frequencies of 750 and 3000 Hz at two masker spectrum levels (30 and 50 dB). The noise masker consisted of two bands of noise, each $0.4f_s$ wide, placed symmetrically or asymmetrically around f_s. In the symmetric conditions, the edge of the noise band closer to f_s was placed at 0.0, 0.2, and $0.4f_s$. In the asymmetric conditions, the closer edge of the noise band on the low- (or high-) frequency side was placed at $0.2f_s$, and the closer edge of the corresponding noise band on the high- (or low-) frequency side was placed at $0.4f_s$, for a total of three symmetric and two asymmetric notch conditions.[2] The signal and masker were both 300 ms in duration (including 10-ms $\cos^2$ rise/fall times).

The data from each combination of signal frequency and masker level were fitted with the roex(p,r) model (Patterson *et al.*, 1982) using a least-squares criterion to derive auditory filter shapes. A modified version of the computer program of Glasberg and Moore (1990) was used to obtain measures of the dynamic range of the filter (r), the slopes of the lower (p_l) and upper (p_u) filter skirts, and the equivalent rectangular bandwidth (ERB). As suggested by Glasberg and Moore, correction for variations in absolute thresholds across frequency was applied to the fitting procedure by implementing an equal-loudness-contour weighting factor for the higher masker level. Additional details of the fitting procedure can be obtained from Hicks and Bacon (1999).

C. Results and discussion

The masked thresholds in the no-aspirin and aspirin conditions exhibited similar trends across subjects (despite different amounts of AIHL for the 300-ms signal, as shown in Table II), and therefore only the averaged data are presented in Fig. 1. Results are shown for each signal frequency as a function of the deviation of the nearer noise band from f_s, normalized to $f_s(\Delta f/f_s)$. Up-pointing triangles represent conditions where the noise bands were placed symmetrically

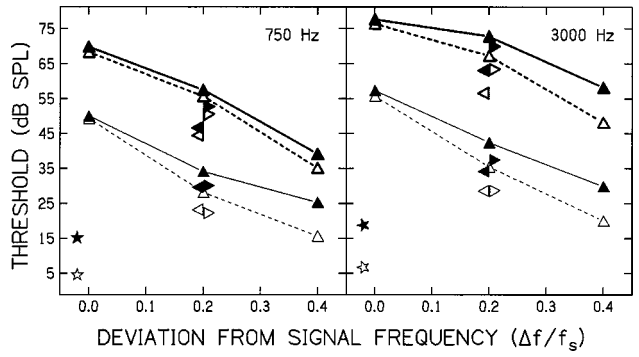


FIG. 1. Masked thresholds in the aspirin (filled) and no-aspirin (unfilled) conditions are shown as a function of the deviation of the nearer noise band from the signal frequency, normalized to the signal frequency. Up-pointing triangles represent conditions where the noise bands were placed symmetrically about the signal frequency, and right- and left-pointing triangles represent conditions where the high- or low-frequency band was further away, respectively. The bold lines are for the 50-dB masker, and the thin lines are for the 30-dB masker. Stars indicate quiet thresholds. The results were averaged across subjects S1–S6 (the standard errors of the means across subjects ranged from 0.35–1.42 at 750 Hz, and from 0.97–3.73 at 3000 Hz).

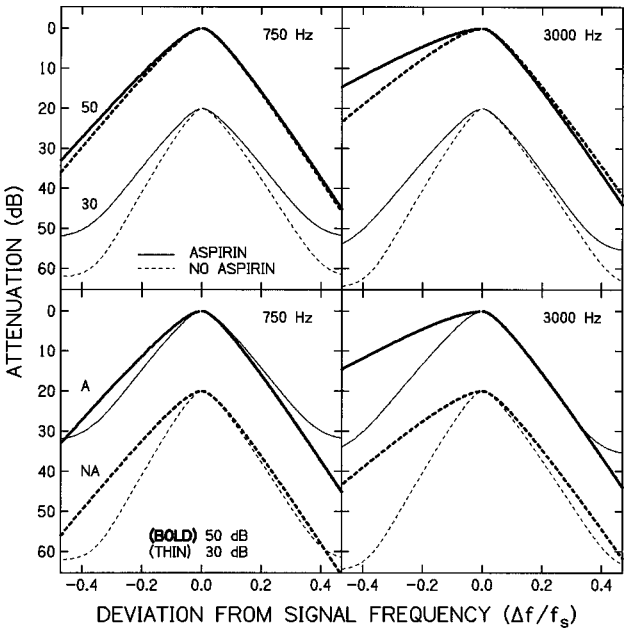


FIG. 2. Auditory filter shapes derived from the mean data of subjects S1–S6. The solid lines show filters obtained in the aspirin conditions (A), and dashed lines show those obtained in the no-aspirin (NA) conditions. The bold and thin lines indicate filters obtained with masker spectrum levels of 50 and 30 dB, respectively. In the top panels, the filters obtained at a masker level of 30 dB were displaced downward by 20 dB. The identical filters are replotted in the bottom panels, but with the filters obtained in the no-aspirin conditions displaced downward by 20 dB.

about f_s, and right- and left-pointing triangles represent conditions where the high- or low-frequency band was further away, respectively. The stars indicate absolute thresholds for the 300-ms signal.

In most cases, thresholds for the aspirin conditions (filled symbols) were higher than those for the no-aspirin conditions (unfilled symbols), although the extent of the difference was dependent on the frequency of the signal and the spectrum level of the masker (30 dB, thin lines; 50 dB, bold lines). For both signal frequencies and masker levels, there was only a slight difference between thresholds for the aspirin and no-aspirin conditions when there was no frequency separation between the noise bands and the signal frequency. Thus, if the width of the auditory filters were estimated using the critical ratio, there would be only a small difference in the estimated bandwidth between the aspirin and no-aspirin conditions. However, for the 3000-Hz signal at both masker levels, and for the 750-Hz signal at the 30-dB masker level, as the noise bands were moved away from f_s, thresholds for the aspirin conditions tended to decrease more gradually than those for the no-aspirin conditions, suggesting broader auditory filters in the aspirin conditions. For the 750-Hz signal at the 50-dB masker level, however, there was relatively little difference between the thresholds for the aspirin and no-aspirin conditions, suggesting that aspirin had little effect on the shape of the filter in this condition. These effects, as well as any asymmetries in the filters, can be seen more easily by examining the derived filter shapes, shown in Fig. 2.

In the top panels of Fig. 2, the filters obtained at a masker spectrum level of 30 dB (thin lines) are displaced downward by 20 dB from those obtained at 50 dB (bold lines) for easier comparison between the aspirin (solid lines) and no-aspirin (dashed lines) conditions. Several measures related to the derived filters (ERB, p_l, p_u, and r) are given in Table III for the individual and mean data. (Note that the mean values represent the parameters of the filters fitted to the mean data, and not the mean of the parameter values across subjects.)

First consider the effects of aspirin at each masker level (top row, Fig. 2). At 750 Hz (left panel), the filters measured at a masker spectrum level of 30 dB were an average of about 25% broader in the aspirin conditions than in the no-aspirin conditions (range=13% to 40%). Both the low- and high-frequency slopes became shallower with aspirin (smaller p_l and p_u values); for some subjects the effect was larger on the low-frequency side (S1, S2, and S4), and for others it was larger on the high-frequency side (S3 and S6). The correlation between the amount of AIHL and the change in filter bandwidth was not significant ($r=0.28$; $p=0.59$); the subject with the largest AIHL (S2, 19 dB) and the subject with the smallest AIHL (S4, 5 dB) had approximately the same change in ERB value after aspirin ingestion (about 21%). In contrast to the results at 30 dB, the filters at 50 dB were more similar to one another in the aspirin and no-aspirin conditions. The increase in filter bandwidth after aspirin ingestion was only about 5% (range=−3% to 10%). A two-factor, repeated-measures ANOVA with aspirin and masker level as factors confirmed these observations. The increases in filter bandwidths with aspirin and with masker level were both significant [$F(1,5)=51.1$; $p<0.001$, and $F(1,5)=19.4$; $p<0.01$, respectively], as was the interaction between the factors [$F(1,5)=20.5$; $p<0.01$]. Tukey *post hoc* tests for significance revealed that the increase in filter bandwidth with aspirin was significant at a masker level of 30 dB ($p<0.001$), but not at 50 dB ($p>0.05$).

For the 3000-Hz signal (top-right panel), the filters measured at both masker levels were an average of about 23% broader in the aspirin conditions than in the no-aspirin conditions (range=−3% to 78%). The largest effect of aspirin

TABLE III. ERB, p_l, p_u, and r values for the filters derived from the individual and mean data obtained in the no-aspirin and aspirin conditions. Asterisks (*) indicate values that may have been overestimated by the fitting procedure.[3]

Subject	Signal frequency (Hz)	Masker level (dB)	No aspirin				Aspirin			
			ERB	p_l	p_u	r	ERB	p_l	p_u	r
S1	750	30	91	41.6	27.1	−37.6	127	24.8	23.7	−27.1
		50	141	19.9	22.8	−71.9	136	18.9	26.4	−75.6
	3000	30	453	28.6	24.6	−47.4	600	21.9	18.7	−29.6
		50	693	13.0	25.8	−61.6	990	6.9	48.2*	−69.4
S2	750	30	90	40.5	28.2	−45.1	109	29.9	26.0	−32.3
		50	113	24.7	28.9	−46.1	112	25.1	29.0	−41.8
	3000	30	426	25.3	31.7	−39.4	762	17.3	16.8	−19.5
		50	531	17.2	32.9	−69.2	663	12.1	36.2*	−57.6
S3	750	30	91	30.8	35.0	−44.4	121	26.9	23.6	−30.5
		50	123	21.7	27.8	−87.7	136	19.8	24.8	−44.1
	3000	30	321	45.6	31.8	−48.8	426	22.9	36.7	−46.2
		50	441	25.0	29.7	−80.6	546	17.8	24.8	−72.6
S4	750	30	94	35.4	29.2	−45.8	115	24.8	27.2	−37.9
		50	114	24.4	28.7	−80.5	115	23.3	29.1	−76.3
	3000	30	408	27.7	31.2	−53.3	393	26.4	36.5	−49.6
		50	831	12.9	16.3	−65.2	777	10.5	28.7*	−56.8
S5	750	30	90	35.4	31.8	−41.6	115	25.3	27.2	−36.1
		50	106	25.1	32.7	−83.5	109	23.6	32.5	−79.5
	3000	30	324	46.7	30.8	−46.8	345	36.7	32.9	−41.51
		50	597	16.1	26.5	−60.8	678	13.3	27.2	−29.6
S6	750	30	97	28.2	34.4	−39.3	110	27.2	27.7	−31.0
		50	116	23.8	28.1	−46.6	123	22.3	26.8	−39.0
	3000	30	354	31.5	36.7	−39.1	471	23.8	27.8	−34.3
		50	717	12.5	25.4	−36.5	906	6.4	22.0*	−55.5
Mean	**750**	**30**	**94**	**34.3**	**30.1**	**−42.1**	**118**	**26.0**	**25.4**	**−32.3**
		50	**119**	**22.8**	**28.1**	**−92.6**	**125**	**21.1**	**27.7**	**−68.5**
	3000	**30**	**384**	**33.5**	**29.5**	**−44.6**	**468**	**24.0**	**27.7**	**−35.8**
		50	**609**	**15.9**	**25.9**	**−80.1**	**765**	**11.0**	**27.1***	**−58.9**

on the average filter shapes was a decrease in the low-frequency slope. However, several subjects also demonstrated decreased high-frequency slopes, whereas still others demonstrated increased high-frequency slopes. This individual variability makes it difficult to determine whether there are differential effects on the low- and high-frequency slopes with aspirin. However, it is clear that the aspirin caused a broadening of the filters at both masker levels, and, unlike the results at 750 Hz, it appears that the increase in bandwidth is positively correlated with the amount of AIHL (30 dB: $r=0.83$; $p<0.05$, 50 dB: $r=0.81$; $p=0.05$). A two-factor, repeated-measures ANOVA with aspirin and masker level as factors revealed significant effects of both aspirin and masker level [$F(1,5)=8.3$; $p<0.05$, and $F(1,5)=15.2$; $p=0.01$, respectively], and no significant interaction [$F(1,5)=0.02$; $p=0.90$].

The broader filters in the aspirin conditions at 750 Hz at the 30-dB masker level and at 3000 Hz at both masker levels are consistent with previous results (Bonding, 1979; Carlyon and Butt, 1993; Beveridge and Carlyon, 1996). The magnitude of the effect of aspirin on frequency selectivity is somewhat larger here than in the previous studies, although there is considerable individual variability in the present and previous studies regarding the extent of the effects of aspirin. The results are also consistent with the previous studies in that even small amounts of AIHL (4–5 dB) can result in broadened filters (see for example, S4 at 750 Hz and S6 at both frequencies). The results at 3000 Hz suggest there is a direct relationship between the amount of AIHL and the reduction in frequency selectivity, although this was not the case at 750 Hz, nor in the results of Carlyon and Butt (1993) at 4000 Hz.

Now consider how aspirin influenced the effects of level on frequency selectivity. In the bottom panels of Fig. 2, the same filters as in the top panels are replotted with the filters obtained in the no-aspirin (NA) conditions displaced downward by 20 dB from those obtained in the aspirin (A) conditions for easier comparison of the effects of level. In the no-aspirin conditions, the average filters were broader at the 50-dB masker level than at the 30-dB masker level, and the difference between the two levels was greater at 3000 Hz than at 750 Hz. This larger effect of level at 3000 Hz is most likely reflective of a greater contribution of nonlinear auditory processing at higher frequencies (Rosen and Stock, 1992; Hicks and Bacon, 1999). In the aspirin conditions, the average filter bandwidths were also wider at the 50-dB masker level than at the 30-dB masker level, although this effect of level was reduced at 750 Hz relative to the no-

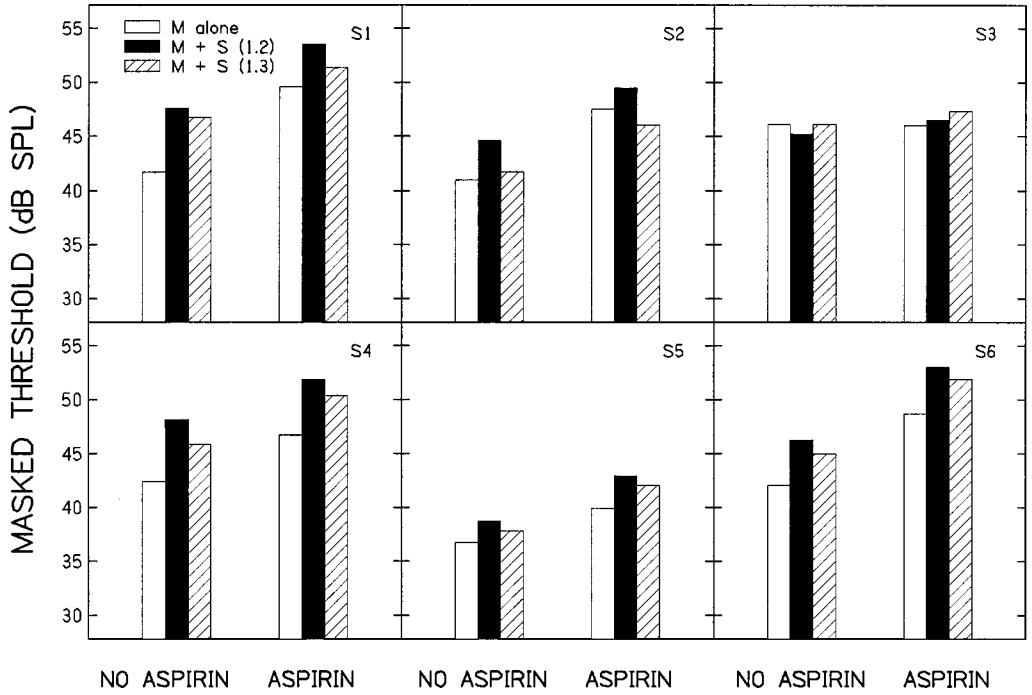


FIG. 3. Masked thresholds for the 750-Hz signal for three masking conditions: the masker alone (M, unfilled bars), and the masker plus suppressor (M+S) at suppressor/masker frequency ratios of 1.2 (filled bars) and 1.3 (hatched bars). Results obtained in no-aspirin and aspirin conditions are shown in the left and right histograms, respectively. Results for the six subjects are shown across panels.

aspirin condition. At 750 Hz, the average bandwidth increased with level by 27% without aspirin, but by only 6% with aspirin. At 3000 Hz, there was no difference between the aspirin and no-aspirin conditions with respect to the effect of level on filter bandwidth. In both cases, the average bandwidth increased by about 60% as the masker spectrum level increased from 30 to 50 dB. Because nonlinear processing appears to be greater at high than at low frequencies (Rosen and Stock, 1992; Hicks and Bacon, 1999), we expected the influence of aspirin on the effect of level to be greater at 3000 Hz. Instead, aspirin had little influence on the effect of level on frequency selectivity at 3000 Hz, although aspirin was clearly influencing the nonlinearity, as indicated by the broadened filters. The large influence of aspirin at both levels at 3000 Hz may reflect the generally greater degree of nonlinear processing at higher frequencies. That is, the nonlinearity may be stronger over a wider range of levels at 3000 Hz than at 750 Hz.

IV. EXPERIMENT 2: TWO-TONE SUPPRESSION

A. Rationale

Two-tone suppression has been measured physiologically at the basilar membrane and in auditory nerve fibers (e.g., Sachs and Kiang, 1968; Schmiedt, 1982; Ruggero *et al.*, 1992; Rhode and Cooper, 1993, and numerous others), and is generally thought to arise from nonlinear cochlear mechanics. Schmiedt *et al.* (1980) have shown that two-tone suppression measured from fibers with a high characteristic frequency (CF) in the gerbil auditory nerve is vulnerable to basal OHC loss after exposure to kanamycin, whereas two-tone suppression measured from mid- or low-CF fibers is resistant to mid-cochlear OHC loss after exposure to intense sound. In humans, Smoorenburg (1980) demonstrated in one individual that psychophysical two-tone suppression measured at a mid-frequency region (2000–2500 Hz) was unaffected by a temporary threshold shift in that region following exposure to intense sound. From these studies, it is not clear how altering OHC function with aspirin will affect measurement of two-tone suppression. Therefore, the purpose of this experiment was to determine how AIHL alters psychophysical measures of two-tone suppression.

B. Stimuli and conditions

The same signal frequencies that were used in experiment 1 were used here. Forward-masked thresholds were obtained in the presence of a single masker tone at f_s, or that masker tone plus a suppressor tone at a frequency of 1.2 or $1.3f_s$. The levels of the masker and suppressor were 50 and 70 dB SPL, respectively. The masker and suppressor (when it was present) were 200 ms in duration and immediately preceded the 15-ms signal (i.e., there was a 0-ms delay at the zero-voltage points). All stimulus durations included 5-ms $\cos^2$ rise/fall times.

C. Results and discussion

Figures 3 and 4 show individual results for signal frequencies of 750 and 3000 Hz, respectively. In each figure, masked thresholds are shown for the no-aspirin (left histograms) and aspirin (right histograms) conditions. The unfilled bars show thresholds when the on-frequency masker was presented alone (M), and the filled and hatched bars

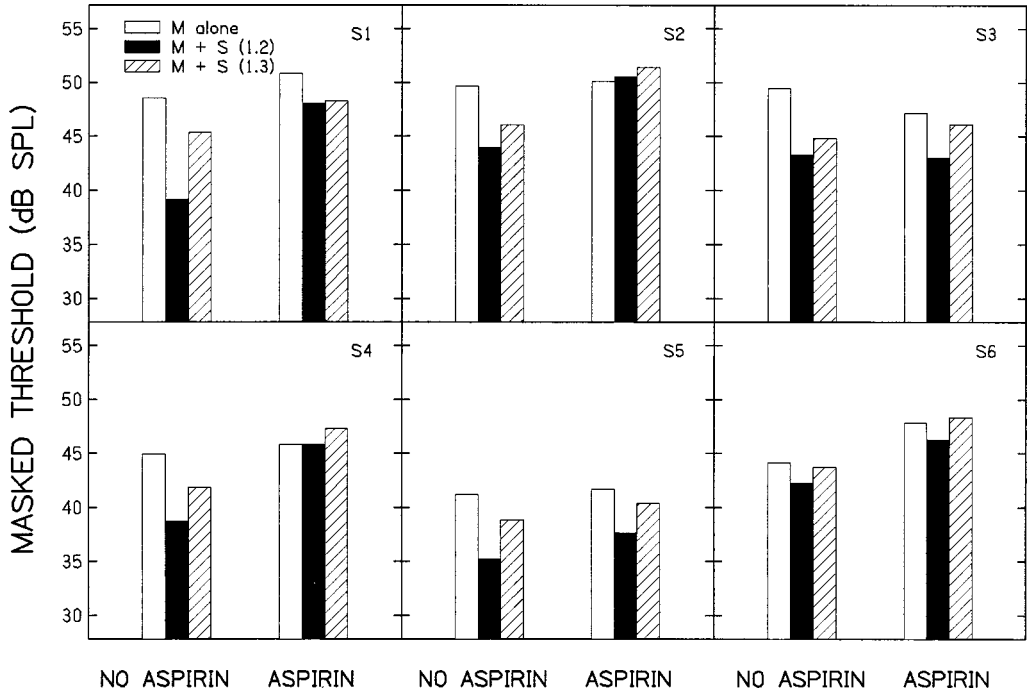


FIG. 4. As Fig. 3, but for the 3000-Hz signal.

show thresholds when the suppressor was present (M+S) at suppressor/masker frequency ratios of 1.2 and 1.3, respectively. Quiet thresholds and the amount of AIHL for each subject can be seen in Table I.

When the signal frequency was 750 Hz, thresholds for the M+S conditions were generally about 4–5 dB higher than those for the M condition, indicating that the suppressor was producing masking. The increase was approximately the same for the no-aspirin and aspirin conditions. However, when the signal frequency was 3000 Hz, the suppressor generally decreased thresholds relative to those for the M condition, indicating a certain amount of suppression.[4] For all subjects, the decrease was larger when the suppressor/masker frequency ratio was 1.2 than when it was 1.3. The decrease was also larger in the no-aspirin conditions than in the aspirin conditions, suggesting smaller amounts of suppression

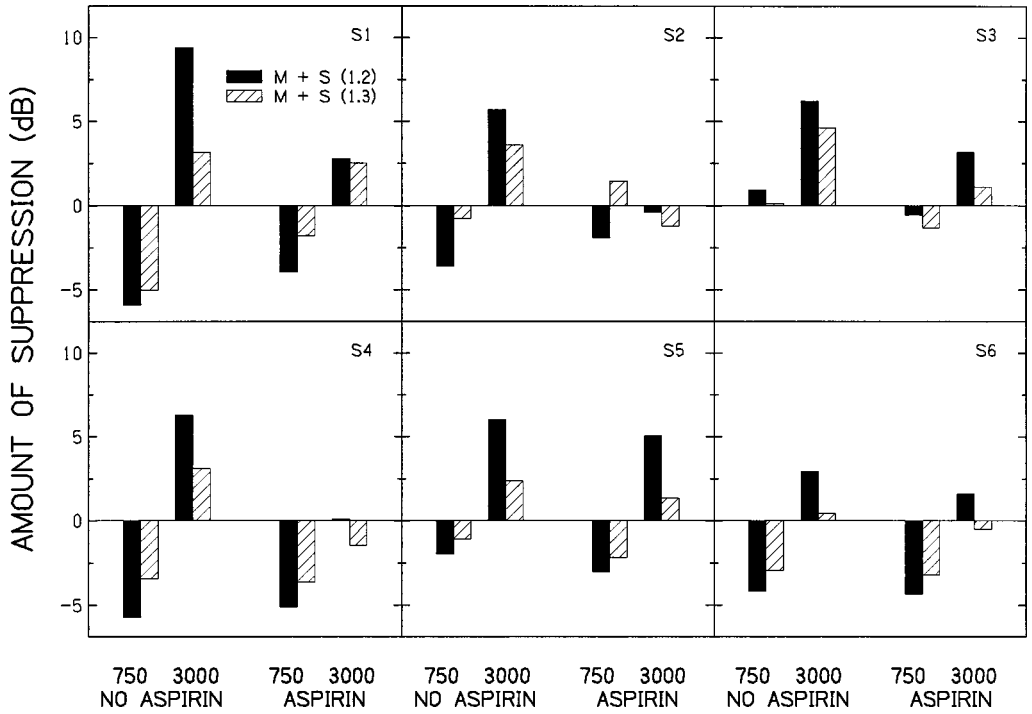


FIG. 5. Amount of suppression for both signal frequencies in the no-aspirin (left histograms) and aspirin (right histograms) conditions. Suppression was calculated by subtracting the thresholds in the M+S conditions from the thresholds in the M condition. The suppressor/masker frequency ratio was either 1.2 (filled bars) or 1.3 (hatched bars). Results for the six subjects are shown across panels.

after aspirin ingestion. For some subjects (S2 and S4), aspirin completely eliminated suppression.

The results in Figs. 3 and 4 have been replotted in Fig. 5 in terms of amount of suppression, calculated by subtracting the thresholds in the M+S conditions from the thresholds in the M condition. In general, when the signal frequency was 750 Hz, there was about −3 and −2 dB of suppression (i.e., an increase in masking) for suppressor/masker frequency ratios of 1.2 and 1.3, respectively, for both the aspirin and no-aspirin conditions, indicating no effect of aspirin. A two-factor, repeated-measures ANOVA with aspirin and frequency ratio as factors confirmed that there were no significant effects of aspirin or ratio [$F(1,5)=0.3$; $p=0.64$, and $F(1,5)=5.4$; $p=0.07$, respectively]. When the signal frequency was 3000 Hz, all subjects showed reduced suppression in the aspirin conditions relative to the no-aspirin conditions for both suppressor/masker frequency ratios. The average amount of suppression was reduced from about 6 to 2 dB for the 1.2 ratio, and from about 3 to less than 1 dB for the 1.3 ratio. An ANOVA revealed that the decrease in suppression with aspirin was significant [$F(1,5)=17.2$; $p<0.01$], as was the difference in the amount of suppression at the two ratios [$F(1,5)=52.2$; $p<0.001$]. There was no significant interaction. The correlations between the amount of AIHL and the change in the amount of suppression after aspirin ingestion were not significant (1.2 ratio: $r=0.62$, $p=0.19$; 1.3 ratio: $r=0.11$, $p=0.84$). Only two subjects, S2 and S4, showed an absence of suppression after aspirin ingestion, whereas the other four subjects continued to show a few dB of suppression, at least for the 1.2 suppressor/masker frequency ratio.

The amount of suppression could also be determined as a change in effective masker level, by dividing the change in threshold (from the M to the M+S condition) by the slope of the growth-of-masking function for the M condition (for details, see Moore and Glasberg, 1983; Lee and Bacon, 1998). Thus, we measured growth-of-masking functions in forward masking for the on-frequency masker (M) over a range of masker levels from 30–70 dB SPL; those slopes are shown in the top of Table IV. However, in the present experiment, the conclusion regarding the effect of aspirin on suppression would be the same for both measures of suppression as the slopes of the growth-of-masking functions obtained with and without aspirin were very similar to one another, and did not reveal any differences across the two frequencies.

In order to determine whether reduced SL as a result of the AIHL could account for the reduction in suppression, the amount of two-tone suppression, obtained with and without aspirin, was compared at equivalent SLs. (After completion of the aspirin experiments, when thresholds had returned to preaspirin values, suppression was remeasured using masker and suppressor levels that were equivalent in SL to those used in the aspirin experiments.) The average amount of suppression (calculated as a change in threshold) at 750 Hz was small and negative for both the aspirin and no-aspirin conditions, and at 3000 Hz, there continued to be less suppression in the aspirin conditions than in the no-aspirin conditions. The differences in the amount of suppression in the aspirin and no-aspirin conditions at *equal* SLs was similar to that at *unequal* SLs (but equal SPLs), indicating that reduced SL in the aspirin condition cannot account for the reduction in suppression. Thus the results of this experiment suggest that aspirin may have less effect at low frequencies, where the nonlinearity may be weak (in fact, so weak as to not show suppression), but have a relatively large effect on suppression at 3000 Hz, where nonlinear processing may have greater influence (see Sec. VI).

TABLE IV. Individual and mean slopes of the growth-of-forward masking functions in the no-aspirin and aspirin conditions for each signal frequency. The top section shows the slopes of the masking functions when the masker frequency=f_s. The middle section shows the slopes when the masker frequency=f_s/1.44. (Slope values were determined from least-squares fits to the data over the range of masker levels where at least 10 dB of masking occurred.) The bottom section shows the transformed slopes determined from the ratio of the slopes for a masker=f_s/1.44 to the slopes for a masker=f_s.

	No aspirin		Aspirin	
Subject	750 Hz	3000 Hz	750 Hz	3000 Hz
	Masker frequency=f_s			
S1	0.74	0.68	0.64	0.61
S2	0.50	0.67	0.41	0.59
S3	0.59	0.73	0.57	0.58
S4	0.67	0.60	0.75	0.84
S5	0.57	0.59	0.60	0.57
S6	0.64	0.60	0.66	0.70
Mean	**0.62**	**0.64**	**0.61**	**0.65**
	Masker frequency=f_s/1.44			
S1	1.10	2.08	0.89	1.09
S2	0.98	1.50	0.74	1.11
S3	0.75	1.97	0.69	1.59
S4	0.92	2.03	0.91	1.57
S5	0.86	1.27	0.84	1.07
S6	0.79	1.38	0.78	1.12
Mean	**0.90**	**1.70**	**0.81**	**1.26**
	Transformed slopes			
S1	1.48	3.06	1.40	1.78
S2	1.97	2.24	1.79	1.88
S3	1.27	2.69	1.23	2.64
S4	1.37	3.36	1.21	1.85
S5	1.51	2.17	1.39	1.87
S6	1.25	2.30	1.18	1.59
Mean	**1.46**	**2.64**	**1.34**	**1.93**

V. EXPERIMENT 3: GROWTH OF FORWARD MASKING

A. Rationale

The purpose of this experiment was to measure growth-of-masking functions in forward masking before and during aspirin ingestion to determine how aspirin alters the slopes of the masking functions. It has been suggested that when the masker frequency is below f_s, the slope of the growth-of-masking function can be used to estimate basilar membrane nonlinearity (Oxenham and Plack, 1997). This idea is based on physiological studies that have shown that the basilar membrane response to a tone at CF (the signal) is compressive at moderate levels, whereas the response to a tone below CF (the masker) is linear. Thus, if subjects monitor the place along the basilar membrane corresponding to f_s, a given

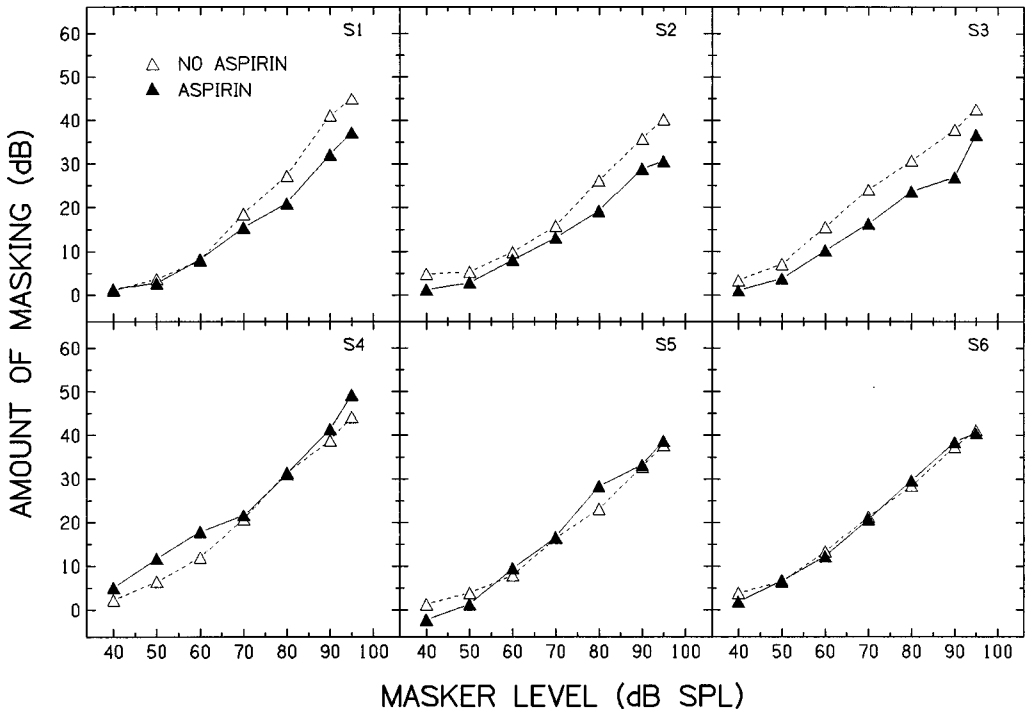


FIG. 6. Growth-of-masking functions in forward masking for the aspirin (filled) and no-aspirin (unfilled) conditions. The signal frequency was 750 Hz, and the masker frequency $= f_s/1.44$. Results for the six subjects are shown across panels.

increase in masker level will require a larger increase in signal level in order to maintain a fixed signal-to-masker ratio in terms of excitation. The resulting growth-of-masking function (signal level versus masker level) will be steeper than if the basilar membrane were responding linearly to both tones. Thus, a steep masking function is consistent with a compressive basilar membrane input/output function. Several studies have demonstrated that the slopes of the growth-of-masking functions measured from hearing-impaired individuals are more gradual than those measured from normal-hearing individuals (simultaneous masking: Stelmachowicz *et al.*, 1987; Murnane and Turner, 1991; Nelson and Schroder, 1997; forward masking: Oxenham and Plack, 1997), suggesting that permanent sensorineural hearing loss reduces the compressive nonlinearity. However, no studies have examined the effects of AIHL to determine whether altering OHC function via aspirin will affect the slope of the masking function. Moreover, aspirin will provide a better test of the role of OHCs (and hence the role of compressive nonlinearity) on the slopes of these masking functions, as the underlying etiology of permanent sensorineural hearing loss is often unknown.

B. Stimuli and conditions

The same signal frequencies that were used in experiments 1 and 2 were used here. Growth-of-masking functions were measured with a pure-tone forward masker below f_s at a signal/masker frequency ratio of 1.44. Masked thresholds were measured over a range of masker levels from 40 to 95 dB SPL. The masker was 200 ms in duration and immediately preceded the 15-ms signal (i.e., there was a 0-ms delay at the zero-voltage points). All stimulus durations included 5-ms $\cos^2$ rise/fall times.

C. Results and discussion

Figures 6 and 7 show the growth-of-masking functions (amount of masking as a function of masker level) for signal frequencies of 750 and 3000 Hz, respectively. Masking functions for each subject obtained with (filled symbols) and without (unfilled symbols) aspirin are shown across panels. Table IV (mid) shows the slopes of the lines fitted to the individual and mean data, determined from least-squares fits to the data over the range of masker levels where at least 10 dB of masking occurred. At 750 Hz, this range was typically 60 to 95 dB, and at 3000 Hz, the range was 70 to 95 dB.

At a signal frequency of 750 Hz, the slope of the masking function in the aspirin condition was only slightly less than that in the no-aspirin condition; the average slopes in the aspirin and no-aspirin conditions were 0.81 and 0.90 dB/dB, respectively. At a signal frequency of 3000 Hz, the slope of the masking function in the aspirin condition was considerably less than that in the no-aspirin condition; the average slopes in the aspirin and no-aspirin conditions were 1.26 and 1.71 dB/dB, respectively. A two-factor, repeated-measures ANOVA with aspirin and signal frequency as factors revealed that the decrease in slope with aspirin was significant $[F(1,5)=13.7;\ p=0.01]$, as was the increase in slope with frequency $[F(1,5)=28.9;\ p<0.01]$. There was also a significant interaction between the factors $[F(1,5)=13.8;\ p=0.01]$; Tukey *post hoc* tests revealed that the slight decrease in slope with aspirin at 750 Hz was not significant

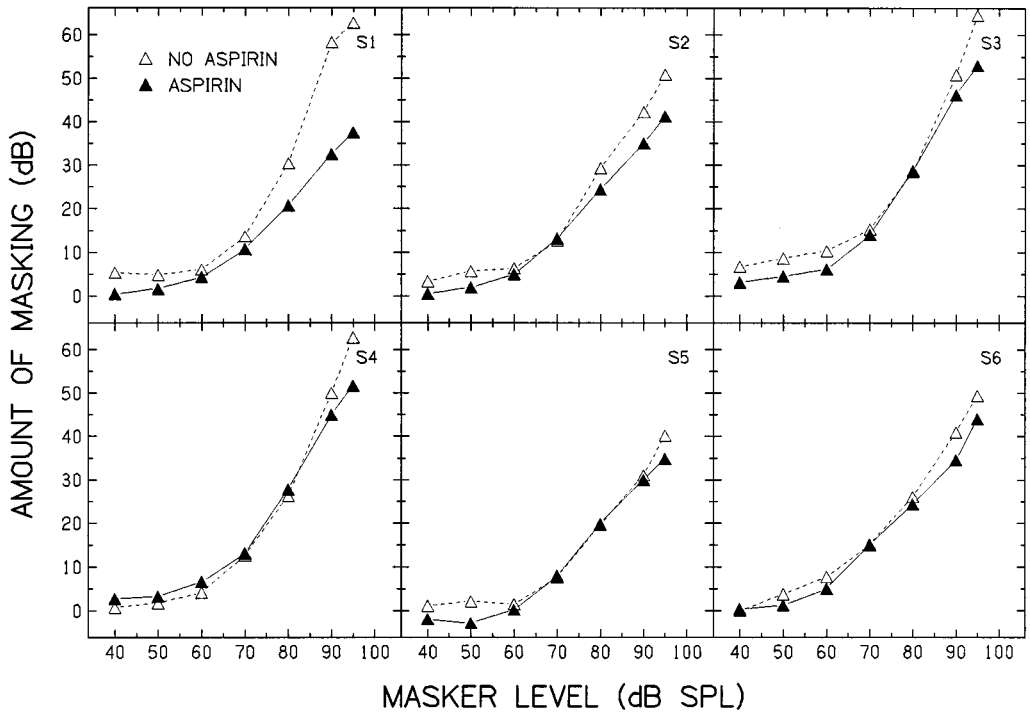


FIG. 7. As Fig. 6, but for the 3000-Hz signal.

($p>0.05$), whereas the decrease in slope at 3000 Hz was highly significant ($p<0.001$). Moreover, there was a tendency for those individuals with the greatest AIHL to exhibit the largest decreases in slope with aspirin [750 Hz: ($r=0.91$, $p<0.05$), 3000 Hz: ($r=0.81$, $p<0.05$)].

As suggested in previous studies, the amount of basilar membrane compression at f_s can be derived by taking the ratio of the slopes of the masking functions for a masker that is equal to f_s and a masker that is lower than f_s (Stelmachowicz *et al.*, 1987; Oxenham and Plack, 1997). This can be done here by taking the ratio of the slopes of the masking functions from this experiment (where masker frequency$=f_s/1.44$) and the masking functions from experiment 2 (where the masker frequency$=f_s$). These transformed slopes are shown in the bottom of Table IV. There was little difference in the slopes of the masking functions (when the masker frequency$=f_s$) between the two signal frequencies or between the aspirin and no-aspirin conditions. As a result, the transformed slopes show the same trends as the slopes of the lines fitted to the data. However, the transformed slopes are steeper because the growth-of-masking functions when the masker frequency was equal to f_s were less than 1.0 (Table IV, top). At 750 Hz, the transformed slopes from the mean data were 1.34 and 1.46 dB/dB in the aspirin and no-aspirin conditions, respectively. At 3000 Hz, the transformed slopes were 1.93 and 2.64 dB/dB in the aspirin and no-aspirin conditions.

To determine whether reduced SL as a result of the AIHL could account for the reduction in slopes, the slopes of the masking functions were recalculated and compared over a fixed range of amount of masking, rather than over a fixed range of masker levels. (Recalculating the functions at equivalent masker SLs was not necessary, given that this would only shift the masking functions in the aspirin conditions horizontally and would not affect the slope values.) At 750 Hz, slopes of 0.83 and 0.89 dB/dB were measured in the aspirin and no-aspirin conditions, respectively, and at 3000 Hz, slopes of 1.20 and 1.64 dB/dB were measured. These values are very similar to those calculated over a fixed range of masker levels, indicating that the reduction in slope caused by the aspirin cannot be accounted for by reduced SL.

VI. SUMMARY AND CONCLUSIONS

The goal of this project was to gain additional insight into the influence of OHCs on three psychophysical measures of nonlinearity by examining the effects of aspirin on these measures. In general, the aspirin was effective in reducing the nonlinearity in all three measures, suggesting that these measures reflect the same (or a similar) active, nonlinear mechanism, namely the compressive nonlinearity provided by the OHCs. Aspirin tended to affect the high frequencies more consistently, although this may reflect the generally greater nonlinear behavior at high frequencies (Rosen and Stock, 1992; Lee and Bacon, 1998; Hicks and Bacon, 1999). At 3000 Hz, the aspirin produced significantly broader auditory filters in experiment 1 (although it did not change the nonlinear effect of level), reduced or eliminated two-tone suppression in experiment 2, and considerably decreased the steep slope of the growth-of-masking functions

in experiment 3. At 750 Hz, where the nonlinear processing appears weak (as evidenced by a lack of suppression in experiment 2 and nearly linear growth of masking in experiment 3), the only significant effect of aspirin was a broadening of the auditory filters at the lower masker level in experiment 1, thus eliminating the small effect of level on frequency selectivity at that frequency. In addition to the effect on the three suprathreshold tasks, aspirin reduced the amount of temporal integration by about 2 dB. The effect was independent of signal frequency, which may indicate that the influence of OHCs at low levels—as opposed to at high levels—is more or less independent of frequency.

It has previously been suggested that temporary AIHL may be useful as a model for permanent sensorineural hearing loss (McFadden *et al.*, 1984a). It is worth examining to what extent the effects of aspirin observed here are similar to the effects of permanent loss observed previously. Previous research has shown that permanent sensorineural hearing loss results in a decrease in frequency selectivity (Festen and Plomp, 1983; Glasberg and Moore, 1986; Nelson, 1991; Dubno and Schaeffer, 1992; Leek and Summers, 1993), a reduction in suppression (Wightman *et al.*, 1977; Leshowitz and Lindstrom, 1977), and a decrease in the slope of growth-of-masking functions (Stelmachowicz *et al.*, 1987; Murnane and Turner, 1991; Oxenham and Plack, 1997; Nelson and Schroder, 1997). We observed similar effects in our subjects with AIHL. Interestingly, we observed these effects in subjects with losses as small as 5 dB. Most of the subjects with permanent hearing loss have considerably greater losses than that. Indeed, frequency selectivity is often thought to not show signs of deterioration until the loss is greater than about 30–40 dB. It may be that aspirin causes a more potent effect on OHCs than sensorineural hearing loss, and it is also possible that some individuals with permanent loss have damage to other structures (e.g., inner hair cells), either alone or in combination with damage to OHCs. Alternatively, or in addition, it may be that the within-subject design is simply more sensitive to effects of OHC dysfunction. For example, in the present study, aspirin produced average thresholds of about 15–25 dB HL (compared to 0–10 dB HL without aspirin). Although this represents a significant threshold shift or hearing loss for these individuals, most individuals with thresholds of 15–25 dB HL would generally be classified into a ''normal-hearing'' group. In fact, this ''normal-hearing'' group could have varying, although mild, degrees of cochlear damage, which would make it more difficult to distinguish their performance from that of a group of subjects classified as ''hearing impaired.'' More research should be done to determine whether similar amounts of AIHL and permanent loss yield similar performance. Currently, we have tested two individuals with sensorineural hearing losses similar in configuration and degree to those with AIHL on the same three experiments described here. Their results are presented in the Appendix and suggest that there are some similarities between AIHL and mild sensorineural hearing loss on these nonlinear measures. Consequently, at least for these measures, aspirin may be a good model for permanent loss. Further, the results of the present study indicate that aspirin can be an effective tool for studying the contribution of OHCs to psychophysical phenomena, and that even small amounts of AIHL can alter performance on psychophysical measures of auditory nonlinearities.

ACKNOWLEDGMENTS

This research was supported by NIDCD Grant No. DC01376 and a research award from the American Academy of Audiology Research Award Program. This research is based on a doctoral dissertation submitted by the first author to the Graduate College at Arizona State University. We thank Marjorie Leek for assistance with deriving the filter shapes in experiment 1, Dennis McFadden for providing an example of his aspirin protocol, and Andrew Oxenham and Van Summers for their helpful comments on an earlier version of this manuscript.

APPENDIX: COMPARISON WITH PERMANENT SENSORINEURAL HEARING LOSS

To determine whether similar degrees of AIHL and permanent loss yield similar performance on the three psychophysical measures examined here, we compared the results from the six individuals with AIHL to results from two individuals with sensorineural hearing losses that were similar in configuration and degree to those with AIHL. All procedures and stimuli were as described in each of the experiments above. Subject ages and quiet thresholds for the 15-

TABLE AI. Ages and quiet thresholds (in dB SPL) for the 15- and 300-ms signals. Results are shown averaged across the six listeners with AIHL and for the two hearing-impaired subjects, HI1 and HI2. The differences between the thresholds (in dB) for the 15- and 300-ms signals are shown in boldface.

Subject	Age (years)	Signal duration (ms)	Signal frequency (Hz)			
			675	750	1500	3000
Average AIHL	27	15	43.6	29.8	28.3	28.7
		300	27.3	15.2	15.6	18.8
		Difference	**16.3**	**14.6**	**12.7**	**9.9**
HI1	67	15	35.3	31.7	36.6	33.8
		300	17.9	17.4	28.6	26.2
		Difference	**17.4**	**14.3**	**8.0**	**7.6**
HI2	42	15	42.1	42.4	41.8	52.7
		300	28.6	27.9	31.5	44.1
		Difference	**13.5**	**14.5**	**10.3**	**8.6**

TABLE AII. Results for the three experiments are shown averaged across the six listeners with AIHL and for the two hearing-impaired subjects. The asterisks (*) in experiment 1 indicate values that may have been overestimated by the fitting procedure.[3] The masked thresholds for experiment 2 are expressed in dB SPL, and the slopes in experiment 3 are expressed in dB/dB.

		Average AIHL		HI1		HI2	
		750 (Hz)	3000 (Hz)	750 (Hz)	3000 (Hz)	750 (Hz)	3000 (Hz)
Experiment 1							
ERB	30 dB	118	468	181	354	194	915
	50 dB	125	765	167	438	177	870
p_l	30 dB	26.0	24.0	23.3	37.6	15.6	16.6
	50 dB	21.1	11.0	21.5	20.3	20.5	9.8
p_u	30 dB	25.4	27.7	13.1	31.1	16.2	10.8
	50 dB	27.7	27.1*	15.3	42.3*	14.6	23.2*
r	30 dB	−32.3	−35.8	−29.6	−40.1	−24.4	−54.2
	50 dB	−68.5	−58.9	−62.1	−79.1	−31.7	−67.9
Experiment 2							
M		46.4	47.2	47.5	48.2	59.3	63.5
M+S (1.2)		49.5	45.0	54.3	49.8	59.0	66.5
M+S (1.3)		48.1	46.9	54.1	50.7	59.3	65.5
Experiment 3							
Slope		0.81	1.26	0.93	1.12	0.81	0.76

and 300-ms signals are shown in Table AI. In addition, boldface entries represent the amount of temporal integration. The amount of integration for HI1 and HI2 was similar to that for the subjects with AIHL at 375 and 750 Hz, but smaller than that at 1500 and 3000 Hz, possibly reflecting the fact that HI1 and HI2 had somewhat greater losses than did the subjects with AIHL at those two higher frequencies.

Several measures from the derived filter shapes (ERB, p_l, p_u, and r) from experiment 1 are given in Table AII for the results averaged across the subjects with AIHL and for the two hearing-impaired subjects. In general, there was only a small effect of level (a decrease of 6%–8%) on filter bandwidth for both hearing-impaired subjects at 750 Hz; at 3000 Hz, the bandwidth decreased slightly (5%) for HI2, and increased by 24% for HI1. Thus permanent loss tended to reduce the effect of level even more than did AIHL, a finding that may reflect the somewhat greater permanent losses. The results for HI1 and HI2 are consistent with the results of Rosen *et al.* (1998), who reported reduced or absent effects of level on filter shapes in two individuals with very mild permanent hearing impairments. The masked thresholds from the three masker conditions [M, M+S (1.2), and M+S (1.3)] of experiment 2 are also given in Table AII. There was no suppression obtained at either signal frequency for either hearing-impaired subject, and the suppressors typically produced masking at 750 Hz, as they did in the case of AIHL. And finally, the slopes of the lines fitted to the masking functions of experiment 3 are given in the bottom of Table AII. For both subjects, the slopes of the functions (in dB/dB) were generally similar at the two signal frequencies and were close to, or slightly less than, 1.0 dB/dB. The largest difference in slope between the subjects with AIHL and those with permanent loss occurred for HI2 at 3000 Hz, possibly reflecting the somewhat greater hearing loss for that subject at that frequency. Thus, preliminary results suggest that there are some similarities between AIHL and mild, sensorineural hearing loss on these nonlinear measures. However, the relatively large amount of individual variability in the aspirin conditions, and even in the results of the two hearing-impaired individuals, makes definitive conclusions difficult. Additional research with individuals with mild hearing losses would help clarify the similarities or differences between these two groups.

[1]Subject S5 showed a clear improvement in threshold (about 7 dB) on day 2 at 1500 Hz; on days 3 and 4 threshold was about equal to that obtained in the baseline (no-aspirin) condition. Interestingly, subject S1 showed a similar improvement at that frequency on day 1 (threshold shift not shown, as thresholds were measured on the first day only for S1); it is unknown whether others would have shown a similar effect. This improvement was thought to reflect a reduction in masking by a spontaneous otoacoustic emission (SOAE) at or near 1500 Hz. That is, thresholds were higher prior to aspirin use due to masking by the SOAE, but then improved during aspirin use as a result of the aspirin eliminating the SOAE (McFadden and Plattsmier, 1984; Long and Tubis, 1988). A search for SOAEs in S1 several months after the conclusion of the experiments revealed an emission at 1514 Hz (as well as at 2142, 2661, and 2820 Hz). Although the presence of SOAEs was not evaluated in S5, it seems likely that a reduction in emission level was responsible for the improvement in threshold with aspirin for both subjects.

[2]Due to the limited amount of time available for testing during aspirin administration, only five notchwidth conditions, rather than the seven from Hicks and Bacon (1999), were tested (the two most-distal asymmetric notchwidth conditions were eliminated). Thus, for comparison with the aspirin data, the normal-hearing data from Hicks and Bacon (1999) were refitted with the same five notchwidth thresholds. It is important to note that the filter shapes were very similar whether they were derived from five or seven thresholds. Stone *et al.* (1992) have also shown that these five notchwidths will give essentially the same filter characteristics as a larger number of notchwidths.

[3]Several p_u values in Tables III and AII may have been overestimated by the fitting procedure: Glasberg and Moore (1990) have suggested that when the p value for one side of a filter is more than twice the value of the other side, the larger value is poorly defined.

[4]Alternatively, the decrease in threshold could reflect a cueing effect (Moore, 1981; Moore and Glasberg, 1982; Neff, 1985,1986). As discussed

in Hicks and Bacon (1999), however, the results presented here are probably not influenced by cueing. This is based on the signal level at threshold in the *M* condition generally being lower than the level of the masker, and the slope of the growth-of-masking function (masker and signal equal in frequency) being less than 1.0 [see Figs. 3 and 4 and Table IV in the present study, and Fig. 2 from Hicks and Bacon (1999) for these same subjects].

ANSI (**1989**). ANSI S3.6-1989, ''Specifications for Audiometers'' (American National Standards Institute, New York).

Bacon, S. P., and Takahashi, G. A. (**1992**). ''Overshoot in normal-hearing and hearing-impaired subjects,'' J. Acoust. Soc. Am. **91**, 2865–2871.

Beveridge, H. A., and Carlyon, R. P. (**1996**). ''Effects of aspirin on human psychophysical tuning curves in forward and simultaneous masking,'' Hearing Res. **99**, 110–118.

Bonding, P. (**1979**). ''Critical bandwidth in patients with a hearing loss induced by aspirin,'' **18**, 133–144.

Carlyon, R. P., and Butt, M. (**1993**). ''Effects of aspirin on human auditory filters,'' Hearing Res. **66**, 233–244.

Day, R. O., Graham, G. G., Bieri, D., Brown, M., Cairns, D., Harris, G., Hounsell, J., Platt-Hepworth, S., Reeve, R., Sambrook, P. N., and Smith, J. (**1989**). ''Concentration-response relationships for salicylate-induced ototoxicity in normal volunteers,'' Br. J. Clin. Pharmacol. **28**, 695–702.

Dubno, J. R., and Schaeffer, A. B. (**1992**). ''Comparison of frequency selectivity and consonant recognition among hearing-impaired and masked normal-hearing listeners,'' J. Acoust. Soc. Am. **91**, 2110–2121.

Festen, J. M., and Plomp, R. (**1983**). ''Relations between auditory functions in impaired hearing,'' Hearing Res. **16**, 251–260.

Fitzgibbons, P. J., and Wightman, F. L. (**1982**). ''Gap-detection in normal and hearing-impaired listeners,'' J. Acoust. Soc. Am. **72**, 761–765.

Florentine, M., Fastl, H., and Buus, S. (**1988**). ''Temporal integration in normal hearing, cochlear impairment, and impairment simulated by masking,'' J. Acoust. Soc. Am. **84**, 195–203.

Glasberg, B. R., and Moore, B. C. J. (**1986**). ''Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments,'' J. Acoust. Soc. Am. **79**, 1020–1033.

Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hearing Res. **47**, 103–138.

Hicks, M. L., and Bacon, S. P. (**1999**). ''Psychophysical measures of auditory nonlinearities as a function of frequency in individuals with normal hearing,'' J. Acoust. Soc. Am. **105**, 326–338.

Kakehata, S., and Santos-Sacchi, J. (**1996**). ''Effects of salicylate and lanthanides on outer hair cell motility and associated gating charge,'' J. Neurosci. **16**, 4881–4889.

Lee, J. M., and Bacon, S. P. (**1998**). ''Psychophysical suppression as a function of signal frequency: Noise and tonal maskers,'' J. Acoust. Soc. Am. **104**, 1013–1022.

Leek, M. R., and Summers, V. (**1993**). ''Auditory filter shapes of normal-hearing and hearing-impaired listeners in continuous broadband noise,'' J. Acoust. Soc. Am. **94**, 3127–3137.

Leshowitz, B., and Lindstrom, R. (**1977**). ''Measurement of nonlinearities in listeners with sensorineural hearing loss,'' in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London), pp. 283–292.

Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Long, G. L., and Tubis, A. (**1988**). ''Modification of spontaneous and evoked otoacoustic emissions and associated psychoacoustic microstructure by aspirin consumption,'' J. Acoust. Soc. Am. **84**, 1343–1353.

McCabe, P. A., and Dey, F. L. (**1965**). ''The effect of aspirin upon auditory sensitivity,'' Ann. Otol. Rhinol. Laryngol. **74**, 312–325.

McFadden, D., and Champlin, C. A. (**1990**). ''Reductions in overshoot during aspirin use,'' J. Acoust. Soc. Am. **87**, 2634–2642.

McFadden, D., and Plattsmier, H. S. (**1984**). ''Aspirin abolishes spontaneous oto-acoustic emissions,'' J. Acoust. Soc. Am. **76**, 443–448.

McFadden, D., Plattsmier, H. S., and Pasanen, E. G. (**1984a**). ''Aspirin-induced hearing loss as a model of sensorineural hearing loss,'' Hearing Res. **16**, 251–260.

McFadden, D., Plattsmier, H. S., and Pasanen, E. G. (**1984b**). ''Temporary hearing loss induced by combinations of intense sounds and nonsteroidal anti-inflammatory drugs,'' Am. J. Otolaryngol. **5**, 235–241.

Moore, B. C. J. (**1981**). ''Interactions of masker bandwidth with signal duration and delay in forward masking,'' J. Acoust. Soc. Am. **70**, 62–68.

Moore, B. C. J. (**1991**). ''Characterization and simulation of impaired hearing: Implications for hearing aid design,'' Ear Hear. **12** (Suppl.), 154–161.

Moore, B. C. J. (**1995**). *Perceptual Consequences of Cochlear Damage* (Oxford U.P., Oxford).

Moore, B. C. J., and Glasberg, B. R. (**1982**). ''Contralateral and ipsilateral cueing in forward masking,'' J. Acoust. Soc. Am. **71**, 942–945.

Moore, B. C. J., and Glasberg, B. R. (**1983**). ''Growth of forward masking for sinusiodal and noise maskers as a function of signal delay; implications for suppression in noise,'' J. Acoust. Soc. Am. **73**, 1249–1259.

Murnane, O., and Turner, C. W. (**1991**). ''Growth of masking in sensorineural hearing loss,'' Audiology **30**, 275–285.

Murugasu, E., and Russell, I. J. (**1995**). ''Salicylate ototoxicity: The effects on basilar membrane displacement, cochlear microphonics, and neural responses in the basal turn of the guinea pig cochlea,'' Aud. Neurosci. **1**, 139–150.

Myers, E. N., Bernstein, J. M., and Fostiropolous, G. (**1965**). ''Salicylate ototoxicity,'' N. Engl. J. Med. **273**, 587–590.

Neff, D. L. (**1985**). ''Stimulus parameters governing confusion effects in forward masking,'' J. Acoust. Soc. Am. **78**, 1966–1976.

Neff, D. L. (**1986**). ''Confusion effects with sinusoidal and narrow-band noise forward maskers,'' J. Acoust. Soc. Am. **79**, 1519–1529.

Nelson, D. A. (**1991**). ''High-level psychophysical tuning curves: Forward masking in normal-hearing and hearing-impaired listeners,'' J. Speech Hear. Res. **34**, 1233–1249.

Nelson, D. A., and Schroder, A. C. (**1997**). ''Linearized response growth inferred from growth-of-masking slopes in ears with cochlear hearing loss,'' J. Acoust. Soc. Am. **101**, 2186–2201.

Oxenham, A. J., Moore, B. C. J., and Vickers, D. A. (**1997**). ''Short-term temporal integration: Evidence for the influence of peripheral compression,'' J. Acoust. Soc. Am. **101**, 3676–3687.

Oxenham, A. J., and Plack, C. J. (**1997**). ''A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing,'' J. Acoust. Soc. Am. **101**, 3666–3675.

Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. (**1982**). ''The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold,'' J. Acoust. Soc. Am. **72**, 1788–1803.

Plomp, R., and Bouman, M. A. (**1959**). ''Relation between hearing threshold and duration for tone pulses,'' J. Acoust. Soc. Am. **31**, 749–758.

Rhode, W. S., and Cooper, N. P. (**1993**). ''Two-tone suppression and distortion production on the basilar membrane in the hook region of cat and guinea pig cochleae,'' Hearing Res. **66**, 31–45.

Rosen, S., and Stock, D. (**1992**). ''Auditory filter bandwidths as a function of level at low-frequencies (125 Hz–1 kHz),'' J. Acoust. Soc. Am. **92**, 773–781.

Rosen, S., Baker, R. J., and Darling, A. (**1998**). ''Auditory filter nonlinearity at 2 kHz in normal hearing listeners,'' J. Acoust. Soc. Am. **103**, 2539–2550.

Ruggero, M. A., Robles, L., and Rich, N. C. (**1992**). ''Two-tone suppression in the basilar membrane of the cochlea: Mechanical basis of auditory-nerve rate suppression,'' J. Neurophysiol. **68**, 1087–1099.

Russell, I., and Schauz, C. (**1995**). ''Salicylate ototoxicity: Effects on the stiffness and electromotility of outer hair cells isolated from the guinea pig cochlea,'' Aud. Neurosci. **1**, 309–319.

Sachs, M. B., and Kiang, N. Y. S. (**1968**). ''Two-tone inhibition in auditory-nerve fibers,'' J. Acoust. Soc. Am. **43**, 1120–1128.

Schmiedt, R. A. (**1982**). ''Boundaries of two-tone rate suppression of cochlear-nerve activity,'' Hearing Res. **7**, 335–351.

Schmiedt, R. A., Zwislocki, J. J., and Hamernik, R. P. (**1980**). ''Effects of hair cell lesions on responses of cochlear nerve fibers. I. Lesions, tuning curves, two-tone inhibition, and responses to trapezoidal-wave patterns,'' J. Neurophysiol. **43**, 1367–1389.

Shehata, W. E., Brownell, W. E., and Dieler, R. (**1991**). ''Effects of salicylate on shape, electromotility and membrane characteristics of isolated outer hair cells from guinea pig cochlea,'' Acta Oto-Laryngol. **111**, 707–718.

Smoorenburg, G. F. (**1980**). ''Effects of temporary threshold shift on combination-tone generation and on two-tone suppression,'' Hearing Res. **2**, 347–355.

Stelmachowicz, P. G., Lewis, D. E., Larson, L. L., and Jesteadt, W. (**1987**). ''Growth of masking as a measure of response growth in hearing-impaired listeners,'' J. Acoust. Soc. Am. **81**, 1881–1887.

Stone, M. A., Glasberg, B. R., and Moore, B. C. J. (**1992**). ''Simplified measurement of impaired auditory filter shapes using the notched-noise method,'' Br. J. Audiol. **26**, 329–334.
Stypulkowski, P. H. (**1990**). ''Mechanism of salicylate ototoxicity,'' Hearing Res. **46**, 113–146.
Wier, C. C., Pasanen, E. G., and McFadden, D. (**1988**). ''Parial dissociation of spontaneous otoacoustic emissions and distortion products during aspirin use in humans,'' J. Acoust. Soc. Am. **84**, 230–237.
Wightman, F. McGee, T., and Kramer, M. (**1977**). ''Factors influencing frequency selectivity in normal and hearing-impaired listeners,'' in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, New York), pp. 295–306.

Evaluation of spectral enhancement in hearing aids, combined with phonemic compression

Bas A. M. Franck,[a] C. Sidonne G. M. van Kreveld-Bos, and Wouter A. Dreschler
Academic Medical Center, Clinical and Experimental Audiology, Meibergdreef 9, 1105 AZ, Amsterdam, The Netherlands

Hans Verschuure
Erasmus University, Department ENT/Audiological Center, Dr. Molenwaterplein 40, 3015 GD, Rotterdam, The Netherlands

(Received 22 January 1998; revised 15 October 1998; accepted 6 May 1999)

In this study, the separate and combined effects on speech perception of compensation of the reduced dynamic range by compression and compensation of the reduced frequency resolution by spectral enhancement is investigated. The study has been designed to compare the effects of signal processing on monosyllabic consonant–vowel–consonant words for hearing-impaired listeners in conditions of quiet, fluctuating noise, and continuous noise. Speech perception of spectrally enhanced speech was compared with unprocessed speech. In addition, a comparison was made between combinations of spectrally enhanced speech and two types of phonemic compression. In the past, the definition ''syllabic compressor'' is often used to indicate fast compressors. However, the time constants of the fast compressors used in this study are so short that syllabic has become an inappropriate term. Moreover, intelligibility tests were performed in which scores were acquired of monosyllabic words, and their constituent ''phonemic'' parts. Therefore, the definitions ''phoneme'' and phonemic will be used throughout this paper. In one condition, spectral enhancement produced significant improvements for vowel perception. But, this was counteracted by deterioration of the consonant scores for all but one subject. In general, the best overall scores for consonant–vowel–consonant words were obtained in the unprocessed condition. After the spectral enhancement, a single-channel phonemic compressor added no improvement. There are indications that a multichannel phonemic compressor and spectral enhancement have opposite effects, because the scores for this combination are, in general, the lowest. [S0001-4966(99)04208-3]

PACS numbers: 43.66.Dc, 43.66.Sr, 43.66.Ts [JWH]

INTRODUCTION

In order to help hearing-impaired people distinguish speech in noisy environments, several algorithms have been proposed. Two different approaches are possible: reduction of the noise with a noise-reduction technique, or enhancement of the speech. Focus on enhancement of the speech signal can result in manipulations in the time structure or in the frequency spectrum of the signal. Two possible types of signal enhancement are phonemic compression and spectral enhancement, respectively.

According to Baer *et al.* (1993a), the recognition of speech sounds is based on the use of spectral shape to detect spectral features, especially formants. The recognition may be facilitated by spectral enhancement. On the other hand, it is questionable whether this spectral enhancement can be beneficial for these persons, because the enhancement process takes place *before* the deteriorated auditory filters do their destructive work.

Phonemic compression is assumed to be beneficial for hearing-impaired individuals for another reason: it puts the fast variations of the speech signal, which lie roughly in a 30-dB range, into the narrower dynamic range of the hearing-impaired subjects. In this way, the essential parts of speech will be audible again. By fast compression, the level difference between the successive phonemes can also be reduced (Verschuure and Dreschler, 1996). Therefore, phonemic compression ought to be beneficial for the perception of consonants, because the temporal masking effects will be reduced. In order to compress quickly, short attack and release times are used. A possible disadvantage could be the reduction of the modulations in the speech signal. To avoid too many distortions in the waveforms, compression ratios are kept low, at a value of 2.

This investigation concentrates primarily on the effect of spectral enhancement on speech intelligibility. But, it also focuses on combinations of spectral enhancement and phonemic compression.

Two phonemic compressors have been used, (i) an eight-channel phonemic compression system described by van Harten-De Bruijn *et al.* (1997). The assumed advantage of this system is the fact that the parameters in each filter band can be set to the needs of the hearing-impaired person in that particular frequency region. (ii) A single-channel phonemic compressor as developed by Verschuure *et al.* (1994). This system is based on the hypothesis that the ears of our specific group of hearing-impaired individuals mainly contain two different zones: a low-frequency zone in which lin-

[a]Electronic mail: B.A.Franck@amc.uva.nl

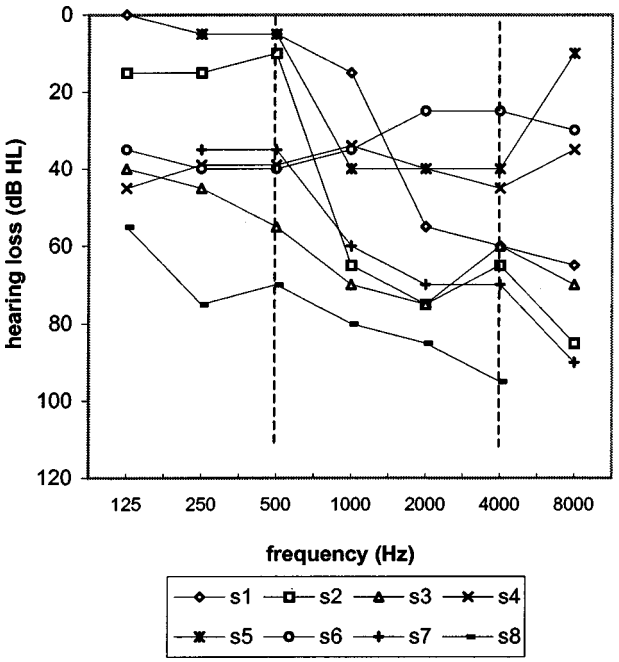


FIG. 1. Hearing losses (dB HL) for eight subjects as a function of frequency (Hz). The subjects are indicated as S1 to S8.

ear amplification can be applied, and a high-frequency zone where a phonemic compressor is required.

I. METHOD

A. Subjects

Eight subjects participated in this project. The subjects were 16 to 18 years old, all studying at the Alexander Graham Bell school in Amsterdam, a school for hearing-impaired pupils. They were paid volunteers. With the exception of two subjects (S1 and S5), all subjects have linear hearing aids. Only one subject (S2) wears two hearing aids.

For each subject, the hearing loss (in dB) of the best ear was measured at 0.125, 0.25, 0.5, 1.0, 2.0, 4.0, and 8.0 kHz with adequate masking of the contralateral ear. They were selected primarily with a cochlear hearing loss. Only one subject (S6) had a mixed loss. The audiograms of the eight participants are presented in Fig. 1. Most audiograms are sloping towards the higher frequencies, except for subject S4 and S6. S5 has a ''cookie-byte'' audiogram between 1.0 and 4.0 kHz.

B. Experimental design

Two different kinds of tests were carried out: psychophysical and intelligibility tests. The three different psychophysical tests, pure-tone audiometry, loudness scaling, and measurement of auditory filters, give information about the residual capacities of the ears of the individual subjects. The outcomes can be correlated with the individual speech scores. Eventually, significant correlations will provide information about the relations between specific psychophysical parameters and the intelligibility of processed speech.

C. Loudness scaling

We measured the individual loudness perception using the Würzburger Hörfeld Skalierung (Helbrück and Moser, 1985). This method prescribes that subjects listen to 1/3-octave noises of different intensities centered at four different frequencies: 0.5, 1.0, 2.0, and 4.0 kHz, presented in random order. Their task was to give an indication of the loudness of the noises. They could respond by giving any integer between 0 and 50, corresponding with not heard and too loud, respectively. The numbers 5, 15, 25, 35, and 45 corresponded to very soft, soft, middle loud, loud, and very loud, respectively. An automatic-fit procedure yields a curve relating subjective loudness to input level (given in sound-pressure level, dB) for each of the four frequency areas. Both the steepness and the horizontal position of the fitted curves carry information about the loudness perception of the subjects. The steepness is directly related to the subject's dynamic range (DR) per frequency. We calculated dynamic range values from the level differences corresponding to the subjective responses 40 and 10 (in categorical loudness units), and the indications of loud to very loud and soft to very soft, respectively. For each subject, the most comfortable level corresponds with the categorical unit 25. This level was used for the intelligibility test.

D. Auditory filters

Frequency selectivity was measured at 0.5 and 3 kHz, using a notched-noise technique and a Von Békésy procedure, described by Leeuw and Dreschler (1994). The parameter is the positioning of the notched-noise bands, relative to the center frequency f_c of the probe tone. As described in Leeuw and Dreschler (1994), this notched-noise method for clinical purposes includes five different noises. The probe tones are, except for the noise without notch, asymmetrically located in the notch. The notched-noise bands are either 600 or 1000 Hz wide. For two noise conditions, the higher frequency noise band is situated 400 Hz higher than the signal, so the lower band is either 200 or 600 Hz lower than the probe tone. For the other two conditions the opposite is true: the lower frequency noise band lies 400 Hz lower, while the higher noise band is either 200 or 600 Hz higher. With these five conditions, it is possible to get a reliable estimate of the auditory filter in a reasonable time. Each condition was tested twice, unless the thresholds obtained differed more than 5 dB. In that case, a third measurement was included.

The thresholds were fitted by means of the program FASTFIT (Glasberg and Moore, 1990), which yielded parameters for the lower and upper skirts of the triangular-shaped auditory filter, PL and PU, respectively. The equivalent rectangular bandwidths (ERB) were calculated from PL and PU.

E. Intelligibility test

In the main tests, the intelligibility measurements, the subjects were asked to listen to sentences. These consist of target words in five different carrier phrases. The extra words (two words, three phonemes) that always precede the target word are especially important in difficult listening situations, i.e., when background noise is present. The target words are

TABLE I. Four processing conditions determined by spectral enhancement and amplification type.

Processing	Amplification		
	Linear	**S**ingle-channel phonemic compression	**M**ultichannel phonemic compression
Unprocessed	UL	…	…
Spectrally **E**nhanced	EL	ES	EM

nonsense words, always consisting of three phonemes: consonant–vowel–consonant (CVC). Four categories of data were obtained: the three phonemes and the overall scores. The male spoken Dutch sentences come from Hans Boer lists, developed by the TNO-IZF Institute in Soesterberg (Steeneken *et al.*, 1990). Each list consists of 51 sentences. In order to avoid fatigue, the subjects listened to no more than four lists per session.

The test is quite elaborate because we are dealing with three dimensions: noise type, signal-to-noise ratio (S/N), and processing condition. We selected four processing conditions from the combinations of spectral enhancement (**U**nprocessed and spectrally **E**nhanced) and the presence of a certain kind of compression (linear (no compression), **S**ingle channel compression, or **M**ultichannel compression, see Table I). This results in either an ES condition, in which the signal is spectrally enhanced and compressed in a single channel, or the EM condition achieved by spectral enhancement and a multichannel compression algorithm. The reference condition EL for these two compression conditions is spectral enhancement without compression. The reference condition UL with respect to spectral enhancement also uses linear amplification, but is unprocessed. We cannot compare one of the conditions including compression (ES and EM) with the unprocessed UL condition directly, because this would presuppose that we used two linear, independent processes (spectral enhancement and phonemic compression). The other possible conditions—US and UM—are not included in the test for reasons of comprehensiveness and necessity. We are not interested in the effects of the phonemic compressors alone, because these have been studied extensively (e.g., Verschuure *et al.*, 1994).

In addition, we used three acoustical conditions in which speech is presented: noiseless (q), fluctuating noise (f), and continuous noise (c). In order to save time, the quiet condition (q) was the only one that was retested. The fluctuating noise was a male voice played backward, reading Dutch sentences. The continuous noise consisted of several speaking persons and was called restaurant noise, simulating that specific kind of background. The speech signals were presented at two signal-to-noise ratios relative to the noises: S/N=+4 and S/N=−1. For both speech and noise signals, the average level (time constant=1000 ms) was measured. In total, 24 different conditions (3 noises×2 S/N ratios×4 processing schemes) were presented to the subjects.

The order of the processing conditions and the order of the noise conditions for each day were randomized for each subject. For the processing conditions, this was done by means of a Latin square.

F. Experimental setup

The subjects listened to computer-generated stimuli through headphones. They used a keyboard to enter the CVC words they heard. Because the spectral enhancement was done off-line, the computer was only used for processing the compression algorithms. The computer output was converted to analog, and fed into the audiometer. During this stage the desired level for each individual could be adjusted, so the subject could hear the stimuli at the most comfortable level (MCL). It appeared necessary to implement a high-pass and a low-pass filter to avoid distortions. The cutoff frequencies were 0.3 and 6 kHz, respectively.

G. Spectral enhancement (E)

The sentences (in the presence or absence of noise) were recorded on audio CD and were processed off-line by spectral enhancement in Cambridge by our HEARDIP partners Baer, Stone, and Moore. For computational reasons, the signals were sampled at a 16-kHz rate and low-pass filtered at 4 kHz. This may be expected to be broad enough for speech. In order to manipulate the signal in the frequency domain, the waveform of the input had to be translated to a spectral representation. Because the signal varied continuously, we needed an averaging process over a specified period of time. Baer, Stone and Moore chose the size of the Hamming window to be 1024 points, 32.0 ms long. This windowed segment was converted to the frequency domain, using the fast Fourier transform (FFT) process. After the spectrum was processed, the spectral information was restranformed into the time domain with an inverse FFT. All segments were added using an overlap-add procedure to get the ultimately processed waveform.

In the spectral enhancement process itself (which is in conformity with the method of processing of experiment 3 in the study of Baer *et al.*, 1993a), only the magnitudes were manipulated. First, the excitation pattern was calculated using the input spectrum. This pattern was used to filter out excessive detail which would not be resolved by the ear anyway. It is a representation of the frequency contents after normal auditory filtering. Then, the new pattern was subjected to a function which enhanced the peaks and valleys. In order to create a gain function based on the readout of these local extremes, a function was used that is composed of the difference between a positive and a negative rounded exponential function. For both functions, the bandwidths are varied with center frequency, based on ERB calculations of auditory filters by the function $\mathrm{ERB}=6.23f^2+93.39f+28.52$ (designed earlier by Moore and Glasberg, 1983). Baer *et al.* (1993a) also performed informal listening experiments to determine specific multiplication factors for these functions, performed some scaling, and specified the gain function based on the enhancement function mentioned. To shape the output spectrum, the gain function was convoluted with the input spectrum. The minima and maxima were enhanced in this way, and the fine structure of the original spectrum was preserved. In the inverted FFT, the adapted magnitude data were combined with the original phase data and transformed into the time domain. Note that the convolution process was

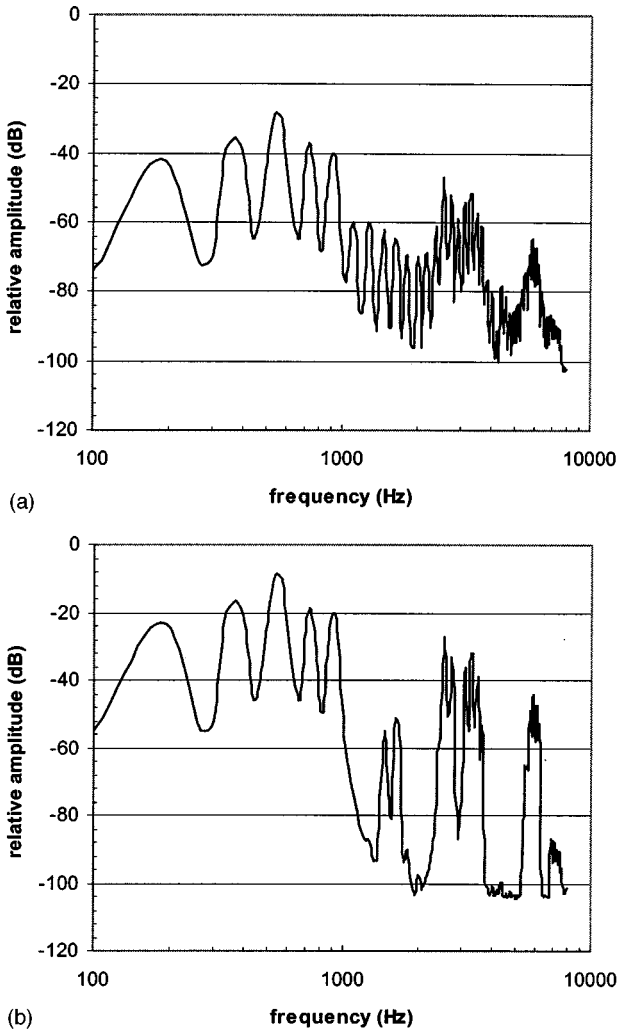


FIG. 2. (a) Spectral representation of the unprocessed vowel /o/ of the Dutch nonsense word ''woom,'' and (b) spectrally enhanced vowel /o/, when the sharpening procedure developed by the Cambridge group Baer, Stone, and Moore (1993b) has been used. For both figures, a window size of 30 ms has been used to calculate the average spectrum.

preceded by a pre-emphasis on the enhancement filters, because the spectrum drops off at a rate of 6 dB octave. In other words, the spectrum of the input signal was prewhitened, so the peaks would tend to have equal weight across the spectrum.

The effect of spectral enhancement is illustrated in Fig. 2. The spectral content of the spectrally enhanced Dutch vowel /o/ is compared with the unprocessed vowel /o/. To exaggerate the effects of spectral enhancement, the amount of enhancement chosen is larger than is actually used. These pictures are mean fast Fourier transforms, using a Hanning window size of 30 ms, and sampled at 16 kHz. A logarithmic scale is used for the abscissa because of cochlear properties.

H. Single-channel phonemic compression (S)

When one of the phonemic compressors is used, the setup contains a filterbank and a compression table. The filterbank that has been constructed contains finite impulse response (FIR) filters. These have a linear phase characteristic. In this setup, only one FIR filter is used with the shape of the mirrored audiogram of the hearing-impaired individuals. Since the majority of the subjects have a high frequency loss, the filter will look like a high-pass filter. An envelope detector controls the compressor. The result of this system is that, for low input levels, the response will look like the mirrored audiogram (compression controlled), whereas for high input levels the response will be almost flat. For intermediate levels, the responses look more like flattened versions of the mirrored audiogram. The compression ratio was set to a factor 2. This compressor uses $\tau_\alpha = 5$ ms and $\tau_r = 15$ ms. For the single-channel phonemic compression system, these settings obtain optimum performance, which appeared to be a compromise between comfort, i.e., compression ratio and knee-point settings, and speech intelligibility, which defines the choice of the time constants and delay time. The time constants were chosen high enough to effectively compress modulations below 20 Hz, which are considered to contain important speech information (Goedegebure, 1999).

Apart from this requirement, fast compression systems should suddenly react on fast level changes of transients. However, these higher and lower levels cannot be reached by the compressor that quickly, so over- and undershoots are created, respectively. Overshoots are most easily perceived, and therefore should be reduced. The delay time used to suppress overshoots is $\tau_{\mathrm{delS}} = 3$ ms. Verschuure and Dreschler (1996) reported on these technical considerations. The single-channel compressor uses a knee-point in the I–O curve, 28 dB below the overall rms value of the speech, which is 22 dB below the maximum input level. For input levels below this knee-point, the action of the system is linear.

I. Multichannel phonemic compression (M)

In multichannel compression systems, the spectral energy within channels determines the compressive activity for each channel. This ought to be advantageous, especially for hearing-impaired subjects with dynamic ranges that are highly frequency dependent (Verschuure *et al.*, 1993). In the multichannel setup, an envelope detector and a nonlinearity table are embedded in each of the eight frequency bands with fixed center frequencies around 543, 900, 1273, 1604, 2021, 2546, 3208, and 4041 Hz. These bands are 1/3 octave wide, except for the lowest frequency band. For computational reasons this band is 1 octave wide. The envelope detector, together with the nonlinearity table, takes care of the multiplication factor of that specific frequency region in order to determine the level of each channel. Frequency bands need additional amplification because the dynamic ranges and MCLs for every hearing-impaired subject depend on frequency. Therefore, we determined the difference between the MCL for each band and the MCL for 500 Hz, for each subject.

Contrary to the single-channel compressor, the eight-channel compressor does not use a knee-point. All input-levels below -62 dB rms and output levels above -20 dB rms have been removed, for all frequency bands. For high input levels, a limiter is implemented to avoid signals be-

TABLE II. Dynamic ranges (DR) and most comfortable levels (MCL) for four different frequencies (0.5, 1, 2, and 4 kHz) for the subjects S1 to S8. These variables are derived from the loudness data (using the Würzburger Hörfeld Skalierung). DR values correspond to the difference between the subjective indications loud to very loud and soft to very soft. MCL values correspond to the indication medium loud.

	Parameter (dB)							
Subject	DR0.5	DR1	DR2	DR4	MCL0.5	MCL1	MCL2	MCL4
S1	26	33	13	20	97	103	109	109
S2	51	26	22	39	97	101	114	109
S3	36	26	27	27	109	113	113	107
S4	32	31	35	30	89	86	92	95
S5	47	34	37	30	67	84	82	76
S6	44	46	62	59	98	95	91	83
S7	21	23	21	28	99	108	115	102
S8	32	26	26	12	109	106	114	115
normal hearing[a]	55	54	59	58	72	71	75	69

[a]Verschuure *et al.*, 1993.

coming uncomfortably loud. The knee-point of the limiter is set at a threshold above the peak levels of speech. Therefore, not much distortion of the speech signals is expected. Like the single-channel compressor, the time constants and compression ratio used for the multichannel compressor are chosen to give optimum performance on both speech intelligibility and comfort (Verschuure *et al.*, 1993; van Harten-De Bruijn *et al.*, 1997). For each channel, the compression ratio of the multichannel compressor is set at a CR=2. The compressor uses attack and release times of $\tau_\alpha = 0.5$ ms and $\tau_r = 10$ ms, in all channels. To reduce overshoots that can be interpreted as plosives, a digital delay was applied with a delay time of $\tau_{\mathrm{del}M} = 0.5$ ms.

II. RESULTS

A. Psychophysical data

The Würzburger Hörfeld Skalierung method measures loudness data in terms of categorical loudness units (CLUs). From these data, two important parameters can be extracted: dynamic ranges (DR) and most comfortable levels (MCL). These parameters are given by $\mathrm{DR}(f) = L_{\mathrm{CLU}=40}(f) - L_{\mathrm{CLU}=10}(f)$, and $MCL(f) = L_{\mathrm{CLU}=25}(f)$. Table II presents for each subject the dynamic ranges and most comfortable levels for four different frequencies, 0.5, 1.0, 2.0, and 4.0 kHz. Reference data for normal-hearing subjects measured with the same procedure are presented for purposes of comparison. Seven of the eight subjects clearly have a smaller dynamic range, especially for higher frequencies. Next to this, the levels at which the loudness of the speech is prescribed as most comfortable are shown for each subject and each frequency. The MCLs are clearly higher than normal.

Using the Von Békésy notched-noise method for measuring the auditory filters centered at 0.5 and 3.0 kHz, five threshold values were obtained for each frequency. The bandwidths (ERB) of the auditory filters have been calculated from the steepness parameters of the filter slopes, according to: $\mathrm{ERB} = 2fc/\mathrm{PL} + 2fc/\mathrm{PU}$ [ERB in Hz, f_c=center frequency in Hz, PL=slope of the lower skirt, PU=slope of upper skirt, Glasberg and Moore (1990)]. All three parameters are given for each subject in Table III. In the last row of the table, we calculated the ERB, PL, and PU values for the average normal-hearing person. A comparison between the

TABLE III. For all subjects, three auditory-filter parameters are given: the upper slope (PU), the lower slope (PL), and the rectangular equivalent bandwidth (ERB). Two center frequencies are used to obtain a typically high-frequency and low-frequency filter: 0.5 and 3 kHz. The ERB values according to the average normal-hearing subject are calculated by an equation suggested by Glasberg and Moore (1990): $\mathrm{ERB} = 24.7(4.37 f_c + 1)$. The filters are assumed to be symmetrical for normal-hearing persons, so PL=PU and therefore $\mathrm{PL} = 4000 f_c/\mathrm{ERB}$. ERB is given in Hz, while f_c is given in kHz. PL and PU are without dimensions. The normal-hearing data are calculated from data proceeding from a study by Verschuure *et al.* (1993).

	Parameter					
Subject	ERB0.5 (Hz)	PL0.5	PU0.5	ERB3 (Hz)	PL3	PU3
S1	112.5	17.0	18.6	1035	6.4	50.0
S2	83.0	22.2	26.3	1038	49.3	6.6
S3	141.0	17.0	12.2	1884	3.4	49.8
S4	102.5	19.9	19.1	645	12.7	34.9
S5	73.5	21.4	37.1	984	12.8	11.7
S6	115.0	21.5	14.7	318	42.9	33.6
S7	95.5	19.6	22.6	591	24.9	17.1
S8	139.0	13.8	15.1	744	18.6	14.2
normal hearing[a]	**78.7**	**25.4**	**25.4**	**348.5**	**34.4**	**34.4**

[a]Calculations of the ERB, PL, and PU values for the average normal-hearing person are based on the formulas $\mathrm{ERB} = 24.7(4.37 f_c/1000 + 1)$ and $\mathrm{PL} = \mathrm{PU} = 4 f_c/\mathrm{ERB}$ that originate from Glasberg and Moore (1990).

TABLE IV. Three-dimensional analyses of variance, using subject (8), algorithm (4), and noise type (5) as parameters. Non-significant interactions are abbreviated by n.s.

Design	Probability
Subject	<0.001
Algorithm	<0.01
Noise type	<0.001
Subject×algorithm	n.s.
Subject×noise type	<0.01
Algorithm×noise type	n.s.

data of the subjects and the reference data shows remarkable deviations, especially at 3 kHz (see ERB3, PU3, and PL3 for subjects 1, 2, and 3). The shallower slopes of subjects 1, 2, and 3 reflect very broad filters (large ERB3 values).

B. Speech perception data

1. Overall scores

A three-way analysis of variance was used for the three experimental variables: subjects (8), algorithms (4), and combinations of noise type and S/N ratio (5). As shown in Table IV, the main effects are all significant: $p<0.001$ for the parameter ''subject'' and ''noise type,'' $p=0.004$ for ''algorithm.'' The large variance between the noises is not surprising, because the noise type ''quiet'' is very different from fluctuating noise and continuous noise. *Post hoc* testing showed that the significant effect of algorithm is due to the difference between the scores in the EM condition and the original condition UL ($p=0.029$).

The parameters subject and noise type interact significantly ($p<0.01$). As indicated later, the scores of three subjects (S1, S3, and S8) are especially bad in the fluctuating noise condition with the lowest signal-to-noise ratio. But most important, the other interactions did not account for large amounts of variance, which facilitates the interpretation.

In Fig. 3 the averaged data of all subjects ($n=8$) are presented. The averaged data show the overall effects of spectral enhancement alone (EL: second bars), and combinations of spectral enhancement and compression (EM and ES, in the third and fourth bars, respectively). To determine the significance of the differences, Wilcoxon (matched-pairs signed rank sum) tests were done.

The scores are highest in the quiet condition, for a signal-to-noise ratio S/N=+4 the scores are moderate, and for S/N=−1 the scores are lowest. When the signal-to-noise ratios are the same, the scores in continuous and fluctuating noise are about equal. So, it seems that the subjects hardly benefit from the pauses in the fluctuating noise.

In the original condition (UL) the scores are always best, although the differences are not always statistically significant. The largest differences appear between the EM condition and one of the other processing conditions. There are significant differences in scores between the EM and UL condition ($p<0.02$), and between the EM and EL condition ($p<0.02$). Except for condition $c(-1)$, the scores for the EL condition and the ES condition are approximately equal.

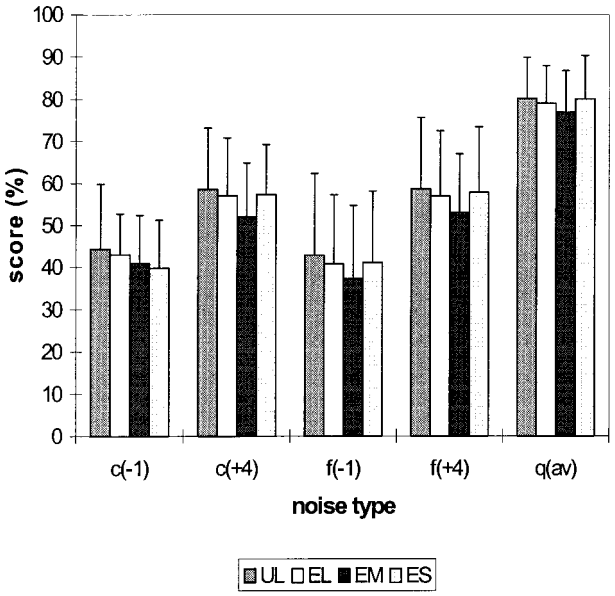


FIG. 3. Averaged CVC scores (for all eight subjects) as a function of noise type. The data are subdivided in different processing conditions: the first bar represents the original condition (UL), the second bar spectral sharpening (EL), the third bar the combination of spectral sharpening and multichannel compression (EM), and the fourth bar the combination of spectral sharpening and single-channel compression (ES). The noise types are abbreviated: continuous noise (c), fluctuating noise (f), and quiet (q). The S/N ratios are indicated between brackets.

2. Phoneme scores

The results can be analyzed in more detail by looking at the scores of the different parts of the CVC words: the initial and final consonants and the vowel. These three phonemes are defined as C1, C2, and V, respectively. We expect confusions to be mainly determined by noise type, so the S/N ratio will no longer be taken into account. Therefore, the two S/N ratios have been averaged both for the fluctuating and continuous noise conditions.

On top of the expected significance of the main effects, variance analyses of the different phonemes show more significant interactions than for the word scores (see Table V). There is a weak effect of ''subject×algorithm'' for C1. The significant interaction ''subject×noise'' type is found for all phonemes. An interaction between algorithms and noise types is only significant for the final consonants. In conclusion, these results suggest that the main trends are found for

TABLE V. Three-dimensional analyses of variance for the individual phonemes: the initial consonant (C1), the vowel (V), and the final consonant (C2). Due to the averaging on S/N ratio, the parameter noise type contains only three different types. Non-significant interactions are abbreviated by n.s.

	Probability		
Design	C1	V	C2
Subject	<0.001	<0.001	<0.001
Algorithm	<0.05	<0.001	<0.01
Noise type	<0.001	<0.001	<0.001
Subject×algorithm	<0.05	n.s.	n.s.
Subject×noise type	<0.05	<0.001	<0.01
Algorithm×noise type	n.s.	n.s.	<0.05

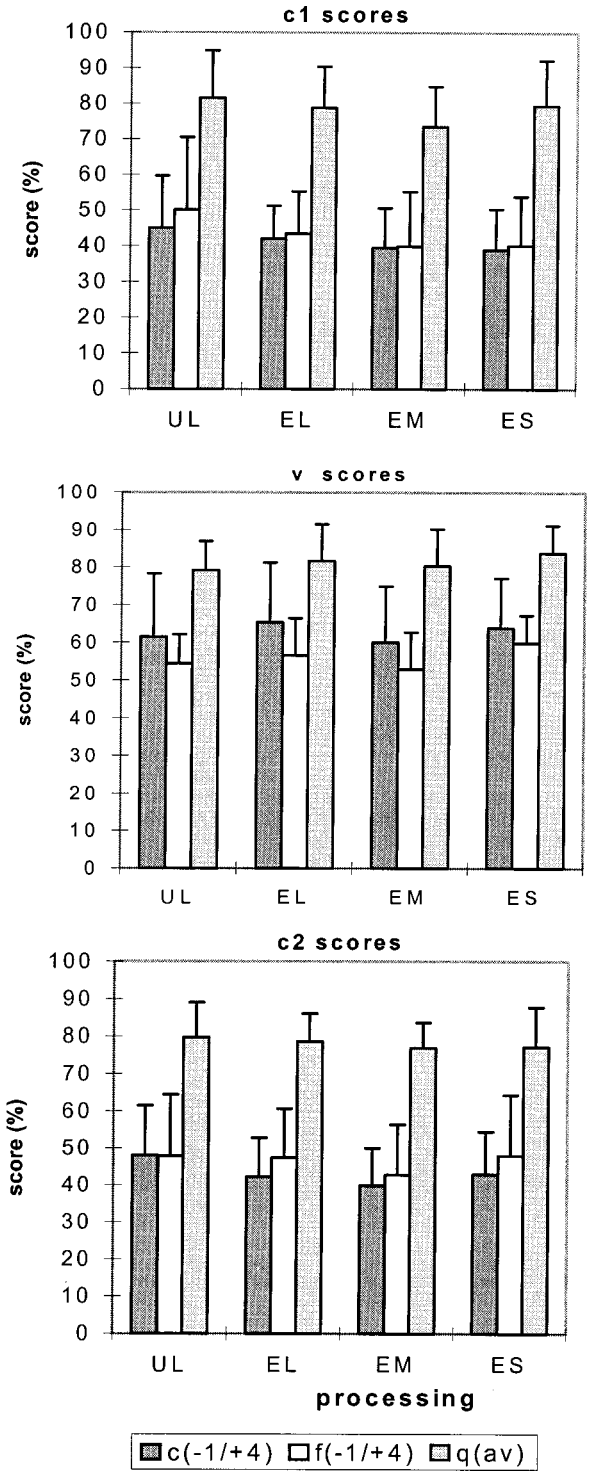


FIG. 4. Three representations concerning the effects of noise type. The noise types are abbreviated: continuous noise (c), fluctuating noise (f), and quiet (q). The averaged S/N ratios are indicated between brackets. On top, the initial consonant scores (C1) are depicted, in the middle the vowel scores (V), and at the bottom the final consonant scores (C2).

each phoneme, but some interaction effects may differ between phonemes.

In Fig. 4(a), (b), and (c) we focus on the differences between quiet, fluctuating, and continuous noise, averaged across the S/N ratio and test–retest for each of the signal-processing conditions. In the original condition (UL) the C1, V, and C2 scores in quiet are very similar. In the conditions with background noise, the identification of initial and final consonants is clearly more difficult than for vowels. This is especially true of the continuous noise condition ($p<0.01$ for both consonants). For fluctuating noise, only initial consonant scores are significantly different from vowel scores ($p<0.05$).

In the quiet condition, the consonant scores for UL tend to be higher than in other algorithms ($p<0.05$ for UL-EM in initial consonants). For vowels, however, there is a trend that higher scores are achieved in the conditions where spectral enhancement is used alone (EL), and where the spectral enhancement is combined with single-channel compression (ES), but these effects are not significant. These trends in the quiet condition are also present in continuous and fluctuating noise. Only for the final consonants are there significant differences in favor of the original condition (in fluctuating noise $p<0.05$ for UL-EM and EL-EM; in continuous noise $p<0.05$ for UL-EL and UL-EM).

For all three noise conditions, there is a trend that the subjects can benefit from spectral enhancement only for vowels. For vowels, we found that the scores for the spectrally enhanced condition alone (EL), and combined with the single-channel compression condition (ES) correspond closely. This might imply that single-channel compression could not add much to spectral enhancement. The combination gives no further improvement. On the other hand, when the multichannel compressor follows the spectral enhancement process, the scores get lower relative to enhancement alone. The outcome is again the same as the original, implying that the improvements made by spectral enhancement are lost due to multichannel compression.

For the vowels, several significant differences in scores appear between conditions with continuous and fluctuating noise. Obviously, it is easier to perceive vowels in continuous noise than in fluctuating noise. For the perception of consonants an opposite trend is present in the data, especially for the C2 scores, but statistical significance was reached only for condition EL ($p<0.05$).

3. Individual data on speech perception

In the group data, some effects of spectral enhancement are present, but they are not very impressive. It is possible that the small effects found are due to averaging over the eight subjects. Individual data on speech perception show much larger differences. Subjects 3, 7, and 8 benefit from spectral enhancement in most conditions, but others do not. The extra benefits from single-channel compression could only be found for the easiest conditions c(+4), f(+4), and q(av) in subjects 1, 2, and 6. For multichannel compression (EM), the CVC scores were always worse than in the spectrally enhanced condition alone (EL), except for subject 1.

C. Analysis of the results

We will apply several methods of analysis to get a better picture of the advantages and disadvantages of using spectral enhancement and two types of dynamic compression. The

TABLE VI. Results of a two-dimensional INDSCAL analysis on the C1 scores (a), the V scores (b), and the C2 scores (c). The dimensions correspond to specific acoustic characteristics. For each phoneme, the weightings for each dimension are given as a function of the noise, and processing conditions.

	Noise condition				Processing condition			
	$c(-1)$	$c(+4)$	$f(-1)$	$f(+4)$	UL	EL	EM	ES
(a) C1 perception								
Plosiveness	0.63	0.68	0.73	0.56	0.77	0.62	0.65	0.57
Sibilance	0.49	0.62	0.44	0.49	0.38	0.59	0.62	0.60
(b) V perception								
First formant	0.84	0.78	0.75	0.71	0.79	0.74	0.82	0.79
Second formant	0.37	0.51	0.46	0.62	0.51	0.54	0.45	0.50
(c) C2 perception								
Sibilance	0.90	0.61	0.57	0.86	0.50	0.89	0.64	0.84
Voicing	0.24	0.69	0.69	0.35	0.80	0.32	0.68	0.37

outcomes can yield indications as to why new ways of signal processing are favorable for certain subjects and not for others.

1. INDSCAL analysis

In order to get a qualitative impression of the perceptual mechanisms that cause the errors in speech intelligibility, we investigated the pattern of confusions for the different phonemes: C1, V, and C2. For that reason, the confusion matrices were transformed into symmetrical similarity matrices using the algorithm suggested by Houtgast in an article by Klein *et al.* (1970). Two- and three-dimensional INDSCAL analyses on the similarity matrices of the C1, V, and C2 phoneme confusions were performed. The addition of a third dimension in the analyses resulted, in all situations, in a marginal increase in the variance accounted for. For that reason, we present the results of the two-dimensional analyses. Three separate analyses per phoneme were carried out, each analyzing a separate parameter: subject, noise condition, and processing type. We concentrated on the conditions with background noise. In a separate analysis, it was found that the confusion patterns in the quiet condition are quite different from the confusion patterns in noise. For that reason, we did not combine quiet and noise conditions in the same analyses.

2. C1 perception

The accounted variance of the two dimensions of the C1 object space is relatively equally distributed among both dimensions: 43% and 31%, respectively. The first dimension separates the plosives from the sonorants, the second dimension is mainly determined by the /s/ (sibilance). INDSCAL provides information about the perceptual weighting of these dimensions in different conditions (as shown in Table VI for the noise conditions and for the processing conditions) and for the different subjects (to be used in the correlation analysis, see below). It is important to note that this information is related to the relative perceptual importance rather than to the absolute perceptual importance. In most cases, an increase in the perceptual weighting of dimension 1 will go together with a decrease in the weighting of dimension 2. Table VI(a) shows a high weight of plosiveness for the unprocessed condition and relatively better use of the sibilance information in conditions with spectral enhancement. The subjects that profited most from spectral enhancement (S3, S7, and S8) appear to be the subjects that rely the most on sibilance information.

3. V perception

In the INDSCAL analysis of the vowel confusions, the first dimension corresponds to the frequency of the first formant, the second to the second formant. The variance accounted for in these two dimensions is 60% and 25%, respectively. The weighting factors of the formant information per condition are presented in Table VI(b). The differences between conditions are small, but some trends appear to be present. Higher signal-to-noise ratios yield relatively higher weightings of second-formant information, which can be due to improved audibility. Spectral enhancement also favors the weighting of second-formant information, but the addition of compression counteracts this effect. In the individual data, it

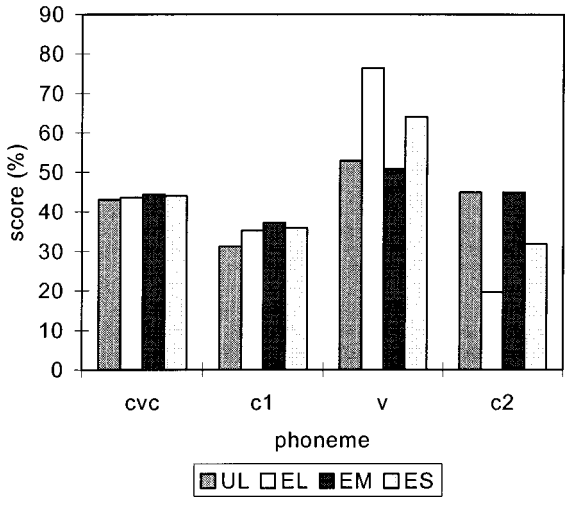


FIG. 5. Scores of the phoneme parts of a CVC word for subject 7 measured in the $c(-1)$ condition: the initial consonant (C1), the vowel (V), and the final consonant (C2). A tradeoff effect between the vowel and the final consonant is clearly discernible for the spectrally enhanced condition alone (EL), and for the combination of spectral enhancement and single-channel compression condition (ES).

TABLE VII. Correlations between psychophysical parameters and intelligibility scores in different noise and processing conditions. Rank correlation coefficients (r_s) greater than 0.8571 (boxed data) indicate a significance value $p<0.01$. r_s-values between 0.7143 and 0.8571 indicate a significance value of $0.01<p<0.05$. The different conditions are abbreviated as follows: **C**ontinuous noise (C), **F**luctuating noise (F), **L**ow S/N ratio (L), **H**igh S/N ratio (H), and **Q**uiet condition (Q). The psychophysical parameters are abbreviated as HL for the hearing loss, MCL for the most comfortable level, DR for the dynamic range, ERB for the equivalent rectangular bandwidth, and PL for the slope of the lower filter flank.

	HLav	HL0.5	HL1	HL2	HL4	MCL0.5	MCL1	MCL2	MCL4	DR0.5	DR1	DR2	DR4	ERB0.5	ERB3	PL0.5	PL3
UL																	
C										0.7686			0.9701			0.9940	
F	−0.7545			−0.8073	−0.8024		0.7381		−0.8743			0.7381	0.9222			0.8503	
L	−0.7545			−0.8073	−0.8024		0.7381		−0.8743			0.7381	0.9222			0.8503	
H	−0.7545			−0.8073	−0.8024		0.7381		−0.8743			0.7381	0.9222			0.8503	
Q	−0.7545			−0.8073	−0.8024		0.7381		−0.8743			0.7381	0.9222			0.8503	
EL																	
C	−0.7545			−0.7711	−0.7785		−0.7857		−0.8982			0.7343	0.8623			0.8264	
F				−0.7576	−0.7229		−0.7665		−0.8193				0.8373			0.7711	
L	−0.8503			−0.7711			−0.7143		−0.8862				0.7306				
H				−0.8675	−0.8024		−0.8095		−0.8383				0.8982			0.8503	
Q									−0.7306					−0.7143			
EM																	
C						−0.7952	−0.7857		−0.7665			0.7143	0.7785	−0.7381		0.8024	
F						−0.7470								−0.7381			
L									−0.8862			0.8095					
H													0.8503			0.8982	
Q						−0.8675	−0.8333						0.7545	−0.7381		0.7785	
ES																	
C									−0.8503			0.8333	0.7904			0.7785	
F				−0.8073	−0.7785		0.7143		−0.8264			0.8810	0.9222			0.8264	
L					−0.7306	−0.7591	−0.8095		−0.8743			0.8333	0.9581			0.9222	
H	−0.7306			−0.7350	−0.7306				−0.7904			0.8333	0.8024			0.7785	
Q	−0.8264			−0.7711	−0.7306		−0.7619		−0.7545			0.7619	0.9341			0.9222	

is striking that subject 6, the only subject with a nonsloping hearing loss, also has high weighting of second-formant information.

4. C2 perception

The analysis of the final consonant confusions shows a distinction of the /s/ in the first dimension. In the second dimension, the voiced and unvoiced consonants are separated (accounted variance is 84% and 11%, respectively). The weighting factors per condition are given in Table VI(c). The weighting factors appear to vary considerably for the different noise conditions in a way we cannot explain. However, for the processing conditions the sibilance dimension becomes even more important in the spectral enhancement conditions, although this effect is partly counteracted by the addition of multichannel compression. It is questionable whether a further increase in sibilance weighting is beneficial for the phoneme scores of the final consonant, while the C2 perception is already dominated by sibilance. The subjects that profited the most from spectral enhancement (S3, S7, and S8) appear to be the subjects that are the least dependent on sibilance information in the final consonant, because they have relatively high weightings for voicing. This may be an explanation for the disappointing effect of spectral enhancement in final consonant perception (Fig. 5 illustrates this effect for S7).

5. Analysis of correlations

We applied the Spearman rank-order correlation method to investigate the relations with psycho-acoustic parameters. These parameters encompass:

(i) Hearing losses for four frequencies (HL0.5, HL1, HL2, and HL4) and the average hearing loss of these frequencies (HLav)

(ii) Most comfortable levels for four frequencies (MCL0.5, MCL1, MCL2, MCL4) and the dynamic ranges for each frequency (DR0.5, DR1, DR2, DR4)

(iii) Slopes of the lower filter skirt for two frequencies (PL0.5, PL3), and the equivalent rectangular bandwidths for two frequencies (ERB0.5, ERB3).

Note that slopes of the upper filter skirt (PH0.5, PH3) are not included, because no strong correlations were found. The same holds for the threshold difference between 0.5 and 4 kHz, which is related to the slope of the hearing loss.

TABLE VIII. Correlations between psychophysical parameters and difference scores in different noise and processing conditions. Rank correlation coefficients (r_s) greater than 0.8571 (boxed data) indicate a significance value $p<0.01$. r_s-values between 0.7143 and 0.8571 indicate a significance value of $0.01<p<0.05$. The different conditions are abbreviated as follows: **C**ontinuous noise (C), **F**luctuating noise (F), **L**ow S/N ratio (L), **H**igh S/N ratio (H), and **Q**uiet condition (Q). The psychophysical parameters are abbreviated HL for the hearing loss, MCL for the most comfortable level, DR for the dynamic range, ERB for the equivalent rectangular bandwidth, and PL for the slope of the lower filter flank.

	HLav	HL0.5	HL1	HL2	HL4	MCL0.5	MCL1	MCL2	MCL4	DR0.5	DR1	DR2	DR4	ERB0.5	ERB3	PL0.5	PL3
Effect of spectral enhancement: EL-UL																	
C										−0.7425			−0.8024			−0.8264	
F	0.9157		0.7530	0.8485			0.7619										
L												−0.7143	−0.7904				
H																	
Q										−0.7186							
Effect of multichannel compression: EM-EL																	
C															0.7665		
F												−0.7381					
L	0.7207																
H		−0.8895				−0.7293					−0.8264						
Q																	
Effect of single-channel compression: ES-EL																	
C																	
F																	0.8333
L																	
H																	0.8333
Q																	
Combined effect of spectral enhancement and multichannel compression: EM-UL																	
C							0.7381			−0.8264			−0.7306			−0.7832	
F	0.8024				0.8982			0.8982			−0.9271						
L	0.8024			0.8675			0.7857		0.8264				−0.7785			−0.7229	
H																	
Q																	
Combined effect of spectral enhancement and single-channel compression: ES-UL																	
C													−0.7586				
F				0.7229	0.7545			0.8469	0.7207		−0.8895						
L							0.7500						−0.8289			−0.7306	
H	0.8503			0.7832	0.9222			0.8469	0.7748		−0.8895						
Q																	

All relations with a two-tailed P value smaller than 5% were considered significant. Significant correlations between intelligibility scores and the psychoacoustic parameters mentioned are presented in Table VII. For each processing condition, we first differentiate between noise types C=continuous, F=fluctuating noise), and after that between S/N ratios [L=low(−1), H=high(+4), Q=quiet].

We were also interested in correlations with the differences for speech intelligibility in different conditions. We discriminate effects of spectral enhancement (EL-UL), effects of additional compression (EM-EL and ES-EL), and combined effects of compression and spectral enhancement (EM-UL and ES-UL). These effects are presented in Table VIII. The extensions C, F, L, H, and Q have the same meaning as in Table VII.

6. Relation with the audiogram

There are only weak negative correlations between the speech scores and the mean audiometric loss. From the analyses per frequency, it can be concluded that the highest correlations are found for the higher frequencies. Detailed analyses on the phoneme level reveal that this can be attributed to the perception of consonants.

The results of the analysis-of-difference scores show that the effects of spectral enhancement for the speech in fluctuating noise are significantly related to the average audiometric loss ($p<0.01$). There is no clear pattern of significant correlations for the additional effect of compression (EM-EL and ES-EL). For the combined effect of spectral enhancement and compression (EM-UL and ES-UL), the significant effects are again present for the fluctuating noise condition ($p<0.05$ for hearing loss at 4 kHz). The signifi-

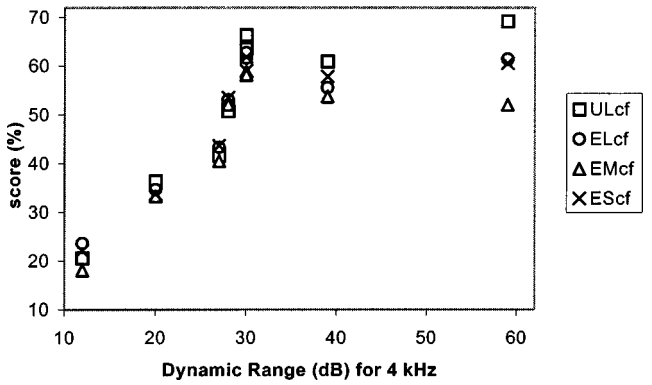


FIG. 6. Speech scores as a function of the dynamic range at 4 kHz (in dB) for four different processing conditions, all averaged on the continuous (C) and fluctuating (F) noise conditions, hence the extension *cf* is used.

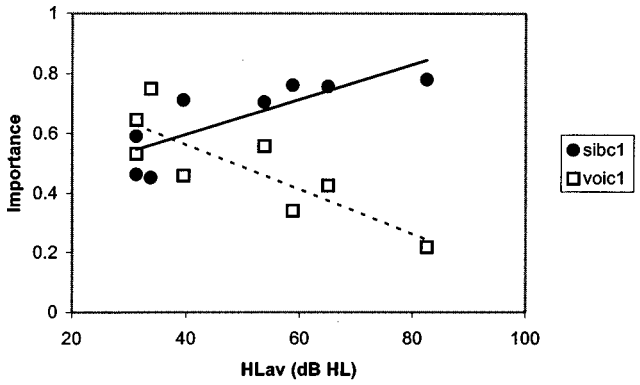


FIG. 7. Importance of two different INDSCAL parameters, sibilance (sibc1) and voicing (voic1), as a function of the average hearing loss (HLav), for the initial consonants. The importance is expressed in a value ranging between zero and 1. The more important the INDSCAL parameter is for a certain subject, the closer the value is to the number 1.

cant correlations appear to stress the importance of the high-frequency cues (sibilance) for large high-frequency losses.

7. Relation with the loudness scaling results

There are only few significant correlations with the MCL values and they show roughly the same pattern as for the pure-tone thresholds: the highest correlations are found at 4 kHz. For the difference scores the correlations are only mildly significant, without a clearly interpretable pattern. There are strong correlations between the speech scores and the dynamic-range parameters at 2 kHz ($p<0.05$) and—even stronger—at 4 kHz (DR4:$p<0.01$). For UL, EL, and ES, there are clear patterns of significant correlations, for the overall scores as well as for the separate phoneme scores (C1, V, and C2), indicating that a high dynamic range is accompanied by high speech scores. Further analysis of the significant relationships shows a characteristic ceiling effect on the scores for dynamic ranges above 30 dB. The ceiling effects appear to be similar for both noise types and for both signal-to-noise ratios. For that reason, we averaged all noise measurements per stimulus condition in order to compare the scores of UL, EL, EM, and ES. In Fig. 6, the scores in the different processing conditions give parallel patterns.

In the analysis-of-difference scores, there are fewer significant correlations for the higher frequencies, but there is a trend for negative correlations in continuous noise conditions and/or at low signal-to-noise ratios. These negative correlations appear to be related to the effects of spectral enhancement (EL-UL) rather than to the additional effects of compression (EM-EL and ES-EL). This implies that the effects of enhancement are smaller for subjects with large dynamic ranges at 4 kHz for the most difficult noise conditions.

8. Relation with the auditory-filter results

The correlations between the speech scores and the auditory-filter parameters (the filter width, ERB, and the steepness of the shallow filter skirt, PL) are only significant at 500 Hz ($p<0.05$). Narrow ERBs and steeper filter skirts go together with higher speech scores. This may be an indication for the involvement of ''upward spread of masking.'' It is striking that the ERB parameters at 500 Hz correlate mainly with the EM scores ($p<0.05$), while there are more significant correlations with the PL parameters for UL, EL, and ES. The ERB parameters appear not to be related to the effects of enhancement and the additional effects of compression, expressed in the difference scores (EL-UL and EM-EL, ES-EL, respectively). The patterns of the speech scores in relation to PL values are more or less the same for all phoneme categories. There is a trend for negative correlations between PL at 500 Hz and the difference scores in continuous noise conditions and/or at low signal-to-noise ratios. Again, the negative correlations appear to be related to the effects of spectral enhancement (EL-UL) rather than to the additional effects of compression (ES-EL and EM-EL). This implies that the effects of enhancement (EL-UL) are smaller for subjects with steep filter slopes at 500 Hz for the most difficult noise conditions. Finally, the PL at 3 kHz appears to be significantly related ($p<0.05$) to the additional effects of compression in the single-channel processing (ES-EL). Improvements in speech scores due to compression are higher for subjects with steep filter slopes at 3 kHz, especially in the relatively easy noise conditions.

9. Relation with the INDSCAL weightings

Correlations between the average hearing loss on the one hand, and the INDSCAL parameters for plosiveness (Spearman's $r_s=-0.7904$) and sibilance ($r_s=0.8743$) of the initial consonant on the other hand, give opposite regression lines. The patterns can be seen in Fig. 7. Obviously, subjects (3, 7, and 8) with higher losses tend to make use of sibilance more than of voicing/plosiveness in the perception of initial consonants. Note that these subjects benefit most from spectral enhancement. The opposite is true for average hearing losses lower than 40 dB.

In the correlations with vowel scores, it appears that the higher the PL value at 3 kHz, the higher the weighing of the second formant.

With respect to the perception of final consonants, correlations between the most comfortable level at 1.0 kHz on the one hand and the INDSCAL parameters for sibilance ($r_s=-0.8095$) and voicing ($r_s=0.8571$), on the other hand, give opposite regression lines. The patterns are shown in Fig.

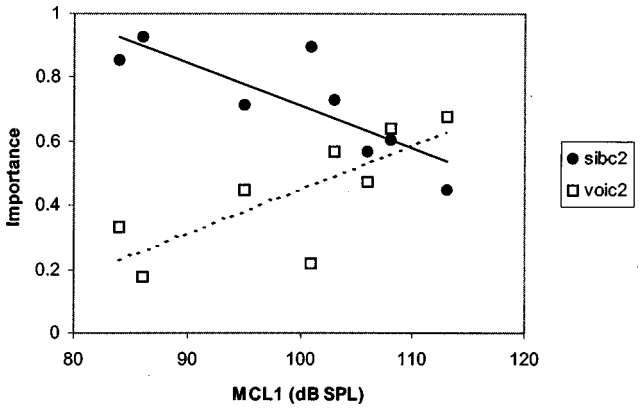


FIG. 8. Importance of two different INDSCAL parameters, sibilance (sibc2) and voicing (voic2), as a function of the most comfortable level at 1000 Hz (MCL1), for the final consonants. The importance value ranges between zero and 1.

8. Obviously, subjects with lower MCL1 levels tend to rely more on sibilance information than on voicing. The opposite is true for MCL1 sound-pressure levels above approximately 110 dB SPL.

III. DISCUSSION

In general, we see that spectral enhancement especially improves vowel perception in particular. It is disappointing that, in spite of better vowel scores, only a few subjects can benefit from spectral enhancement for the CVC words. Obviously, the positive effect on the vowels is counteracted by the negative effect on the consonants. This tradeoff effect, illustrated for subject 7 in Fig. 5, is the reason why almost no spectrally enhanced condition shows any improvement of the overall CVC scores.

The finding that spectral enhancement is particularly beneficial for vowel perception is not a real surprise, because the key to perceive these phonemes is to distinguish the first three formants. Since this is a task that mainly uses information from the spectral domain, enhancement can easily lead to a better vowel recognition. But, apparently, information on consonants is lost for most subjects by using the enhancement. A possible reason can be the low-pass filtering at 4 kHz in the spectral enhancement process that was done offline in Cambridge: parts of bursts of noise for voiceless stops and the random-noise patterns in the higher frequency regions of fricatives get lost. From another point of view, improvement of consonant perception should possibly be sought primarily in the time domain, because temporal cues are important for the discrimination of consonants.

Improvements due to spectral enhancement show significant correlations with the measured dynamic range at 1 kHz; spectral enhancement gives some improvement only for subjects with small dynamic ranges at 1 kHz. Subject 2 is the only exception to the rule, maybe due to a completely different auditory filter at 3 kHz. It is the only filter which has a steep lower skirt and a flat higher skirt.

Considering the single-channel compressor, a beneficial effect was found in the easiest conditions only for subjects S1, S2, and S6—the noise conditions with an S/N=+4 and $q(av)$. In general, the single-channel compressor does not show any negative effects, but does not add intelligibility to the spectrally enhanced signal. We assume that, contrary to multichannel compression, the single-channel compressor does not change the short-term spectra of the speech signals. Given the fact that in an earlier study (Goedegebure *et al.*, 1998) the results of single-channel compression were more promising than the additional effects of phonemic compression this study shows, it may be assumed that the single-channel compressor works best when it is not combined with spectral enhancement.

The absence of beneficial effects due to phonemic compression can possibly be explained by the specific choices made for the attack and release times. We have chosen those compressor parameters that came out best in pilot studies. In combination with spectral enhancement, these choices are not self-evident. It is also possible that stronger effects of phonemic compression can be obtained if the compression takes place *before* spectral enhancement is applied. Hohmann (1989) suggested that specific characteristics in either the frequency or the time domain of the speech signal could not be changed without influencing the characteristics in the other domain. As long as the effects of certain kinds of processing are not well analyzed both in the time and frequency domain, it is hard to predict the results. The question is, to what extent can we compress and/or spectrally enhance the speech signal and retain a minimum of detrimental effects.

IV. CONCLUSIONS

Spectral enhancement improves vowel perception for most subjects, but gives no overall benefit in our subjects and in our experimental conditions. In the individual results, we found that for three subjects the original condition is best, that for three subjects spectral enhancement is best, and that for five subjects there are clear improvements for at least one of the phonemes. Unfortunately, there is a tradeoff effect between increased vowel scores and decreased consonant scores.

For the combined effect of spectral enhancement and single-channel compression, the scores are like the spectral enhancement scores, indicating that the single-channel compressor gives no improvement on top of spectral enhancement. In the multichannel condition the scores are lowest, indicating that multichannel compression and spectral enhancement have opposite effects.

Remarkably, consonant scores are higher in fluctuating noise than in continuous noise, while vowel scores are lower in fluctuating noise. The dynamic range at 4 kHz is the best indicator to predict the speech scores. For initial as well as final consonants, clear relations were found between the INDSCAL parameters for plosiveness/voicing and sibilance on the one hand, and average hearing loss and most comfortable loudness level at 1 kHz on the other. The relations found underline the fact that subjects with higher hearing losses make use of different strategies to understand speech.

ACKNOWLEDGMENTS

The investigation described is part of the European HEARDIP project. The coordination organization TIDE, the

''Technological Innovative Program for the Disabled and Elderly,'' has sponsored the project. We would like to thank Simon Wiersma for his technical support, and André Goedegebure who took considerable trouble in the implementation of the phonemic compressors. We would also like to thank Thomas Baer, Michael Stone, and Brian Moore for processing the speech material in Cambridge, using the spectral enhancement algorithm described in the text.

Baer, T., Moore, B. C. J., and Gatehouse, S. (**1993**). ''Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairment: effects on the intelligibility, quality, and response times,'' J. Rehabil. Res. Dev. **30**, 49–72.

Goedegebure, A., Hulshof, M., Maas, A. J. J., and Verschuure, J. (**1999**). ''Improved speech understanding in hearing-impaired listeners by smoothed fast compression,'' Audiology (submitted).

Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hearing Res. **47**, 103–138.

Harten-De Bruijn, H. E. van, Kreveld-Bos, C. S. G. M. van, Dreschler, W. A., and Verschuure, J. (**1997**). ''Design of two syllabic non-linear multichannel signal processors and the results of speech tests in noise,'' Ear Hear. **18**, 26–33.

Hellbrück, J., and Moser, L. M. (**1985**). ''Hörgeräte Audiometrie: Ein computer-unterstütztes psychologisches Verfahren zur Hörgeräteanpassung,'' Psychol. Beiträge **27**, 494–509.

Hohmann, V. (**1989**). ''Störungen der Lautstärkewahrnehmung bei Innenohrschwerhörigkeit und ihre Kompensation durch digitale Hörgeräte,'' Candidate diploma for the doctorate, University of Göttingen.

Klein, W., Plomp, R., and Pols, L. C. W. (**1970**). ''Vowel spectra, vowel spaces and vowel identification,'' J. Acoust. Soc. Am. **48**, 999–1009.

Leeuw, A. R., and Dreschler, W. A. (**1994**). ''Frequency-resolution measurements with notched noises for clinical purposes,'' Ear Hear. **15**, 240–255.

Moore, B. C. J., and Glasberg, B. R. (**1983**). ''Suggested formulae for calculating auditory-filter bandwidths and excitation patterns,'' J. Acoust. Soc. Am. **74**, 750–753.

Steeneken, H. J. M., Geurtzen, F. W. M., and Agterhuis, E. (**1990**). ''Speech data-base for intelligibility and speech quality measurements,'' IZF-report A13, TNO Institute for Perception, Soesterberg, The Netherlands.

Verschuure, J., Dreschler, W. A., Maré, M. J., Boermans, P. P., and Capellen, M. van (**1993**). ''Compression and expansion in hearing aids especially in conditions with background noise,'' IOP Report 889cd14.

Verschuure, J., Prinsen, T. T., and Dreschler, W. A. (**1994**). ''The effects of syllabic compression and frequency shaping on speech intelligibility in hearing impaired people,'' Ear Hear. **15**, 31–21.

Verschuure, J., and Dreschler, W. A. (**1996**). ''Dynamic compression hearing aids,'' in *Psychoacoustics, Speech and Hearing Aids*, edited by B. Kollmeier (World Scientific, Singapore), pp. 153–164.

Auditory localization of nearby sources. Head-related transfer functions

Douglas S. Brungart[a] and William M. Rabinowitz[b]
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Masssachusetts 02139

(Received 29 June 1998; revised 29 March 1999; accepted 26 May 1999)

Although researchers have long recognized the unique properties of the head-related transfer function (HRTF) for nearby sources (within 1 m of the listener's head), virtually all of the HRTF measurements described in the literature have focused on source locations 1 m or farther from the listener. In this study, HRTFs for sources at distances from 0.12 to 1 m were calculated using a rigid-sphere model of the head and measured using a Knowles Electronic Manikin for Acoustic Research (KEMAR) and an acoustic point source. Both the calculations and the measurements indicate that the interaural level difference (ILD) increases substantially for lateral sources as distance decreases below 1 m, even at low frequencies where the ILD is small for distant sources. In contrast, the interaural time delay (ITD) is roughly independent of distance even when the source is close. The KEMAR measurements indicate that the direction of the source relative to the outer ear plays an important role in determining the high-frequency response of the HRTF in the horizontal plane. However, the elevation-dependent characteristics of the HRTFs are not strongly dependent on distance, and the contribution of the pinna to the HRTF is independent of distance beyond a few centimeters from the ear. Overall, the results suggest that binaural cues play an important role in auditory distance perception for nearby sources. © *1999 Acoustical Society of America.*
[S0001-4966(99)04209-5]

PACS numbers: 43.66.Qp [DWG]

INTRODUCTION

Over the past 100 years, considerable efforts have been made to characterize the relationship between the location of a sound source in space and the sound pressure generated by that source at the eardrums of a human listener. Traditionally, this relationship has been represented by the head-related transfer function, or HRTF. Many researchers have measured the HRTF for relatively distant sound sources, located 1 m or more from the listener. (Møller *et al.*, 1995; Gardner and Martin, 1995; Pralong and Carlile, 1994; Carlile and Pralong, 1994; Middlebrooks *et al.*, 1989; Wightman and Kistler, 1989; Mehrgardt and Mellert, 1977; Shaw, 1974). However, despite the recognition by early researchers that the HRTF varies substantially with distance at distances less than 1 m, very little is known about the characteristics of the HRTF for nearby sources.

Nearly all previous work examining the HRTF for close sources has been based on a rigid-sphere model of the head. Stewart (1911a, 1911b) derived mathematical expressions for the interaural level difference (ILD) and interaural time delay (ITD) on a rigid-sphere head with ears located at diametrically opposite points on its surface, and found that the ILD increases substantially as a nearby source approaches the head, while the ITD remains roughly independent of distance even for close sources. Hartley and Frey (1921) expanded Stewart's calculations, and noted that a combination of distance-invariant ITDs and distance-varying ILDs could provide binaural distance cues for nearby sources. More recently, interaural time and intensity differences have been calculated with an updated sphere model (Brungart and Rabinowitz, 1996; Brungart *et al.*, 1996), and validated with acoustic measurements of the HRTF on a bowling-ball head (Duda and Martens, 1998). A common feature of all of the sphere model predictions is that the HRTF changes substantially with distance at distances less than 1 m, but is virtually independent of distance for sources beyond 1 m. Therefore, it is useful to introduce terminology to differentiate between these near and far regions. In our discussion of auditory localization near the head, we use the term ''proximal region'' to describe the region within 1 m of the listener's head, and the term ''distal region'' to describe the region beyond 1 m.[1]

The rigid-sphere model of the head can provide useful insights into the general behavior of the HRTF for close sources, but it does not fully represent the complexities of the actual HRTF. In order to obtain a clearer picture of the distance dependence of the HRTF, an approximation that accounts for the irregular shape of the head and outer ear as well as diffraction by the neck and torso is needed. Measurements on an anthropomorphic manikin can provide this type of approximation, without the difficulties involved in measuring the HRTF on human subjects. With the exception of some ITD and ILD measurements on a wax dummy head in an early study by Firestone (1930), no comprehensive measurements of the HRTF on an anthropomorphic head (human or manikin) are available in the literature. This paper describes an extensive set of HRTF measurements made with a KEMAR manikin at distances ranging from 0.12 to 1 m. The results are compared to the predictions of the sphere model,

[a]Currently at Human Effectiveness Directorate, Air Force Research Laboratory, WPAFB, OH. Electronic mail: douglas.brungart@he.wpafb.af.mil
[b]Currently at Bose Corporation, Framingham, MA.

and are discussed in relation to their implications for the localization of nearby sources.

I. METHODS

A. HRTF calculations for a rigid-sphere model of the head

Useful approximations of the HRTF for a nearby source, and the corresponding interaural time and intensity differences, can be obtained from mathematical descriptions of the acoustic properties of a rigid sphere. Our model was adapted from the work of Rabinowitz and colleagues (1993), who examined the frequency scalability of the HRTF for an enlarged head. The present variation of the model maintains a fixed head size and varies distance, rather than varying head size at a fixed distance (Brungart and Rabinowitz, 1996). A more detailed review of the application of the model, as well as a comparison of the model's predictions to acoustic measurements made on the surface of a bowling ball, are available in Duda and Martens (1998).

The model approximates the head as a rigid sphere, of radius a, with ''ears'' located at diametrically opposed points on the surface of the head. The sound source is a point velocity source radiating spherical acoustic waves, and is located at distance r from the center of the head and at angle α from the interaural axis. The expression for the complex sound pressure at the ear, denoted by $\mathbf{P}_s$, is given by

$$\mathbf{P}_s(r,a,\alpha,f) = \frac{c\rho_0 u_0}{2\pi a^2} \sum_{m=0}^{\infty} \left(m+\frac{1}{2}\right) \mathrm{L}_m(\cos\alpha) \times \frac{\mathrm{H}_m(2\pi fr/c)}{\mathrm{H}'_m(2\pi fa/c)} e^{-j\pi/2}, \quad (1)$$

where f is the sound frequency, c is the speed of sound, ρ_0 is the density of air, u_0 is the volume velocity of the (infinitesimal) source, L_m is the Legendre polynomial function, and H_m is the spherical Hankel function. To obtain the HRTFs from the model, the pressure at the ear must be divided by the reference free-field pressure at the center of the head $\mathbf{P}_{ff}$, which is simply the pressure of a point source of strength u_0 at distance r:

$$\mathbf{P}_{ff}(r,f) = \frac{2\rho_0 f u_0}{r} e^{j(-2\pi fr/c + \pi/2)}. \quad (2)$$

A remaining complication is the transformation of the angle α between the location of the source and the location of the ear to the more standard spherical coordinates used to define HRTFs. Throughout this paper, the coordinate system has its origin at the midpoint of the interaural axis. Azimuth (θ) will be defined as 0 degrees directly in front of the head, 90 degrees directly to the left, and -90 degrees directly to the right. Elevation (ϕ) will be 0 degrees in the horizontal plane, 90 degrees directly above, and -90 degrees directly below. In this coordinate system, $\alpha = \arccos(\sin(\theta)\cos(\phi))$ and the transfer function at the left ear is

$$\mathbf{H}_L(r,a,\theta,\phi,f) = \frac{\mathbf{P}_s(r,a,\arccos(\sin(\theta)\cos(\phi)),f)}{\mathbf{P}_{ff}(r,f)}. \quad (3)$$

B. Procedures for HRTF measurements

1. Facilities

The HRTF measurements were made in a corner of the anechoic chamber of the Armstrong Laboratory at Wright-Patterson Air Force Base.[2] All of the HRTF measurements were made with a KEMAR manikin, which was fitted bilaterally with Zwislocki couplers and 1.2-cm pressure microphones (Bruel & Kjaer 4134) at the eardrum locations. The microphone outputs were connected to a Bruel & Kjaer 5935 dual microphone power supply, and then passed through a patch panel into the control room. The KEMAR was mounted on a metal stand equipped with optically encoded stepper motors that allow programmable control of the manikin's azimuth and elevation within a fraction of a degree.

In order to avoid the complications involved in near-field measurements with a conventional cone loudspeaker (due to their size, directionality, and complex near-field characterstics), a special sound source approximating an acoustic point source was developed for these measurements. An electrodynamic horn driver (Electro-Voice DH1506) was connected to a 3-m-long section of Tygon tubing with internal diameter of 1.2 cm and walls 1.5 mm thick. For convenience, the end of this tube was routed through a PVC pipe sleeve, 2.5 cm in diameter and 64 cm in length, with the end of the Tygon tube projecting 2 cm from the end of the pipe and with foam material sealing the space between the tube and the interior of the sleeve. The sleeve was used to clamp the point source to a tripod stand that was used to position the source during the measurements. The relatively small diameter of the tube (1.2 cm) allows its opening to act approximately as an acoustic point source in that it is relatively nondirectional even at high frequencies. At 15 kHz, for example, the 3-dB beamwidth of the source was found to be 120 degrees. Furthermore, the small size allows a precise placement of the source and eliminates the potential problem of secondary reflections off the source. The source's frequency response is irregular (Fig. 1); however, its shape is consistent and easily removed from the transfer function measurements, and its effective frequency range is from 200 Hz to 15 kHz.

Reference measurements were made with a 1.2-cm (0.5 in.) free-field microphone (Bruel & Kjaer 4165) located at the position of the center of the manikin head with the manikin removed. Measurements were made at 0.125, 0.25, 0.50, and 1.00 m before and after the HRTF measurements. Changes in the frequency response of the source over the course of the measurements were found to be negligible (within ± 1.5 dB), and the signal measured at the reference microphone was essentially independent of distance except for the inverse relation of overall amplitude to distance.

The measurements were controlled by a computer located in a small room adjacent to the chamber. The two rooms were connected by a patch panel which passed the left and right microphone signals, the sound source signal, and the motor controller signal. The computer was also connected via GPIB bus to a dynamic signal analyzer (Hewlett Packard HP35665A) which was used both to generate the source signal and to measure the transfer functions. The mi-

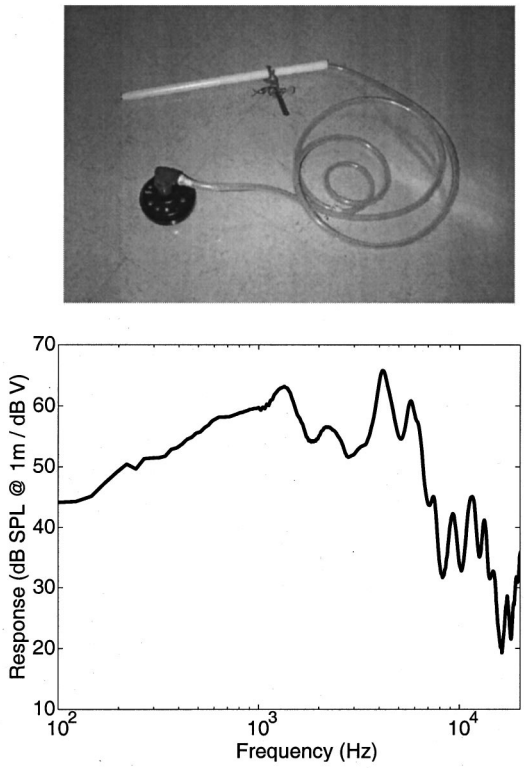


FIG. 1. Acoustic point source. The top panel illustrates the construction of the point source, which essentially consists of an electrodynamic horn driver connected to a 3-m-long flexible tube with a small amount of acoustic foam inserted into the opening to act as a terminating impedance and cut down on standing waves in the tube. The end of the tube is encased in a rigid PVC sleeve to allow easy placement of the tube opening. The sleeve in this picture is curved to allow easy manual placement of the source, and equipped with an electromagnetic sensor to record the location of the tip. In the HRTF measurements described in this paper, only the 64-cm section of the sleeve was used. The bottom panel shows the frequency response of the point source measured with a free-field microphone at a distance of 1 m. The measurement was made in an anechoic chamber using a periodic chirp stimulus and a 1024-point FFT. One-third octave smoothing has been applied to the response. Note the drop-off in the response at frequencies above 6 kHz.

crophone signals were connected to the analyzer inputs. The source signal was amplified (Crown D-75) before being passed through the patch panel to the sound driver.

2. Measurement procedure

The analyzer was operated in transfer-function mode, which measures the ratio of the signal spectra at its two inputs. The analyzer's source output was connected to one input, and one of the ear microphones was connected to the second input. A periodic chirp source signal was used to maximize the signal-to-noise ratio. The periodic chirp is a periodic sine-wave sweep with its period equal to the length of the time record of the fast Fourier transform (FFT). Because of its periodicity, no windowing is required for FFT calculations with the chirp stimulus. At each source position, 64 FFT measurements were combined using root-mean-square (rms) averaging.

All measurements were made at two frequency ranges: from 100 Hz to 12.9 kHz and from 7.8 to 20.6 kHz. Each measurement consisted of a 400-point FFT with 32-Hz resolution. The first 240 points of the low-frequency FFT (from 100 to 7748 Hz) and the first 360 points of the high-frequency FFT (from 7780 to 19 268 Hz) were concatenated into a single 600-point transfer function with 32-Hz resolution from 100 Hz to 19.2 kHz. The two-part measurement was used for two reasons. It allowed higher resolution over the frequency range than a single measurement and, more importantly, it allowed independent ranging of the analyzer's input gain at low and high frequencies. The transfer function of the point source drops off considerably above 7 kHz (see Fig. 1). By dividing the measurement into two parts, the analyzer could be set to optimize the signal-to-noise ratio in both frequency bands.

After each measurement, the amplitude and phase of the transfer function at each frequency value were saved into separate files. The magnitude of the HRTF at each location was found by dividing the amplitude spectrum at that location by the free-field reference measurement corresponding to the source distance. The ITDs were calculated from the difference between the phase responses at the left and right ears.

3. Calibration

Prior to each set of measurements, the KEMAR manikin was adjusted to place the center of the interaural axis directly over the axis of rotation of the stand. When the source is close to the head, correct placement is particularly important because small deviations between the center of the head and the center of rotation will cause changes in the distance to the source as the manikin is rotated in azimuth and severely corrupt the measurements. The centering of KEMAR was accomplished automatically via a series of acoustic measurements with the source placed in front of the manikin at 0-degree elevation. The manikin was first centered in azimuth by rotating the head until the magnitude of the ITD was reduced below 2 μs. Then the roll of the manikin was verified by rotating KEMAR 180 degrees in azimuth and again verifying that the ITD remained near zero. With a nearby source, any left or right tilt of the manikin will prevent the source from falling on the median plane at both 0 and 180 degrees in azimuth. Finally, the manikin head was centered in elevation by verifying that the time delay from the source to the left ear at 0 degrees in azimuth was equivalent (within 5 μs) to the delay from the source to the left ear at 180 degrees in azimuth. If the manikin were tilted forward or backward, the distances (and delays) for these two measurements would differ. After the elevation was adjusted, the centering measurements were repeated until the manikin was acceptably centered in azimuth, roll, and elevation.

II. THE HRTF FOR PROXIMAL-REGION SOURCES IN THE HORIZONTAL PLANE

A. Sphere model predictions

The magnitudes of the HRTFs predicted by the sphere-based mathematical model of the head are shown in Fig. 2. The principal characteristics of these transfer functions can be summarized by four observations:

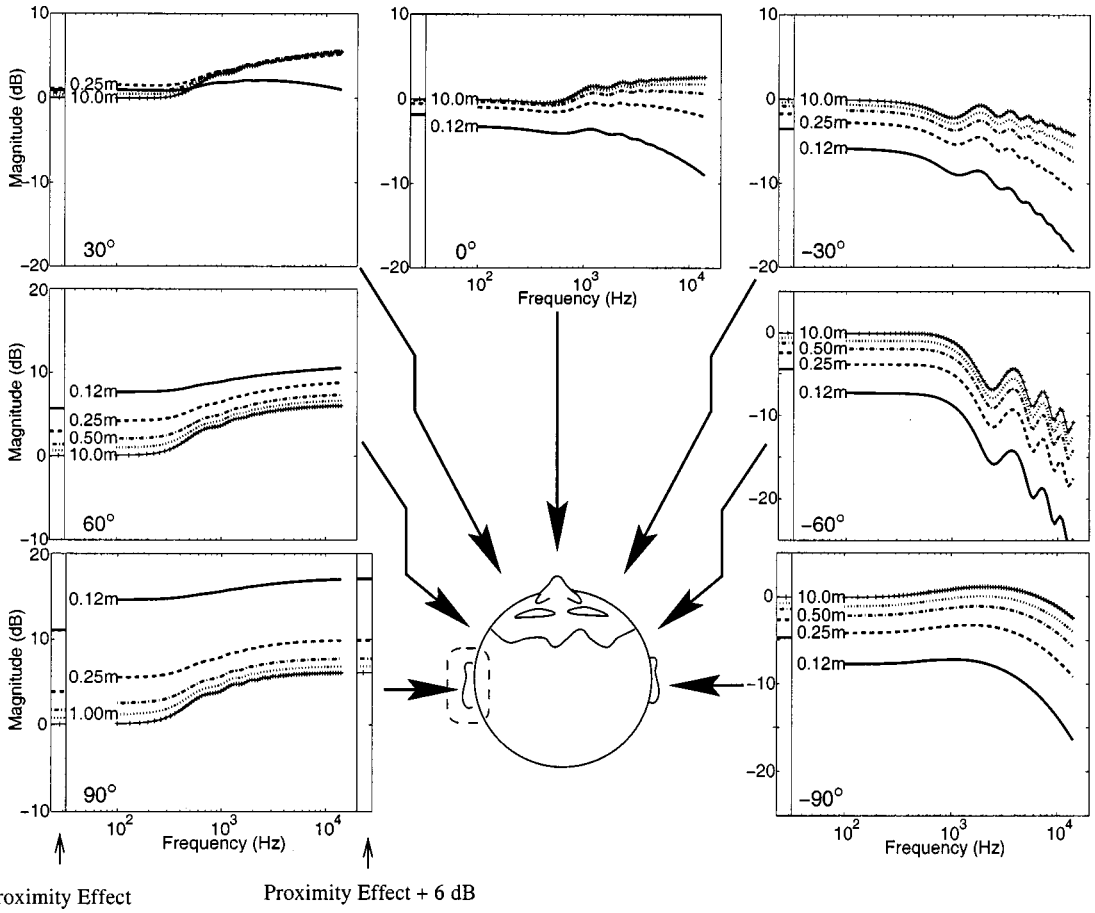


FIG. 2. The HRTF for horizontal-plane sources from 0.125 to 10 m when the head is modeled as a rigid sphere 18 cm in diameter. The HRTFs were calculated by dividing the pressure at the left ear by the free-field pressure at the center of the head [Eq. (3)]. Results are shown for source locations at 30-degree intervals in azimuth in the front hemisphere only, as the sphere model is symmetric across the frontal plane. Note that each panel represents a 30-dB range of responses, but that the values change to accommodate the data. The bars at the left side of each graph show the source proximity effect, which is the gain of the HRTF ignoring diffraction by the head. In the 90-degree panel, the bar at the right side of the panel shows the source proximity effect plus the 6-dB high-frequency pressure doubling effect, and illustrates that the combination of these two effects fully explains the high-frequency asymptotes of the HRTFs at this location. See text for details.

(1) The HRTF increases as the source rotates toward the ear, and decreases as the source rotates away from the ear.
(2) The HRTF increases with frequency when there is a direct path between the sound source and the ear, and decreases with frequency when the ear lies in the acoustic shadow of the head.
(3) The HRTF decreases with distance when there is a direct path from the source to the ear, and increases with distance when the ear is shadowed by the head.
(4) The HRTF changes rapidly as distance decreases below 0.5 m, but slowly (<1 dB) as distance increases from 1 to 10 m.

1. Analysis of sphere model results

Several of the characteristics of the sphere-model HRTFs can be explained intuitively with relatively simple acoustic concepts. There are also some low-frequency diffraction effects in the HRTFs which defy intuitive explanation. These are explained in more detail below.

a. Head shadowing. Head (or sphere) shadowing is the attenuation in the HRTF that occurs when the head obscures the direct path from the sound source to the ear. The magnitude of the shadow is related to the size of the head relative to the wavelength of the sound, so the attenuation at the shadowed ear increases with frequency. As a result, the HRTFs for contralateral sources resemble low-pass filters (see the HRTFs for -30, -60, and -90 degrees in Fig. 2).

As a result of the convex shape of the spherical head, the shadowed region of the sphere's surface will always extend into the hemisphere closest to the source when the source distance is finite. A crude illustration of this phenomenon is provided in Fig. 3, which shows a sound source located 20 cm from the center of a spherical head 20 cm in diameter. Note that the direct path from the source to the surface of the sphere is blocked at all points to the right of point T, where the tangent line ST intersects with the surface of the sphere. Thus the direct path SA from the source to point A on the sphere is blocked by the surface of the sphere, and point A is shadowed by the head. In this example, the shadowed zone extends over two-thirds of the surface of the sphere (the entire region where the arc from point C closest to the source is greater than 60 degrees). The size of this shadowed zone increases as the source moves closer to the head, and decreases as the source moves farther away.

As a consequence of this type of shadowing, it is possible for both the ipsilateral and contralateral ears to fall

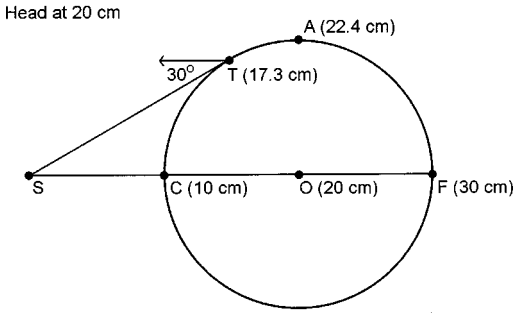


FIG. 3. Head shadowing and the source-proximity effect for a nearby source. In this figure, sound propagates from an infinitesimal point source (S) to the surface of a spherical head with a diameter of 20 cm, centered at 0. Note that the head shadow, which is defined as the region where the direct path from the source to the ear is occluded by the head, includes the entire contralateral hemisphere and extends into the ipsilateral hemisphere where the direct path from the source is tangent to the surface of the sphere (point T). Thus point A is in the head-shadowed region in this example. Also, note that point A is more than twice as far away from the source as the closest point (C) and that (F) is three times as far away as the closest point (C). Thus, even if the effects of diffraction by the sphere are ignored, there would be a substantial difference in sound level between points C, A, and F due to source proximity effects.

inside the shadowed zone of the sphere. In the case of the spherical head used in the model (with a radius of 9 cm), the ipsilateral ear lies in the head shadow for a sound source at 30 degrees at all distances less than 18 cm. Thus, at 30 degrees in Fig. 2, the ear lies in the acoustic shadow of the head at a source distance of 0.12 m, and is unshadowed at all greater source distances. As a result, the high-frequency component of the sound reaching the ear is substantially attenuated when the source is at 12 cm, and the magnitude of the 12-cm transfer function switches from greatest to smallest at this point. The amount of attenuation due to the head shadow also increases as the ear moves deeper inside the shadowed region. Since the size of the shadowed region increases as source distance decreases, this results in increased high-frequency attenuation for nearby sources at the shadowed ear. At all points to the right of 0 degrees, the source is shadowed at all distances, and the ordering of the transfer functions is reversed (i.e., the magnitude of the transfer function decreases with distance in the unshadowed zone (60 and 90 degrees), and increases with distance in the shadowed zone (0, -30, and -60 degrees).

b. Source proximity. The inverse relationship between pressure and distance for a spherically radiating sound wave can substantially influence the HRTF when the source is near the head. This *source-proximity* effect can be viewed as the portion of the HRTF which is not a result of diffraction by the head, i.e., as the HRTF for a pressure sensor suspended in free space at the location of the ear. Since the HRTF is defined as the ratio of the pressure at the ear to the pressure at the center of the head, the source-proximity effect is simply the ratio of the distance from the source to the center of the head to the distance from the source to the ear. An example of the source-proximity effect is provided in Fig. 3. In this figure, which assumes that a head with diameter 20 cm is located 20 cm from the source, the intensity of the sound reaching its surface would be 4.8 dB stronger at the point closest to the source (C in the top panel of Fig. 3) than at the point where the direct path from the source is tangent to the sphere (T). If the direct sound could pass through the head unimpeded, the sound reaching the point farthest from the source (F) would be attenuated 9.5 dB relative to the sound reaching the point closest to the source.

The magnitude of the source-proximity effect is shown along the left side of each panel in Fig. 2. At 10.0 m, the effect is negligible (near 0 dB) for all source directions. The effect increases as source distance decreases, and at 0.12 m the source-proximity effect produces more than 10 dB of gain at 90 degrees and almost 5 dB of attenuation at -90 degrees. The proximity effect is frequency independent. In particular, it helps interpret the large changes in the HRTF at low frequencies that are evident only for nearby sources. However, the source-proximity effect cannot explain all of the low-frequency behavior of the HRTF. In general, the ordering of the low-frequency HRTFs by distance is consistent with the source-proximity effect at each azimuth location, but the magnitude of the low-frequency gain or attenuation is greater than that predicted by source proximity.

c. High-frequency pressure doubling. The magnitude of the HRTF at the ipsilateral ear generally increases with frequency due to high-frequency reflections off the surface of the sphere. When the source is located at 90 degrees, the sound wave impinging on the head is perpendicular to the surface of the sphere at the location of the ear, and at sufficiently high frequencies the sound wave is specularly reflected off the surface of the sphere back in the direction of the source. The direct and reflected sound waves combine in phase to produce a 6-dB pressure gain at the location of the ear. This 6-dB increase in high-frequency gain is evident in the 10 m HRTF at 90 degrees in Fig. 2. In fact, for all of the distances, the high-frequency magnitude of the HRTF at 90 degrees is essentially equivalent to the combination of the source-proximity effect and the high-frequency pressure doubling effect, as shown at the right side of the panel in Fig. 2. As the source rotates away from 90 degrees, the sound waves from the source are no longer perpendicular to the surface of the sphere at the ear and only a portion of the sound wave is reflected at the ear canal opening, resulting in a high-frequency gain slightly less than 6 dB at 60 and 30 degrees. Note that high-frequency pressure doubling does not occur in the contralateral HRTFs.

d. Acoustic bright spot. When the ear is located directly opposite the source (-90 degrees in Fig. 2), all possible sound paths from the source around the sphere to the ear are cylindrically symmetric and, consequently, all components of the diffracted sound wave combine in phase at the ear. This in-phase combination results in a local maximum in the HRTF at all frequencies, a phenomenon known as the acoustic ''bright spot.'' As a result the high-frequency responses of the HRTFs at -90 degrees are substantially greater than the corresponding components of the HRTFs at -60 degrees at each source distance. The effects of constructive and destructive interference on the contralateral hemisphere of the head are seen as ripples of the high-frequency HRTF response.

e. Other diffraction effects. The low-frequency response of the sphere-model HRTFs cannot be explained fully with

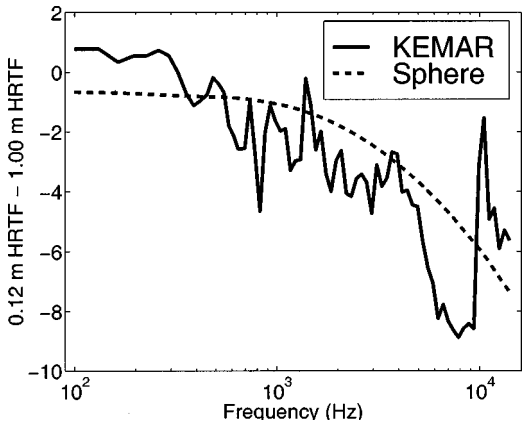


FIG. 4. The difference between the mean HRTFs at 0.12 and 1.0 m. The sphere model and KEMAR HRTFs at 0.12 and 1.0 m were averaged (in dB units) across all locations in the horizontal plane. The results illustrate that the HRTF, on average, decreases in magnitude more quickly at high frequencies than at low frequencies as the source approaches the head. This effect is more pronounced in the KEMAR measurements than in the sphere model. The decrease in spectral content at high frequencies as distance decreases could potentially serve as a spectral distance cue in the proximal region.

the simple concepts described above. As noted before, the source-proximity effect only partially accounts for the low-frequency HRTF responses. The remaining differences result from diffraction by the sphere. Note that the diffraction effects tend to increase the low-frequency gain at the ipsilateral ear and decrease the gain at the contralateral ear. The magnitude of these diffraction effects increases as distance decreases.

2. Low-pass filtering of proximal-region sources

The combination of the low-frequency diffraction effects in the ipsilateral hemisphere and the head-shadowing effects in the contralateral hemisphere produces a consistent change in the shape of the HRTF with source distance. The high-frequency response of the HRTF is generally lower relative to the low-frequency response of the HRTF when the source is close than when the source is more distant. This effect occurs at all source locations, in both the ipsilateral and contralateral hemispheres. The high-frequency response is generally 4–6 dB lower relative to the low-frequency response when the source is at 0.12 m than when the source is at 1 m (Fig. 4). This relationship implies that a sound source at a fixed angle relative to the head will appear to be low-pass filtered as the source approaches the head. Although this effect is modest, it could be used as a monaural distance cue in the proximal region, and it is consistent with previous observations that sound sources appear to ''darken'' in timbre as they approach the head (Begault, 1987; von Bekesy, 1960).

3. Contour plots of sphere-model HRTFs

Contour plots provide a different perspective of the sphere-model HRTFs (Fig. 5, left panels). These plots are particularly useful for viewing the rippled nature of the acoustic bright spot in the contralateral HRTFs. At 2500 Hz, there is a single peak around −90 degrees (location B). As frequency increases, this peak decreases in width and additional peaks form on either side until, at high frequencies, the central peak is very sharp and is surrounded by multiple ridges on either side (location A). The increases in the high-frequency response of the ipsilateral HRTFs due to pressure doubling are also apparent in the contour plots (location C).

B. KEMAR measurements

The HRTFs measured with the KEMAR manikin are shown in Fig. 6. Numerous distal-region HRTFs, both from manikins and from human listeners, are available in the literature, and the directional characteristics of the distal-region HRTFs are well documented. This discussion will focus on a comparison between the HRTFs calculated with the sphere model and those measured with KEMAR, and on the distance-dependent characteristics of the KEMAR HRTFs. It is important to note that these HRTFs measured with the KEMAR manikin may differ from those measured on any individual subject. However, the general pattern of the distant-dependent changes in the HRTFs with the KEMAR manikin should be representative of those in the HRTFs of human listeners. Although it is unlikely that the high-frequency characteristics of the KEMAR HRTFs will match those of any particular human listener, the relative changes in those HRTFs (such as systematic shifts in the azimuth or frequency of notches in the HRTF with distance) provide some insights about the general distant-dependent behavior of the HRTF for nearby sources.

1. Comparison of KEMAR measurements with sphere model

The sphere model best fits the KEMAR measurements at low frequencies (below 1 kHz). For comparison, the magnitude of the sphere-model HRTF at 100 Hz is shown beside the KEMAR measurements in Fig. 6. The fit of the model to the measurements is best at the contralateral ear and worst near the boundary between the shadowed and unshadowed zones (30 and 150 degrees). There are also some discrepancies between the model and measurements at 0.12 m near 90 degrees.

As frequency increases, the KEMAR HRTFs begin to diverge from those of the sphere model. At 2.9 kHz, the quarter-wavelength resonance of the ear canal causes a large peak in the KEMAR HRTFs at all directions and distances. At higher frequencies, the KEMAR transfer functions exhibit a complex series of direction-dependent peaks and notches which arise from the geometry of the pinnae and are not reflected in the sphere model. Five major characteristics of the sphere HRTFs are preserved in the KEMAR HRTFs.

(1) The magnitude of the HRTFs generally increases with frequency when there is a direct path from the source to the ear. In part, this increase is probably a result of the reflections that cause the 6-dB gain in the sphere model. The pinna, which is shaped somewhat like a cone, also provides some gain at high frequencies. Note that the overall high-frequency gain at the ipsilateral ear is greater in the KEMAR measurements than in the sphere model.

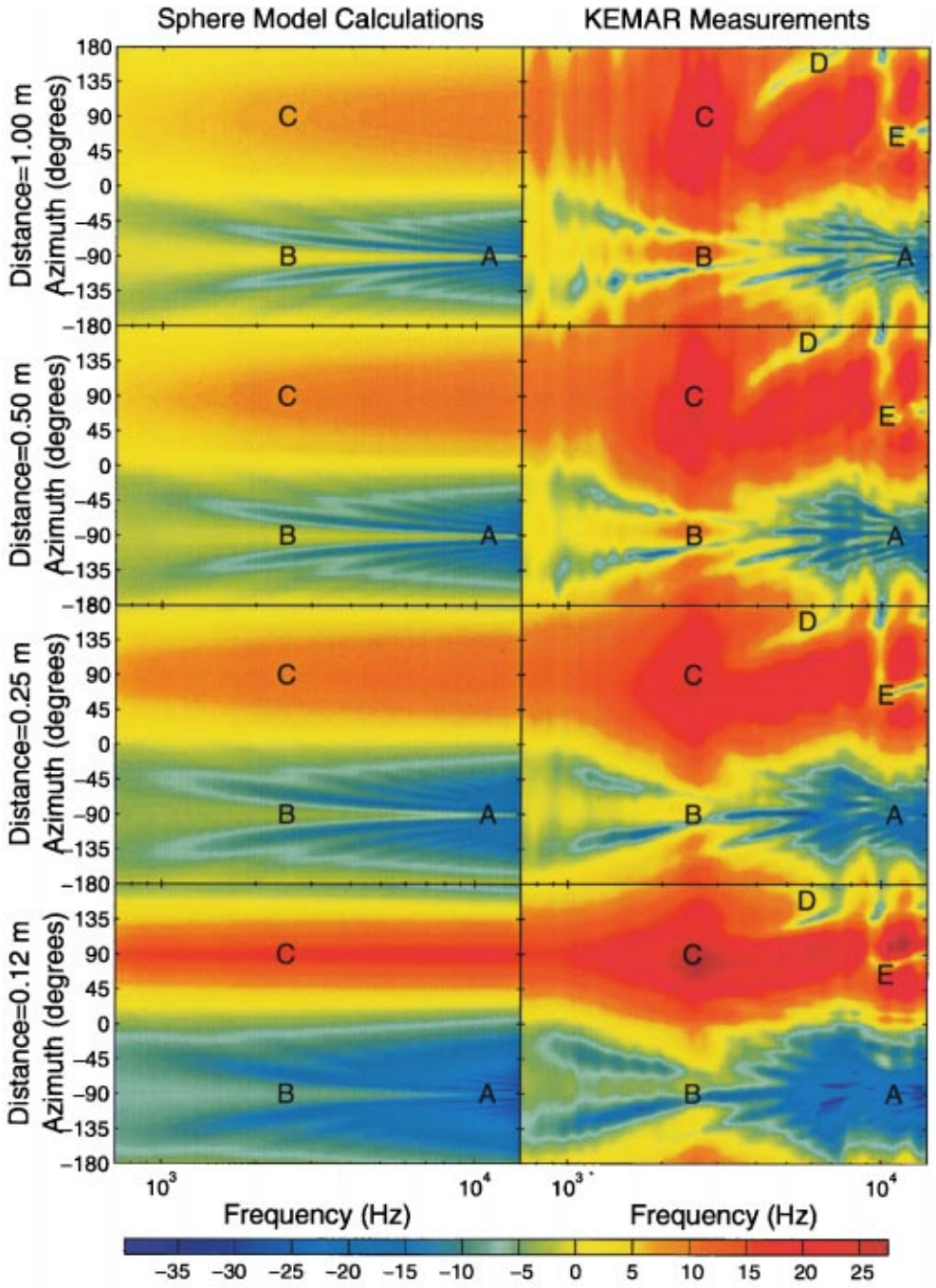


FIG. 5. Contour plots of the HRTFs predicted by the sphere model (left column) and measured with KEMAR (right column). Azimuth is shown at 3-degree intervals, and frequency is shown at 1/12-oct intervals from 500 Hz to 14 kHz. The magnitude of the transfer function at each point (in dB) is represented by color, as shown by the legend across the bottom of the figure. Five reference points (labeled A–E) are present on the contour plots (see text for details).

(2) The high-frequency responses of the HRTFs are attenuated when the ear is in the acoustic shadow of the head.
(3) The overall gain of the HRTFs increases as distance decreases when a direct path exists between the source and the ear, and the overall attenuation of the HRTFs increases as distance decreases when the ear is shadowed by the head. Note that, as in the sphere model, the ear is first shadowed by the head at 30 and 150 degrees when the source is at 0.12 m, and that the ordering of the HRTFs at high frequencies reverses at these locations.
(4) Overall, the magnitude of the HRTF increases more rapidly at low frequencies than at high frequencies as the source approaches the head (Fig. 4). Thus, the sound reaching the eardrums is effectively low-pass filtered as the source approaches the head. This effect is more dramatic in the KEMAR HRTFs than in the sphere model, and may serve as a monaural distance cue in the proximal region.
(5) Although its structure is more complicated, the acoustic bright spot seen in the sphere model is also evident in the KEMAR measurements. This is best seen in the contour plots of the KEMAR measurements (Fig. 5). The peak at intermediate frequencies occurs slightly to the left of −90 degrees in the KEMAR measurements due to the

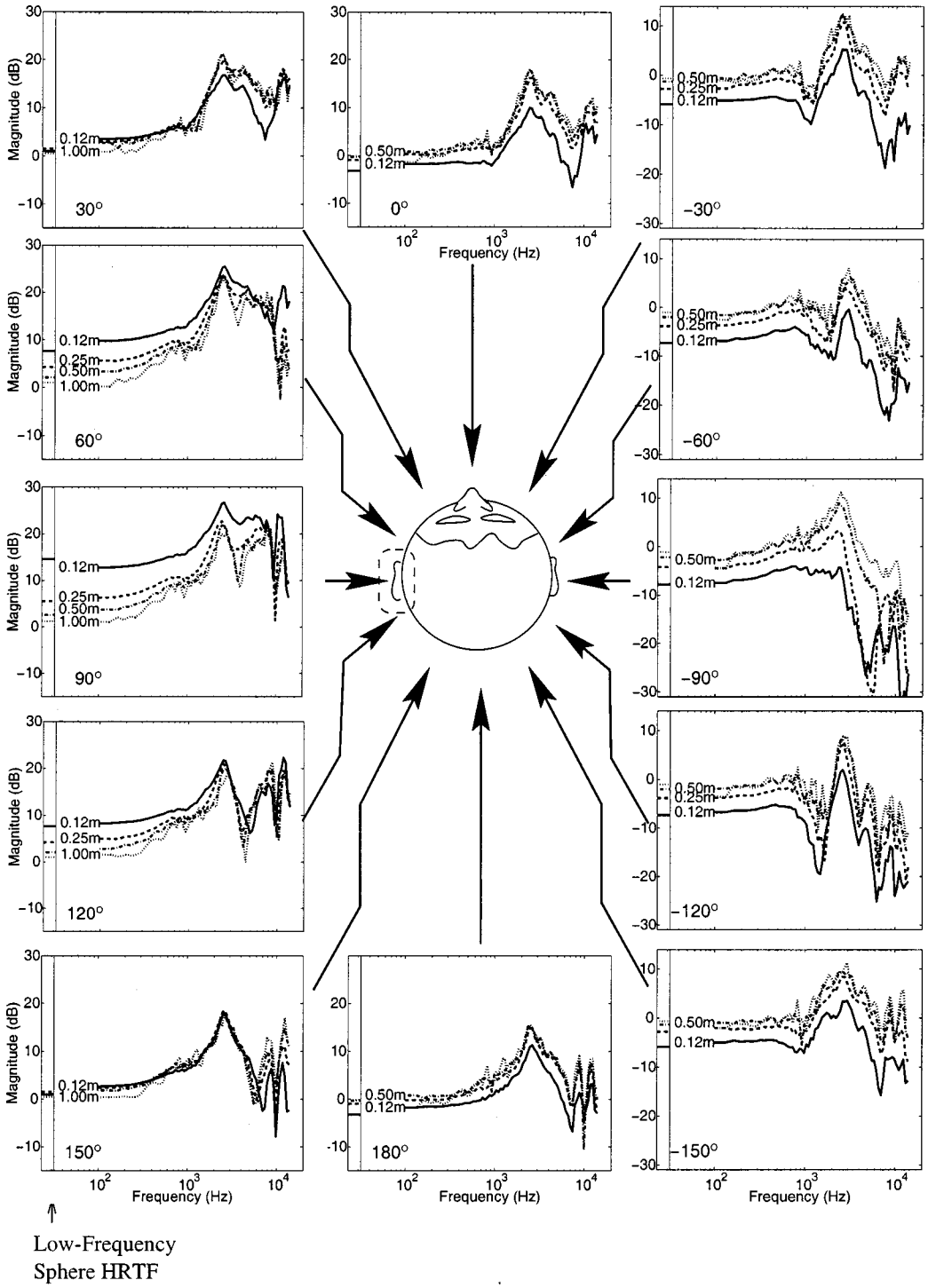


FIG. 6. The HRTF for horizontal-plane sources from 0.125 to 1.0 m measured with the KEMAR manikin. The HRTFs were calculated by dividing the pressure at the left ear by the free-field pressure at the center of the head. The bars at the left side of each panel show the low-frequency asymptote of the HRTF calculated with the sphere model (as shown in Fig. 2).

asymmetries of the manikin head (location B). At higher frequencies, the periodic interference pattern around the bright spot is seen in the KEMAR measurements, but it is more erratic than in the sphere model. Note that one notch from the interference pattern appears to extend into the ipsilateral hemisphere at 6.5 kHz (location D).

2. The parallax effect in the high-frequency HRTF

When a sound source is distant, it is located in roughly the same direction relative to the head and to the left and right ears. However, when the source is close to the head, the angle from the source to the ear can differ substantially from the angle of the source relative to the head (Fig. 7). This discrepancy between the angles of the source relative to the head and ear (or relative to the left and right ears) is known as acoustic parallax (Suzuki *et al.*, 1998). The parallax effect changes systematically with the distance of the sound source; the closer the source comes to the head, the more its angle relative to the ispilateral ear is pushed away from the interaural axis. As a result, the high-frequency features of the

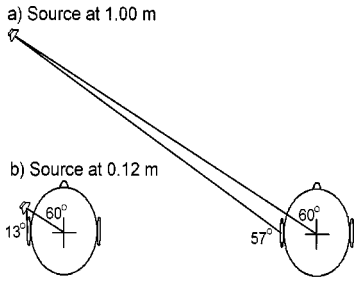


FIG. 7. The parallax effect for nearby sound sources. When a sound source is relatively distant, the direction of the source relative to the ear is approximately the same as the direction of the source relative to the center of the head. For example, with a head width of 18 cm, a source at 60 degrees azimuth in the horizontal plane (relative to the head) is located at 57 degrees relative to the left ear. However, when a sound source is near the head, the angles from the ear to the source and the head to the source can differ substantially. A source 0.12 m from the head at 60 degrees azimuth relative to the head is located at 13 degrees relative to the ear. To the extent that the high-frequency features of the HRTF depend on the direction of the source relative to the pinna, this parallax effect can cause a lateral shift in the HRTF. See text for details.

ipsilateral HRTF associated with the directional properties of the outer ear tend to occur at increasingly lateral azimuth locations as the source approaches the head.

A direct illustration of this effect is provided by the high-frequency HRTFs in Fig. 8. In this figure, two prominent features of the HRTF have been labeled: a notch in the HRTF for sources in the rear hemisphere that increases in frequency from 3.5 kHz to 7 kHz as the source moves behind the head (point D), and a dip in the HRTF for frequencies from 9 to 13 kHz at a narrow range of azimuths around 65 degrees (point E). These points are also marked on the color contour plots in Fig. 5. At 1.00 m, the reference points D and E are centered directly over the HRTF features. As distance decreases, however, both features are shifted laterally (toward 90 degrees) in azimuth. As distance decreases from 1.00 to 0.12 m, the notch at point D is shifted systematically from 160 to 140 degrees, while the dip at point E is shifted from 65 degrees to approximately 80 degrees.

Since the parallax effect is based on the directional frequency response of the pinna, it is important to note that this frequency response varies much less with distance than does the frequency response of the head. This is a direct result of the much smaller dimensions of the pinna—since the pinna is only about 1/5 as large as the head, the ''near field'' of the pinna should be much smaller. The contribution of the pinna to the proximal-region HRTF has been measured by comparing the KEMAR HRTF with the pinna in place to the KEMAR HRTF with the pinna removed (Brungart, 1998). The measurements verify that the frequency response of the ear does not vary substantially with distance until the source is within 4 cm of the ear ($\approx$15 cm from the center of the head).

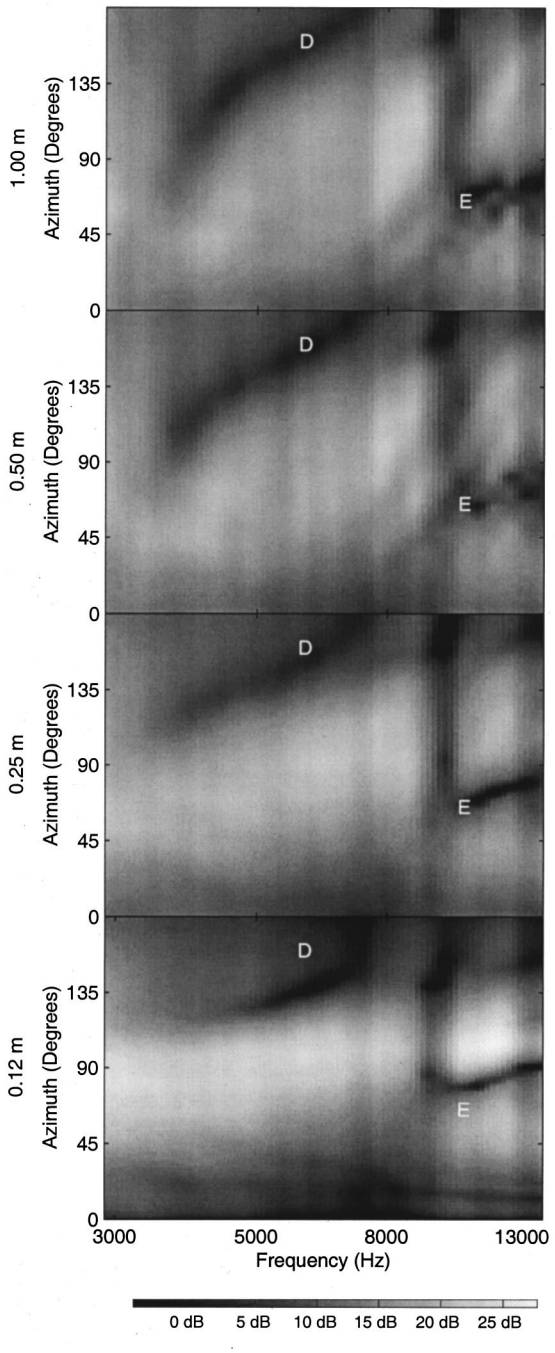


FIG. 8. Illustration of the parallax effect in the high-frequency HRTF. These contour plots show the high-frequency ipsilateral HRTF as a function of frequency and azimuth. The magnitude in (dB) is represented by the gray-scale shading, as indicated in the scale at the bottom of the figure. Two reference points are indicated on each of the panels: D (at 5800 Hz and 160 degrees) and E (at 10 300 Hz and 65 degrees). Note that these points are also indicated in the color panels of Fig. 5. As distance decreases, the features of the HRTF are shifted away from the fixed reference points and toward 90 degrees.

III. INTERAURAL DIFFERENCES FOR PROXIMAL-REGION SOURCES IN THE HORIZONTAL PLANE

A. Interaural level differences

Interaural level differences were calculated at 0.125, 0.25, 0.50, and 1.0 m with the KEMAR HRTFs, and at each of these distances plus 10.0 m with the sphere-model HRTFs. The data are shown in polar form in Fig. 9. The important characteristics of the ILDs can be summarized as follows:

(1) The ILD is always near 0 dB in the median plane and generally increases as the source moves lateral to the head. This result follows directly from the directional

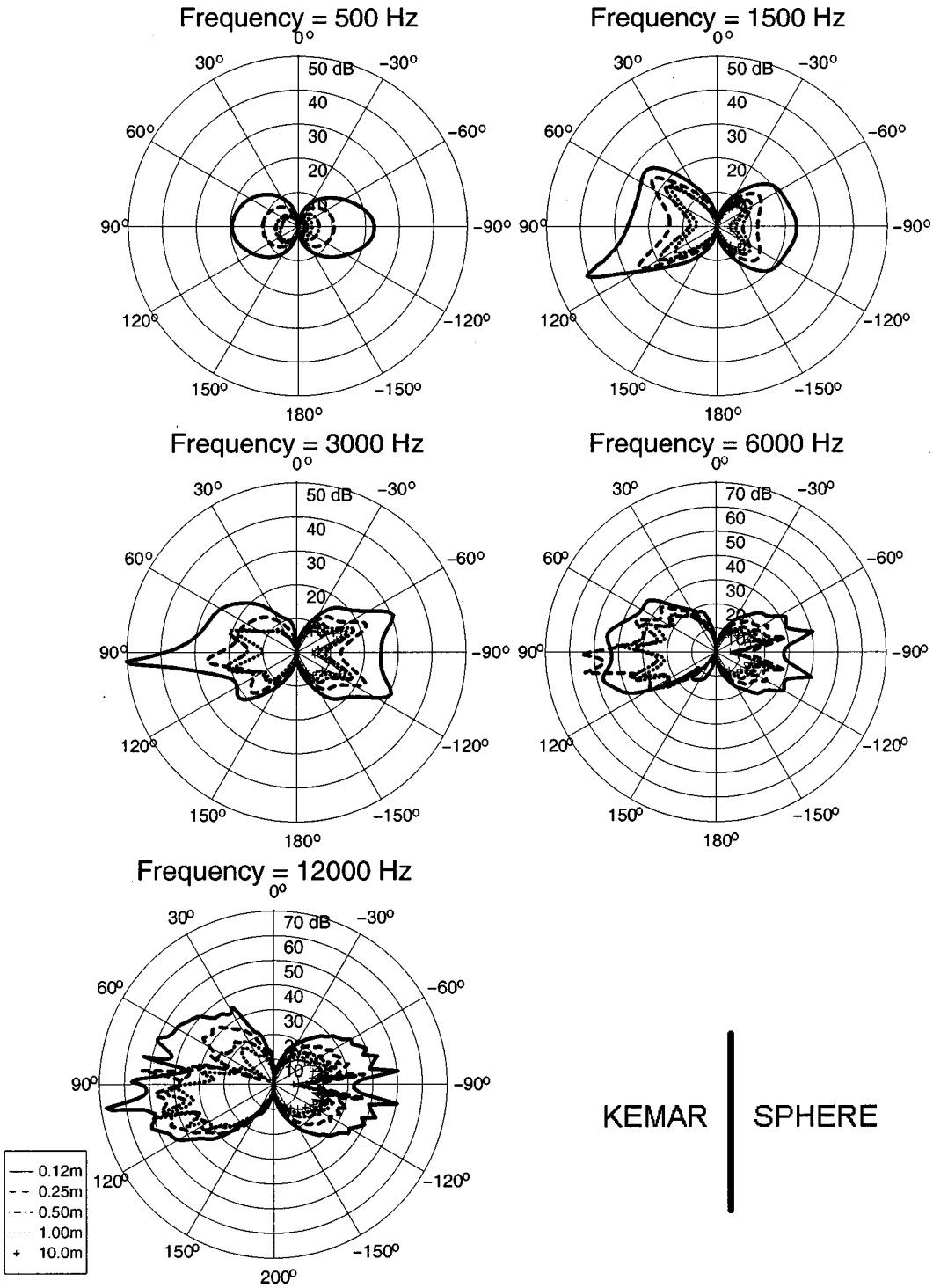


FIG. 9. Interaural intensity differences. Source location in azimuth is represented by angle, and the magnitude of the IID (in dB) is represented by the radius at each angle. Results are provided at five frequencies, ranging from 500 Hz to 12 kHz. The left side of each polar plot shows the IID from the KEMAR measurements, while the right side shows the IID calculated from the sphere model. Note that the radial scale in the plots at 6000 Hz and 12 000 Hz is larger than in the lower-frequency plots.

dependence of the HRTF, which increases in magnitude as the ear rotates toward the source and decreases in magnitude as the ear rotates away from the source.

(2) The ILD generally increases as frequency increases. This behavior results from the tendency of the HRTF to increase with frequency when there is a direct path from the source to the ear and to decrease with frequency when the ear is shadowed by the head. Both effects contribute to the enlarged ILD at high frequencies.

(3) The ILD increases as distance decreases, and dramatically so as the source distance drops below 0.5 m. This distance dependence occurs because the magnitude of the HRTF increases as distance decreases at the ipsilateral ear, and decreases as distance decreases at the contralateral ear. This increase in ILD may be most salient perceptually at low frequencies. For HRTFs measured at distances greater than 1 m, low-frequency ILDs are very small. At 0.12 m, however, they can be on the order of

20 dB and may provide a substantial cue for distance. High-frequency ILDs also increase as distance decreases, but they can also be quite large when the source is distant, and listeners may be less sensitive to a change in ILD from 40 to 60 dB (at high frequencies) than to one from 2 to 20 dB (at low frequencies). Low-frequency ILDs are a unique feature of the proximal-region HRTFs which clearly differentiate them from distal-region HRTFs.

(4) The acoustic bright spot directly opposite the location of the source causes a local minimum in the ILD near ± 90 degrees at intermediate frequencies (1500 and 3000 Hz). In the sphere plots, this minimum also occurs at higher frequencies, but in the KEMAR HRTFs the irregular shape of the head causes the bright spot to break into an erratic series of peaks and nulls which influence the ILD around 90 degrees. At 500 Hz, the bright spot does not influence the ILD.

(5) As was the case with the HRTFs, the sphere model most accurately reflects the behavior of the ILDs measured with KEMAR at low frequencies. The ILD at 500 Hz is a smooth function both for the sphere model and for the KEMAR measurements. At higher frequencies, asymmetries of the KEMAR head cause its ILD to deviate from the sphere model. At these frequencies, the KEMAR ILDs are generally larger than the ILDs predicted by the sphere model. This discrepancy is primarily the result of the directional properties of the pinna, which provides substantial mid- to high-frequency gain at the ipsilateral ear.

B. Interaural time delays

The interaural time delay for a source in the horizontal plane was calculated at 0.125, 0.25, 0.50, 1, and 10 m with the sphere model and measured at 0.12, 0.25, 0.50, and 1 m with the KEMAR manikin (Fig. 10). These time delays were calculated directly from the average slope of the unwrapped phase from 100 to 6500 Hz of the interaural transfer function (the left HRTF divided by the right HRTF). The time delays from the sphere model are necessarily symmetric across both the median and frontal planes. At 10 m, the time delay peaks at approximately 700 μs at 90 degrees. As distance decreases, the magnitude of the time delay increases slightly. At 90 and -90 degrees, where the time delay increase is largest, it changes by about 100 μs as the distance decreases from 10 to 0.125 m. The majority of this magnitude increase occurs as the source moves from 0.25 to 0.125 m and the remainder from 0.50 to 0.25 m; there is virtually no dependence between the time delay and distance beyond 0.50 m.

The KEMAR ITDs from the measurements are generally similar to those from the sphere model, although there is some asymmetry between the front and rear hemispheres. Also, the ITDs with KEMAR exhibit a broader peak when the source is near the interaural axis. As with the time delays predicted by the sphere model, the KEMAR time delays increase by approximately 100 μs as distance decreases from 1 to 0.12 m for a lateral source.

There is, however, one important difference between the KEMAR ITDs and those predicted by the sphere model. In the manikin measurements, the magnitudes of the ITDs increase substantially in the front hemisphere when the distance of the source decreases from 25 to 12 cm. The majority of this increase occurs within a few degrees of the median plane, resulting in a substantial increase in the slope of the ITD at 12 cm near 0 degrees azimuth (bottom panel of Fig. 10). The slope of the ITD as a function of azimuth increases from 11 to 17 μs/degree as a sound source directly in front of the listener moves from 25 to 12 cm. At locations more than 10 degrees from the median plane, the slope of the ITD as a function of azimuth is approximately the same at 12 and 25 cm, and the magnitude of the ITD at 12 cm remains approximately 50 μs greater than the ITD at 25 cm over a large portion of the front hemisphere. The increase in ITD for a very close source directly in front of the face is most likely caused by the interference of the nose with the propa-

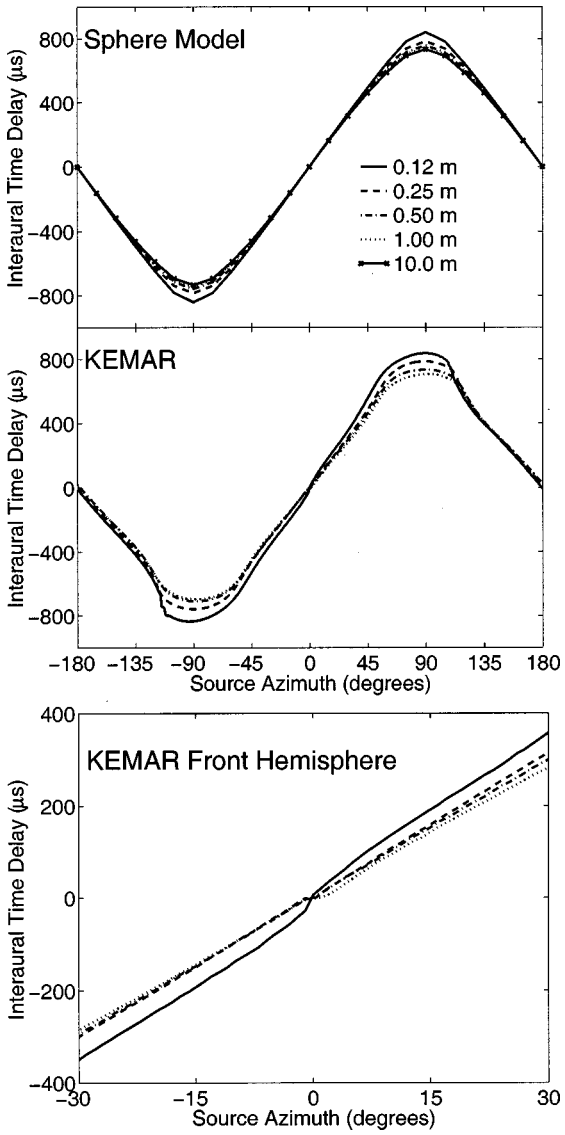


FIG. 10. The ITDs for the sphere model of the head (upper panel) and KEMAR (middle and lower panels). The delay was determined from the best linear fit of the unwrapped phase difference between the left and right ears (see text for details). The sphere model results assume a head diameter of 18 cm. The bottom panel is an enlarged view of the middle panel around 0 degrees. Positive delays indicate a lag at the right ear, and negative delays indicate a lag at the left ear.

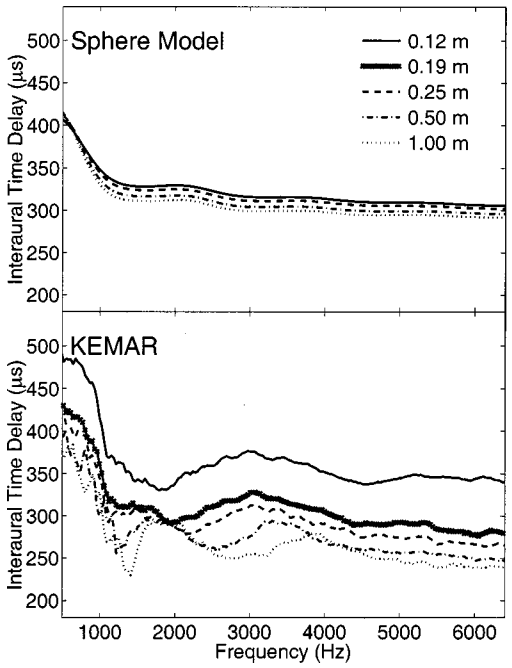


FIG. 11. The ITD (phase delay) for a source at 30 degrees azimuth calculated with the sphere model (upper panel) and measured with KEMAR (lower panel), as a function of frequency. Note that the plot at 19 cm is shown only for the KEMAR measurements.

gation of the sound around the head, since at 12 cm the sound source is located within 1 cm of the tip of the nose when the source is at 0 degrees. Since listeners are most sensitive to changes in ITD when the ITD is near zero (Hershkowitz and Durlach, 1969), it is likely that the increase in the rate of change of the ITD as a function of azimuth at locations directly in front of the face also results in increased sensitivity to changes in the azimuth of a sound source in that region.

The frequency-averaged values of the ITD shown in Fig. 10 provide a reasonable overall measure of the changes in the ITD with distance. However, the ITD is not constant across all frequencies, so it is useful to examine the frequency dependence of the ITD as a function of distance. The frequency dependence of the ITD for a nearby source at 30 degrees azimuth is shown in Fig. 11. These time delay values were calculated directly from the unwrapped interaural phase spectrum at each frequency. Both the sphere model calculations and the KEMAR measurements indicate that the time delay increases substantially at low frequencies (below 1500 Hz), a phenomenon reported by Kuhn (1977). In the sphere model, the time delay for a source at 1 m is $\approx 290\,\mu$s at high frequencies (>14 kHz) and $\approx 430\,\mu$s at low frequencies. Kuhn's model of ITD, based on Woodworth's geometrical model of the ITD, would predict an ITD of 278 μs at high frequencies and 420 μs at low frequencies for a distant sound source. Thus the sphere-model calculations are basically in agreement with the predictions of Kuhn's model. Note that the sphere-model calculations indicate very little change in the time delay with the distance of a sound source at 30 degrees azimuth.

The ITD measurements with the KEMAR manikin for a source at 30 degrees (bottom panel of Fig. 11) indicate a much larger distance dependence in the ITD measurements. In particular, they indicate an increase in ITD of approximately 70 μs as the distance of the source decreases from 25 to 12 cm. In this plot, an additional measurement at 19 cm has been added to illustrate how quickly the time delay increases as the source gets close to the face: more than half of the total increase in ITD from 1 m to 12 cm occurs between 19 and 12 cm. Note that the frequency dependence of the ITD is largely independent of the distance of the source. The ITDs increase systematically at low frequencies at each source distance. The ITD measured with KEMAR at 1 m is similar to the ITD for a clothed KEMAR manikin, as reported by Kuhn (Kuhn, 1977). Both the sphere model and the KEMAR measurements indicate that the dependence of the interaural time delay on distance is substantially weaker than the dependence of the interaural level difference on distance. At the closest distances, the magnitude of the ILD increases dramatically, while the ITD generally increases by no more than 10%–12% (with the exception of very close sources directly in front of the face). The reason for this discrepancy is simple: the ITD depends (roughly) on the arithmetic difference between the source distance to the ipsilateral ear and the distance to the contralateral ear. The ILD (ignoring head-diffraction effects) depends on the ratio of these distances. As the source approaches the head, the ratio of distances to the ipsilateral and contralateral ears increases much faster than the arithmetic difference between the distances, so the ILD increases more dramatically than the ITD. As discussed later, this disparity between the distance dependence of ILD and ITD is even greater when perceptual factors are considered.

IV. ELEVATION DEPENDENCE OF THE PROXIMAL-REGION HRTF

The high-frequency characteristics of the HRTF at the ipsilateral ear change substantially with elevation. This is illustrated in Fig. 12, which shows the high-frequency HRTF at three elevation locations and three source distances. The pattern of peaks and notches in the transfer functions is substantially different at each elevation location. At -30 degrees elevation, for example, there is a notch at approximately 7 kHz in the HRTF that stretches across the entire ipsilateral hemisphere which is not found at either of the other two elevations. Similar patterns have been reported in previous HRTF studies (Carlile and Pralong, 1994; Shaw, 1974) and they will not be discussed in detail here. It is apparent, however, that these patterns do vary with elevation, and that they could provide a salient cue for evaluating the elevation of a sound source.

The HRTF patterns appear relatively independent of source distance. Specifically, in Fig. 12, the major features of the HRTFs are considerably more consistent across the three distances when elevation is fixed (within each row of the figure) than across the three elevations when distance is fixed (within each column). If this result generalizes to all elevations, it would imply that elevation cues are roughly independent of distance and that the same mechanisms that mediate elevation perception in the distal region may also be used in the proximal region.

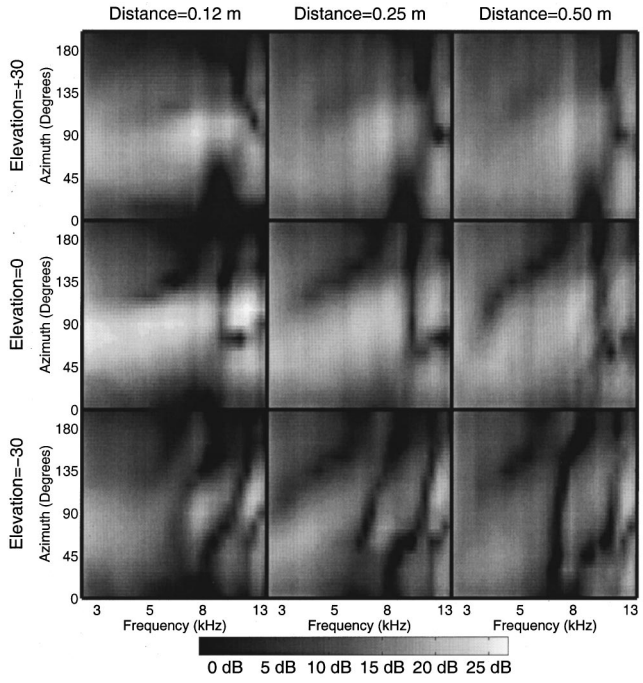


FIG. 12. Surface plots of the left-ear HRTFs measured with the KEMAR manikin at +30, 0, and −30 degrees in elevation. As in Fig. 8, the darkness of the plots indicates the magnitude of the HRTF at each point: brighter areas indicate greater gain in the HRTF. The plots are limited to high frequencies at the ipsilateral ear, where the most salient elevation cues occur, and are shown with 1/12-oct resolution in frequency and 15 degrees resolution in azimuth (interpolation has been used to smooth the plots).

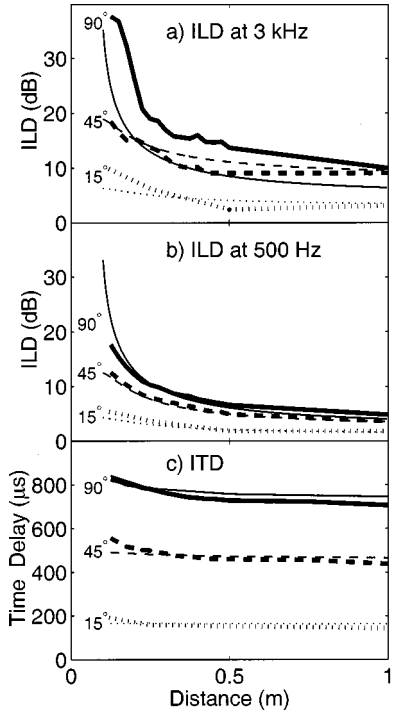


FIG. 13. Effects of source distance on interaural differences for the sphere model (thin line) and KEMAR (thick line). Results for (a) IID at 3 kHz, (b) IID at 500 Hz, and (c) ITD for three azimuths (15, 45, and 30 degrees) at 0-degree elevation.

V. PERCEPTUAL IMPLICATIONS OF THE PROXIMAL-REGION HRTFS

The distance-dependent attributes of the proximal-region HRTFs and consequent changes in interaural differences may have important perceptual implications. The ILDs increase rapidly as distance decreases, especially at distances less than 0.5 m, while the ITDs increase only slightly. Perceptual issues may further increase this disparity. Hershkowitz and Durlach (1969) found that listeners could discriminate changes in ILD on the order of 0.8 dB over a wide range of baseline ILDs. Therefore, the distance-dependent changes in ILD from 0.125 to 1 m (Fig. 13) span a range of up to 15 jnd's at 500 Hz and 30 or more jnd's at 3 kHz. In contrast, the jnd for ITD was approximately 15 μs at ITDs below 400 μs and increased rapidly for ITDs greater than 400 μs (Hershkowitz and Durlach, 1969). The changes in ITD in the proximal region span a few jnd's, at most. Thus, subjects can be expected to perceive large changes in ILD as the distance of a nearby source changes while the perceived ITD should remain relatively constant.

The combination of perceptually invariant ITDs and strongly distance-dependent ILDs suggests a possible strategy for determining the distance of a nearby source in the horizontal plane based on binaural information from the HRTFs. The ITD information, which is relatively independent of distance, could be used to identify the azimuthal direction of the sound source. With source azimuth determined, ILD could then be used to estimate the distance of a sound source, provided the source is outside the median plane (where the ILD is near zero at all distances). The dependence of ILDs on azimuth and distance predicts the following characteristics in proximal-region distance estimation performance:

(1) Distance accuracy would be best for sources to the side of the listener and worst for sources directly in front, behind, or anywhere in the median plane.
(2) At a fixed azimuth, the jnd in percentage distance would increase as distance increases. In Fig. 13, the slope of the curve relating ILD to distance increases substantially as the source approaches the head. Thus the percent decrease in distance necessary to produce a fixed increase in ILD decreases as distance decreases. If the jnd in distance is defined as the percent decrease in distance necessary to produce a single jnd in interaural intensity, the distance jnd will decrease with distance.
(3) Provided a source is sufficiently broadband to allow the listener to perceive both ITD and ILD information, distance perception would not require information about the spectral shape or intensity of the source. The key advantage to binaural depth perception, in contrast to depth perception based on intensity, spectral cues, or reverberation, is that the binaural information is derived from the difference signal between the left and right ears and does not depend on the characteristics of the source (except its frequency range).

Note that this model is similar in concept to one suggested by Hirsch (1968), further explored by Greene (1968), and expanded by Molino (1970). Hirsch's model demonstrated the possibility of determining the distance of a sound source based on the relationship between the ILD and ITD. His model, which ignored diffraction by the head and as-

sumed that the ears were detectors in free space, predicted that distance could be calculated directly from the ratio of ITD to ILD. Molino's expanded model, based on a spherical head, required the azimuth location of the source to be known *a priori*. Greene and Molino used threshold data for ITD and ILD to calculate the predicted accuracy of distance perception using this model and, predictably, found that distance perception in the distal region would be very poor. Molino noted that predicted accuracy would be much greater in the proximal region, due to the dramatic increase in ILDs in that region. The present data indicate not only that the changes in ILD in the proximal region are large, and should be easily perceptible, but also that the relative invariance of the ITD in the proximal region may allow listeners to determine azimuth directly from the ITD, without any other external knowledge of source direction. Outside of the horizontal plane, the situation becomes more complex, but it is still possible to determine the azimuth, elevation, and distance of a sound source from the HRTFs. Recall that the KEMAR measurements at 30 and -30 degrees indicate that the high-frequency, elevation-dependent features in the HRTF, which are believed to be important elevation cues, are roughly independent of distance. If these high-frequency cues could be used to determine the elevation of the source, and compensate for the elevation-dependent changes in the ILD and ITD, the azimuth and distance of the source could still be accurately determined from interaural cues. This model would imply greater distance accuracy in the horizontal plane than in the median plane, since ILDs decrease as a source moves directly above or below a listener.

Another possible distance cue indicated by the HRTF measurements is a modest increase in the relative low-frequency gain as distance decreases. As a result, the signal reaching the eardrums from a nearby source is effectively low-pass filtered as the source approaches the head (Fig. 4). This low-pass filtering may explain the ''darkening'' of a very near sound source reported by von Bekesy (1960). Begault (1987) also noted the ''darkening'' in the timbre of very near sound sources, and suggested that the tendency to perceive larger increases in loudness at low frequencies than at higher frequencies from an equivalent increase in sound pressure level (the so-called Fletcher–Munson curve) might provide an explanation. Since the pressure level increases at all frequencies as the source approaches the ear, the Fletcher–Munson curve suggests that the perceived increase in loudness would be greater at low frequencies. The ''darkening'' of proximal-region stimuli reported in the literature is probably a combination of the boost in low-frequency gain due to proximal-region acoustic effects shown in the HRTFs and the nonuniform perception of increasing loudness across frequency.

The implications of the proximal-region properties of the HRTF in directional localization are less obvious than those in distance localization. There is evidence that low-frequency time delays (Wightman and Kistler, 1992) dominate the perception of azimuth when they are available. Since time delays vary only slightly as distance decreases, it is not likely that azimuth perception will be significantly different in the proximal region than in the distal region when the spectrum of the source extends into low frequencies. However, the increase in the slope of the ITD as a function of azimuth for nearby sources directly in front of the nose might increase the sensitivity to changes in the azimuth of the source in that region. Note that, when time delay information is not available from low-frequency time delays or high-frequency envelope delays (Middlebrooks *et al.*, 1989), it is possible that localization ability may degrade substantially. Without time delay information, there is no obvious mechanism for determining the relative contributions of distance and direction to the ILD. In other words, there is no way to determine whether a certain ILD is the result of a distant sound source near the interaural axis or a nearby sound source near the median plane. The consequences of this confusion on localization performance are unclear. The increase in ILDs at close distances will, however, decrease the jnd in azimuth to the extent that the jnd is limited by the change in ILD. The change in ILD with head orientation could also provide a strong dynamic distance cue when exploratory head motions are allowed. In addition, the lateral shift in the high-frequency pinna cues for nearby sources caused by the acoustic parallax effect may introduce a lateral bias if the auditory system is using these cues in the perception of azimuth.

The high-frequency pinna cues that are believed to determine elevation were found to be roughly independent of distance over the limited range of elevations measured. These cues are believed to dominate the perception of elevation, so it is unlikely that the localization of elevation is strongly dependent on distance in the proximal region.

VI. CONCLUSIONS

The major conclusions of this study can be summarized as follows:

(1) The dominant distance-dependent feature of both the sphere model and KEMAR HRTFs in the proximal region is an increase in the interaural level difference with decreasing source distance across all frequencies. When a source is near the interaural axis, the change in ILD can span 20–30 dB (25 or more jnd's) as distance moves from 1 to 0.12 m, providing a potentially strong binaural distance cue in the proximal region. It is particularly interesting that large ILDs occur at low frequencies exclusively in the proximal region. In the distal region, low-frequency ILDs are small at all source directions.
(2) In contrast to the ILD, the ITD is roughly independent of distance in the proximal region.
(3) Both the sphere model and the KEMAR measurements indicate that the average low-frequency gain of the HRTF increases more rapidly than the average high-frequency gain as the source approaches the head. Thus, the sound reaching the ears is effectively low-pass filtered as the source approaches the head. This filtering may serve as a spectral distance cue in the proximal region.
(4) There is a substantial parallax effect in the ipsilateral HRTF due to the discrepancy between the orientation of the source relative to the ear and the orientation of the

source relative to the head. This parallax effect causes a lateral shift of 20° or more in the high-frequency features of the HRTF as the source approaches the head.

(5) For a fixed azimuth, the high-frequency characteristics of the HRTFs vary systematically with elevation but are relatively invariant to changes in distance. As a result, elevation localization may be similar in the proximal and distal regions.

In summary, it is clear that the distance-dependent attributes of the HRTF in the proximal region provide possible binaural distance cues which are unavailable for more distant sources. There are also substantial differences in the azimuth-dependent characteristics of the HRTF, particularly in the ILD, when the source is close the head, but their effects are probably mitigated by the relative insensitivity of the perceptually dominant interaural time delay to source distance. The high-frequency elevation-dependent characteristics of the HRTF also appear to be relatively independent of distance. Later papers in this series will discuss experiments measuring localization accuracy in the proximal region and whether listeners are able to make use of the available distance cues in that region.

ACKNOWLEDGMENTS

The authors would like to thank Nat Durlach, Steve Colburn, and William Peake for their assistance in the preparation of this manuscript. The work was also partially supported by AFOSR Grant Nos. F49620-96-1-0202 and F49620-98-1-0108.

[1]In the past, we have used the terms ''near field'' and ''far field'' to describe these somewhat arbitrary regions. However, because the terms ''near field'' and ''far field'' have very specific meanings in physical acoustics (and have been used loosely to describe a number of other phenomena in other areas of acoustics), we have decided to eliminate any ambiguities about our terminology by introducing the terms ''proximal region'' and ''distal region'' and precisely defining them in the context of auditory localization.

[2]This chamber ($10\times10\times10\ m^3$) currently contains the Auditory Localization Facility (ALF). The ALF is a large, wire-frame geodesic sphere used in localization experiments at the Armstrong Laboratory (McKinley *et al.*, 1994). Because of the ALF, HRTF measurements were made in a corner of the chamber. Acoustic measurements with a free-field microphone verified that the presence of the ALF did not substantially affect the anechoic conditions for the measurements in this study.

Begault, D. R. (**1987**). ''Control of Auditory Distance,'' Ph.D. thesis, University of California, San Diego.

Brungart, D. (**1998**). ''Near-Field Auditory Localization.'' Ph.D. thesis, Massachusetts Institute of Technology.

Brungart, D., and Rabinowitz, W. (**1996**). ''Auditory localization in the near-field,'' in Proceedings of the Third International Conference on Auditory Display, Santa Fe Institute.

Brungart, D. S., Rabinowitz, W. M., and Durlach, N. (**1996**). ''Auditory localization of a nearby point source,'' J. Acoust. Soc. Am. **100**, 2593.

Carlile, S., and Pralong, D. (**1994**). ''The location dependent nature of perceptually salient features of the head-related transfer function,'' J. Acoust. Soc. Am. **95**, 3445–3459.

Duda, R., and Martens, W. (**1998**). ''Range dependence of the response of a spherical head model,'' J. Acoust. Soc. Am. **104**, 3048–3058.

Firestone, F. (**1930**). ''The phase difference and amplitude ratio at the ears due to a source of pure tone,'' J. Acoust. Soc. Am. **2**, 260–270.

Gardner, W., and Martin, K. (**1995**). ''HRTF measurements of a KEMAR,'' J. Acoust. Soc. Am. **97**, 3907–3908.

Greene, D. (**1968**). ''Comments on 'Perception of the range of a sound source of unknown strength,'' ' J. Acoust. Soc. Am. **44**, 634.

Hartley, R., and Fry, T. (**1921**). ''The binaural location of pure tones,'' Phys. Rev. **18**, 431–442.

Hershkowitz, R., and Durlach, N. (**1969**). ''Interaural time and amplitude JNDs for a 500-Hz tone,'' J. Acoust. Soc. Am. **46**, 1464–1467.

Hirsch, H. (**1968**). ''Perception of the range of a sound source of unknown strength,'' J. Acoust. Soc. Am. **43**, 373–374.

Kuhn, G. (**1977**). ''Model for the interaural time differences in the azimuthal plane,'' J. Acoust. Soc. Am. **62**, 157–167.

McKinley, R., Ericson, M., and D'Angelo, W. (**1994**). ''Three dimensional audio displays: development, applications, and performance,'' Aviat., Space Environ. Med. **65**, A31–A38.

Mehrgardt, S., and Mellert, V. (**1977**). ''Transformation characteristics of the external human ear,'' J. Acoust. Soc. Am. **61**, 1567–1576.

Middlebrooks, J., Makous, J. C., and Green, D. M. (**1989**). ''Directional sensitivity of sound-pressure levels in the human ear canal,'' J. Acoust. Soc. Am. **86**, 89–107.

Molino, J. (**1970**). ''Perceiving the range of a sound source when the direction is known,'' J. Acoust. Soc. Am. **53**, 1301–1304.

Møller, H., Sørensen, M., Hammershøi, D., and Jensen, C. B. (**1995**). ''Head-related transfer functions of human subjects,'' J. Audio Engineering Society, **43**, 300–320.

Pralong, D., and Carlile, S. (**1994**). ''Measuring the human head related transfer functions: A novel method for the construction and calibration of a miniature 'in ear' recording system,'' J. Acoust. Soc. Am. **95**, 3435–3444.

Rabinowitz, W., Maxwell, J., Shao, Y., and Wei, M. (**1993**). ''Sound localization cues for a magnified head: implications from sound diffraction about a rigid sphere,'' Presence **2**, 125–129.

Shaw, E. (**1974**). ''Transformation of sound pressure level from the free field to the eardrum in the horizontal plane,'' J. Acoust. Soc. Am. **56**, 1848–1861.

Stewart, G. (**1911a**). ''The acoustic shadow of a rigid sphere with certain applications in architectural acoustics and audition,'' Phys. Rev. **33**, 467–479.

Stewart, G. (**1911b**). ''Phase relations in the acoustic shadow of a rigid sphere; phase difference at the ears,'' Phys. Rev. **34**, 252–258.

Suzuki, Y., Kim, H.-Y., Takane S., and Sone, T. (**1998**). ''A modeling of distance perception based on auditory parallax model,'' in Proceedings of the 16th International Congress on Acoustics and the 135th Meeting of the Acoustical Society of America.

von Bekesy, G. (**1960**). *Experiments in Hearing* (Wiley, New York).

Wightman, F., and Kistler, D. (**1989**). ''Headphone simulation of free-field listening. I: Stimulus synthesis,'' J. Acoust. Soc. Am. **85**, 858–877.

Wightman, F., and Kistler, D. (**1992**). ''The dominant role of low-frequency interaural time differences in sound localization,'' J. Acoust. Soc. Am. **91**, 1648–1660.

Individual differences in external-ear transfer functions reduced by scaling in frequency

John C. Middlebrooks
Kresge Hearing Research Institute, University of Michigan, 1301 East Ann Street, Ann Arbor, Michigan 48109-0506

(Received 1 September 1998; revised 10 March 1999; accepted 26 May 1999)

This study examined inter-subject differences in the transfer functions from the free field to the human ear canal, which are commonly know as head-related transfer functions. The directional components of such transfer functions are referred here to as *directional transfer functions* (DTFs). The DTFs of 45 subjects varied systematically among subjects in regard to the frequencies of spectral features such as peaks and notches. Inter-subject *spectral differences* in DTFs were quantified between 3.7 and 12.9 kHz for sound-source directions throughout the coordinate sphere. For each pair of subjects, an optimal frequency scale factor aligned spectral features between subjects and, thus, minimized inter-subject spectral differences. Frequency scaling of DTFs reduced spectral differences by a median value of 15.5% across all pairs of subjects and by more than half in 9.5% of subject pairs. Optimal scale factors showed a median value of 1.061 and a maximum of 1.38. The optimal scale factor between any pair of subjects correlated highly with the ratios of subjects' maximum interaural delays, sizes of their external ears, and widths of their heads.
 [S0001-4966(99)01409-5]

PACS numbers: 43.66.Pn, 43.64.Ha, 43.66.Qp [DWG]

INTRODUCTION

The physical structures of the head and external ears transform the spectrum of a sound as it travels from the source to the ear canal. The transfer function from the source to any of several points in the ear canal has come to be known as the head-related transfer function (HRTF; e.g., Morimoto and Ando, 1980; Butler, 1987; Wightman and Kistler, 1989a). The HRTF contains prominent spectral features, such as peaks and notches, that vary according to source direction (e.g., Batteau, 1967; Blauert, 1968/69; Shaw, 1974). Direction-specific spectral features provide cues to the directions of sound sources (reviewed by Middlebrooks and Green, 1991). When a listener hears through headphones sounds that have been filtered with HRTFs measured from his or her own ears, a ''virtual acoustic environment'' results. The listener reports sensations of sounds that appear to originate from well-defined directions in the space surrounding the subject (Wightman and Kistler, 1989a, b; Bronkhorst, 1995; Hartmann and Wittenberg, 1996; Møller *et al.*, 1996). Subjects can localize such virtual targets with accuracy that approaches the accuracy of localization of actual free-field sound sources. Conversely, when targets are synthesized using HRTFs that have been recorded from the ears of a different subject, the virtual environment is less vivid and localization is less accurate. In particular, virtual localization with nonindividualized HRTFs is characterized by a high incidence of large errors in vertical and front/back localization (Morimoto and Ando, 1980; Wenzel *et al.*, 1993; Møller *et al.*, 1996). The goal of the present study was to identify systematic inter-subject differences in HRTFs that might account for this degradation of the virtual environment.

Previous reports from several research groups have mentioned inter-subject differences in HRTFs (e.g., Shaw, 1966; Mehrgardt and Mellert, 1977). For instance, Wightman and Kistler (1989a) reported that inter-subject standard deviations within 1/3-octave bands of HRTFs could reach nearly 8 dB for bands centered between 7 and 10 kHz. In our previous work, we have noted that the directional sensitivity (i.e., direction-specific amplitude gain) of a particular subject's ear at one frequency can closely resemble the directional sensitivity of another subject's ear at a different frequency (Middlebrooks *et al.*, 1989). Moreover, we found that the frequencies at which subjects showed similar directional sensitivity tended to rank inversely with the overall sizes of subjects. Intuitively, one can think of lower-frequency sounds, with longer wavelengths, as interacting with larger heads and ears in a similar way that higher-frequency sounds, with shorter wavelengths, interact with smaller heads and ears. These observations suggest that it might be possible to reduce inter-subject differences in HRTFs by scaling them in frequency. Reductions in inter-subject differences in HRTFs, in turn, might lead to improvements in virtual localization using nonindividualized HRTFs.

This report considers the directional components of HRTFs, which are referred to as *directional transfer functions* (DTFs). Two hypotheses are tested. The first is that DTFs differ systematically among subjects in regard to the frequencies of spectral features. The second is that the optimal scale factor between the DTFs of two subjects can be predicted by the relative physical sizes of the subjects. The results largely confirm both hypotheses. The companion study (Middlebrooks, 1999) tests the hypothesis that reductions in inter-subject differences produced by frequency scaling can result in improvements in virtual localization.

I. METHODS

A. Subjects

Subjects were recruited from the student population of the University of Michigan and from members and associates of the Kresge Hearing Research Institute. Subjects ranged from 21 to 50 yr in age (30.2 ± 6.3, mean $\pm$ standard deviation). Twenty-three were male, and 22 were female. On the basis of predominant racial heritage, 34 of subjects were European, 9 were Asian, and 2 were African. No subjects were recruited or excluded because of properties of their ears, but none of the subjects exhibited any conspicuous abnormalities of external ears, and none reported any hearing disorder. Some effort was made to sample subjects representing a broad range of physical sizes. Subjects ranged from 152 to 193 cm in height (171 ± 10.4). During measurements, subjects removed earrings and eyeglasses and fastened long hair so that the paths to the ears were clear.

B. Apparatus and sound generation

Acoustical measurements were made inside a double-walled anechoic chamber. The usable space inside the sound-absorbent cones was $2.6 \times 3.7 \times 3.2$ m. The chamber contained a mechanism that could position two loudspeakers on the surface of an imaginary sphere that was 1.2 m in radius. The mechanism was driven by computer-controlled stepping motors. The loudspeakers were separated by 160° on the sphere, and one loudspeaker or the other could be positioned at any azimuth throughout 360° and at any elevation from 70° below to 90° above the horizontal plane. The subject stood on a platform that was adjusted in height to place the center of his or her interaural axis at the center of the coordinate sphere.

The longest acoustical wavelength of interest in this study was 113.3 mm (at 3 kHz) and the largest head width was 153 mm (Sec. II D). Thus the 1200-mm distance between the loudspeaker and the center of the subject's head corresponded to roughly 10.6 wavelengths and 7.8 head widths. For that reason, the stimulus conditions are regarded as a reasonable approximation of the acoustical far field.

The loudspeakers were Pioneer model TS-879 two-way coaxials. Audio signals were generated and recorded using an Intel-based personal computer that was equipped with a 16-bit analog interface and floating-point array processor from Tucker-Davis Technologies (TDT). Output and input sample rates were 50 000 samples per second.

C. Measurements of transfer functions

Transfer functions were measured by recording from subjects' ear canals with miniature microphones while presenting broadband sounds from various locations. Recordings were made simultaneously from Knowles model 1934 microphones placed in both ears approximately 5 mm deep to the entrances of the ear canals. The transfer function from the free field to a microphone in an ear canal varies substantially with changes in the position of the microphone in the standing-wave pattern in the canal. Nevertheless, we have shown previously that, at any particular frequency, measurements of the directional gain of the ear are largely insensitive to the position of a microphone at any point deep to the ear canal entrance (Middlebrooks *et al.*, 1989). Moreover, we have described a procedure, reviewed below, by which one can isolate a directional component of the transfer function that is essentially independent of microphone position in the ear canal (Middlebrooks and Green, 1990).

Probe stimuli consisted of ''Golay codes,'' which are complementary pairs of binary sequences that have correlation structure equivalent to many repetitions of a single impulse (Golay, 1961; Foster, 1986; Zhou *et al.*, 1992). At each sound-source location, a single pair of 512-point codes was presented at an output rate of 50 000 samples per second. That resulted in a 256-point spectrum that ranged from 0 to 25 kHz with resolution of 97.7 Hz. Normally, for the purpose of rapid data acquisition, complementary Golay codes were presented immediately one after the other. There was initial concern that reverberation from the response to the first code might contaminate the response to the second code, but preliminary control experiments proved that contamination to be negligible. Probe stimuli were presented from 400 loudspeaker locations that ranged through 360° of azimuth and in elevation from -70 to 90°. The locations were distributed in azimuth and elevation so that each location represented a roughly equal area on the coordinate sphere and so that neighboring locations were separated by angles of no more than 10° measured at the center of the sphere. At each location, the probe sound level was adjusted so that the amplified microphone signal into the analog-to-digital converter was 50%–99% of the clipping level of the converter.

On the day of each set of measurements, the transfer functions of the two loudspeakers were measured using the Golay probe stimuli and a $\frac{1}{2}$-in. precision microphone (ACO Pacific) positioned in the center of the testing chamber in the absence of the subject. The transfer functions from the loudspeakers to the ear canal were computed by dividing the spectra measured with the ear canal microphones by the transfer functions of the loudspeakers. The influence of the transfer functions of the ear canal microphones was eliminated along with the lumped ''common component'' described below.

Measurements from each subject were completed in approximately 8.5 min. The subject was instructed to stand stationary during the measurements, to hold his or her mouth closed with the jaw slack so that the teeth did not touch, and to refrain from swallowing. A handrail was provided to help the subjects remain stationary, but otherwise there was no head rest or other restraint. Subjects were observed with a video monitor. In a few instances in which a subject coughed or moved or reported swallowing, measurements were halted and begun again. With those exceptions, however, movement of subjects appeared not to contribute appreciably to variance in measurements. This was confirmed by the test–retest reliability of results (Sec. II C).

Each external-ear transfer function was treated as the product of two transfer functions. The first, which contained components that were specific to each sound-source direction, was referred to as the *directional transfer function* (DTF; Middlebrooks and Green, 1990). The second con-

tained components that were common to all directions. Contributors to the common component included the ear canal resonance and the microphone transfer function. The magnitude spectrum of the common component was estimated for each set of measurements in each ear by computing the root-mean-squared (rms) of the sound pressure at each frequency averaged across all 400 sound-source locations. The minimum-phase spectrum of the common component was given by the Hilbert transform of the logarithm of the magnitude spectrum (Oppenheim and Schafer, 1975; Middlebrooks and Green, 1990). The complex common component was formed by the common magnitude spectrum and the minimum phase spectrum. The DTF for each sound-source location was calculated by dividing the complex transfer function measured at each location by the complex common component.

D. Data analysis

The linear spacing of frequency components in measured DTFs sampled higher octaves more densely than lower octaves. To avoid over-representing higher frequencies and for ease in scaling DTFs in frequency, DTFs were processed with a bank of band-pass filters to sample frequency components at equal intervals on a base-2 logarithmic (i.e., octave) scale. The filter bank consisted of 85 triangular filters. The 3-dB bandwidth of the filters was 0.057 octave, filter slopes were 105 dB per octave, and center frequencies were spaced in equal intervals of 0.0286 octave from 3 to 16 kHz. The interval of 0.0286 octave was chosen to give intervals of 2% in frequency, which provided suitable resolution in preliminary tests of frequency scaling.

We scaled DTFs in frequency by applying constant offsets in logarithmic frequency. An upward scaling of 2%, for instance, was accomplished by offsetting each component upward by 0.0286 octave. Scale-factor axes in the illustrations are drawn on logarithmic scales to indicate the equal logarithmic steps in which scale factors were manipulated. Also, computation of correlation coefficients and of root-mean-squared (rms) differences between sets of scale factors utilized the base-2 logarithms of scale factors. The advantage of a logarithmic scale is that the logarithm of a factor and of its inverse are symmetrical about the origin. For instance, linear factors of 2 and $\frac{1}{2}$ correspond to $+1$ and -1 octave. For convenience to the reader, the results of computations conducted using logarithms are expressed as linear factors.

The set of tested sound-source locations sampled some regions of space somewhat more densely than others. To provide uniform sampling and to permit rotation of the coordinate sphere (needed for Sec. II B), a spatial interpolation procedure was employed. In that procedure, 393 sectors were defined on the coordinate sphere. Each sector had a solid angle of 0.0305 steradians. On the horizontal plane, the sectors were 10° wide in azimuth and 10° tall in elevation. At higher and lower elevations, the dimensions of sectors were adjusted to maintain a constant solid angle. The centers of sectors ranged through 360° of azimuth and from -60 to 90° in elevation. A DTF at the center of each sector was computed by forming a weighted average of the filter-bank-processed DTFs measured for all points within 16° of the center of the sector. The spatial interpolation was described in more detail and its validity tested in a previous report (Middlebrooks, 1992). Except when stated otherwise, all computations that involved only one ear were conducted with DTFs that were measured from subjects' right ears and that were processed with the filter bank and spatially interpolated as described above.

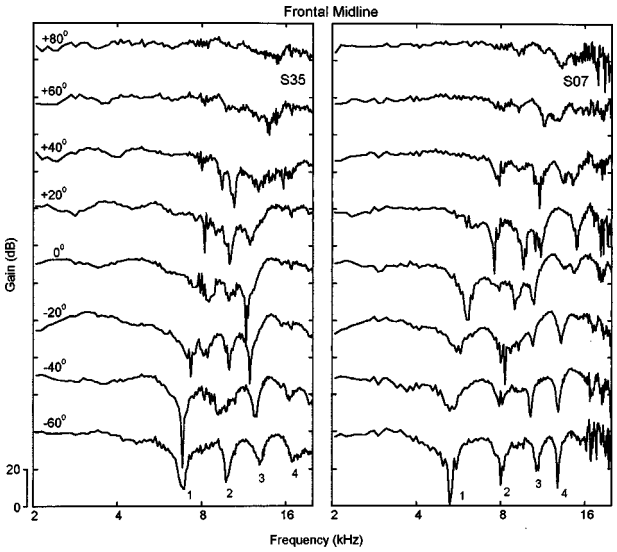


FIG. 1. Directional transfer functions (DTFs) for directions on the frontal midline at the elevations indicated relative to the horizontal plane. Left and right panels show DTFs measured from the right ears of subjects S35 and S07, respectively. Numbers 1 through 4 indicate spectral notches discussed in the text.

II. RESULTS

A. Scaling of DTFs among subjects

Directional transfer functions (DTFs) were recorded from the ears of 45 subjects. Examples of DTFs from the right ears of two subjects are shown in Figs. 1, 2, and 3. Those DTFs were measured for sound sources at eight elevations along the frontal (Fig. 1) and rear (Fig. 2) midlines and at eight frontal azimuths in the horizontal plane (Fig. 3). The left and right columns of plots represent measurements from subjects S35 and S07, respectively. The DTFs of these subjects, as well as those of all other subjects in the sample, showed patterns of spectral features (e.g., spectral peaks, notches, and slopes) that varied systematically as a function of sound-source elevation and azimuth. As sound sources were shifted in elevation from low to high, the center frequencies of spectral features tended to shift from low to high frequencies (Figs. 1 and 2). Changes in sound-source azimuth resulted in upward shifts in the frequencies of spectral features as source azimuth shifted from the side opposite the recorded ear, past the frontal midline, to the side of the recorded ear (Fig. 3).

The DTFs shown in both columns of Figs. 1 through 3 contain similar patterns of spectral features, but those spectral features tend to lie at higher frequencies for subject S35 (left column) than for subject S07 (right column). For descriptive purposes in Fig. 1, numbers identify four notches in the DTFs for elevation -60°. In the left column, notches 1

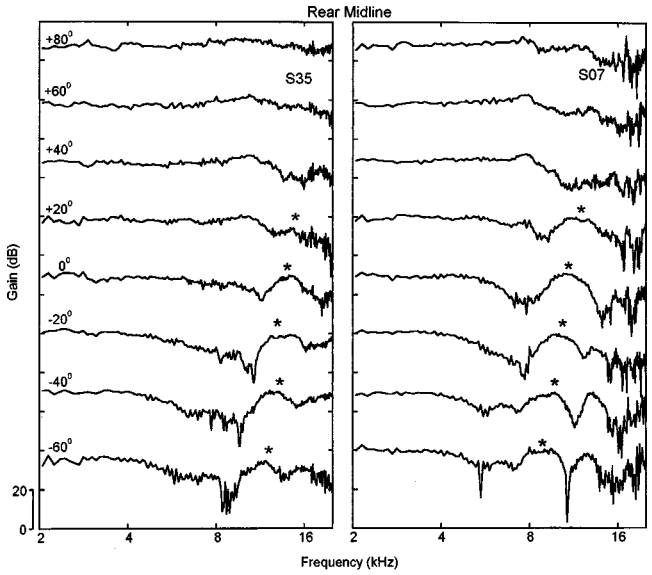


FIG. 2. DTFs for directions on the rear midline at the elevations indicated relative to the horizontal plane. The asterisks indicate spectral peaks discussed in the text.

through 4 lie between 6.9 and 16.8 kHz, whereas the corresponding notches lie about 25% lower in the right column, between 5.2 and 12.8 kHz. In the data for the rear midline (Fig. 2) asterisks indicate spectral peaks that can be seen in the DTFs for elevations −60 through +20°. Again, those peaks lie at higher frequencies for subject S35 than for subject S07. In Fig. 3, asterisks indicate the lowest-frequency notches. As the sound-source location was shifted from azimuths 0° to right 80°, the center frequencies of those notches shifted from 8.4 to 10.5 kHz in S35's DTFs and from 6.1 to 8.0 kHz in the S07's DTFs.

Quantitative comparisons among DTFs from various subjects used DTFs that had been filtered with a bank of 85

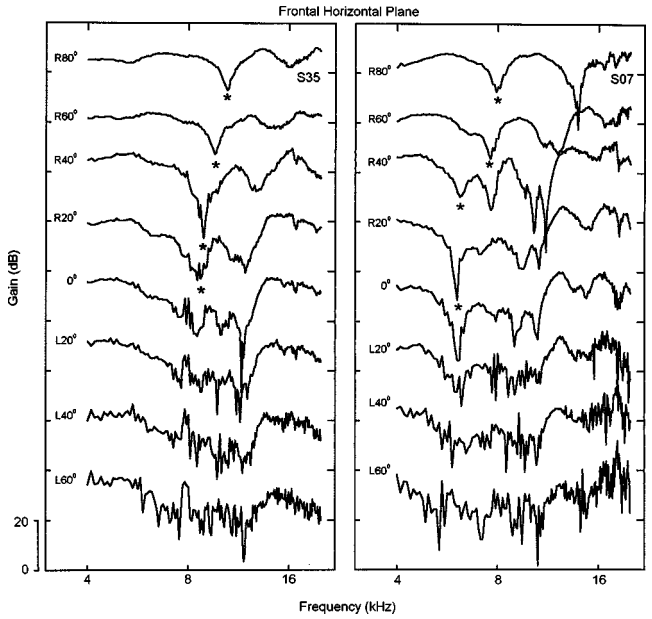


FIG. 3. DTFs for directions on the frontal horizontal plane at the azimuths indicated relative to the midline. L and R indicate left and right. The asterisks indicate spectral notches discussed in the text.

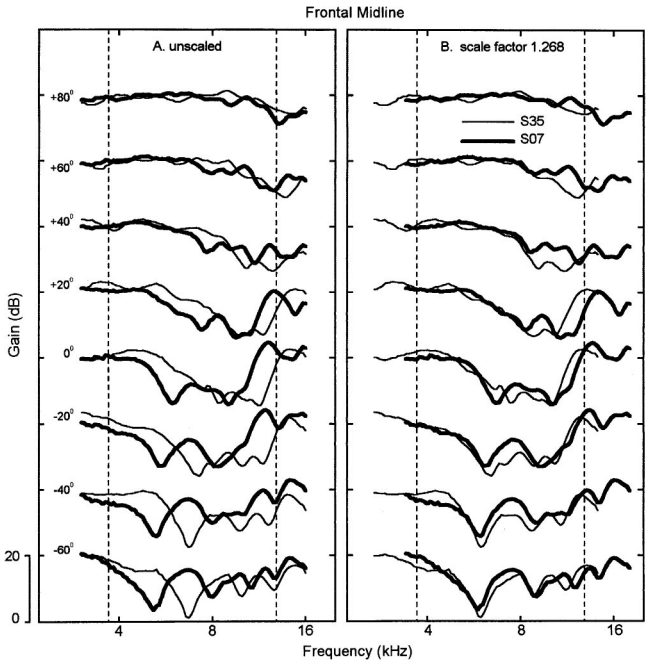


FIG. 4. DTFs for directions on the frontal midline at elevations indicated relative to the horizontal plane. These DTFs were processed with a bank of narrow-band filters, as described in Sec. I D. Thin and thick lines represent measurements from subjects S35 and S07, respectively. In the left panel, the DTFs are shown unscaled, as they were measured, whereas in the right panel subject S07's DTFs are scaled frequency by 1.126 and subject S35's DTFs are scaled by 1/1.126, a combined relative scaling of 1.28. The vertical grid lines indicate the frequency band over which quantitative comparisons were made.

bandpass filters (see Sec. I). Figure 4(A) shows DTFs from subjects S7 and S35 filtered in that way. One can see several instances in which a peak in the DTF from one subject corresponds in frequency to a notch in the DTF from the other subject. In Fig. 4(B), DTFs from subject S07 are scaled upward by a factor of 1.126, and those from subject S35 are scaled downward by 1.126. The combined inter-subject scaling by $1.268(1.126^2)$ resulted in a conspicuously better alignment of spectral features between the two subjects. Note that the DTFs that are represented in Fig. 4 contained frequency components in equal logarithmic steps. Frequency scaling upward or downward by a factor of 1.126 was accomplished by shifting components upward or downward by six 0.0286-octave steps.

Inter-subject differences between DTFs were quantified by a metric that will be referred to as the *inter-subject spectral difference*. That metric was computed across the 64 filter-bank components from 3.7 to 12.9 kHz. That frequency band is indicated in Fig. 4 by vertical dashed lines. The dB amplitudes of the 64 components of DTFs for each direction were subtracted, component by component, to form a difference spectrum. Then the variance across the 64 components of the difference spectrum was computed. Finally, the variance was averaged across 393 tested directions for each pair of subjects. The variance computation that was used to represent inter-subject spectral differences captured inter-subject differences that varied with frequency and eliminated differences, such as differences in overall gain, that were constant across frequency. For that reason, the variance com-

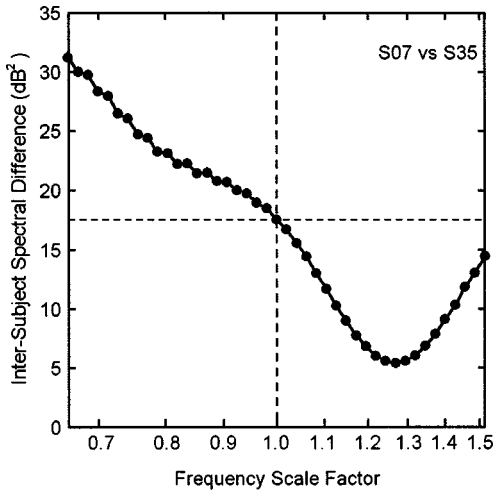


FIG. 5. Inter-subject spectral difference in sets of DTFs from subjects S07 and S35 as a function of relative frequency scale factor.

putation was preferred over a measure of the sum of the squared differences, or the square root thereof (the ''Euclidean distance''), which might have emphasized differences in overall gain at the expense of differences in spectral shape.

Figure 5 shows the inter-subject spectral difference between subjects S07 and S35 as a function of relative frequency scale factor. DTFs were scaled incrementally by factors of 1.02 (or 1/1.02) by making alternating upward logarithmic frequency shifts of 0.0286 octave in one set of DTFs and downward log frequency shifts in the other set. As a set of DTFs was scaled upward, high-frequency samples passed out of the high end of the analysis band and low-frequency amplitudes entered the low end of the band, as indicated Fig. 4(B). The plot in Fig. 5 shows a clear minimum spectral difference at a scale factor of 1.268. The inter-subject spectral difference dropped from 17.5 dB^2 prior to scaling to 5.4 dB^2 when subject S07's DTFs were scaled upward by 1.268 relative to those of subject S35. That is a 69.1% reduction in inter-subject spectral difference. The scale factor that produced the minimum spectral difference between any given pair of subjects will be referred to as the *optimal scale factor*.

Although inter-subject spectral differences normally were measured across DTFs for all sound-source directions, the scale factor that minimized spectral difference usually was quite similar when measured for DTFs for individual source directions. Figure 6(A) shows inter-subject spectral differences for one pair of subjects as a function of scale factor. The thick line shows the spectral difference averaged across all sound-source directions (the usual procedure). The optimal scale factor was 1.24. The thin lines show the spectral differences computed for three individual directions. Two of the thin lines show minima that correspond closely to the minimum of the thick line. The third thin line showed a second, slightly deeper, minimum at a scale factor less than one (indicated by the arrow). The histogram in Fig. 6(B) shows the distribution of optimal scale factors for the same pair of subjects computed for each of the source directions. The seven tallest bars in the histogram encompass 77% of the directions, and optimal scale factors computed for those directions fell within a range of 1.17–1.32. The histogram shows a second peak near a scale factor of 0.79. That peak corresponds to measurements that were made at a few direc-

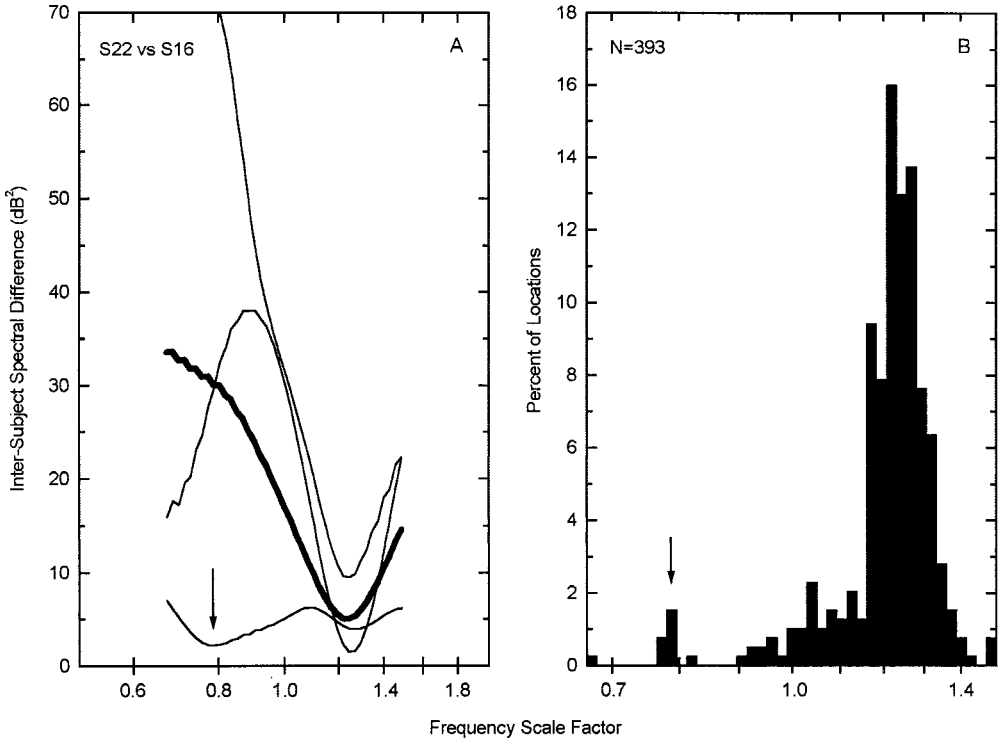


FIG. 6. (A) Inter-subject spectral difference between subjects S16 and S22 as a function of relative frequency scale factor. Values were computed for 3 individual sound-source directions (thin lines) or averaged across 393 sound-source directions (thick line). (B) Distribution of optimal scale factors for subjects S16 and S22, measured for 393 individual directions. The isolated mode around 0.79 represents 10 directions behind the subjects' heads at which the signal-to-noise ratio was low.

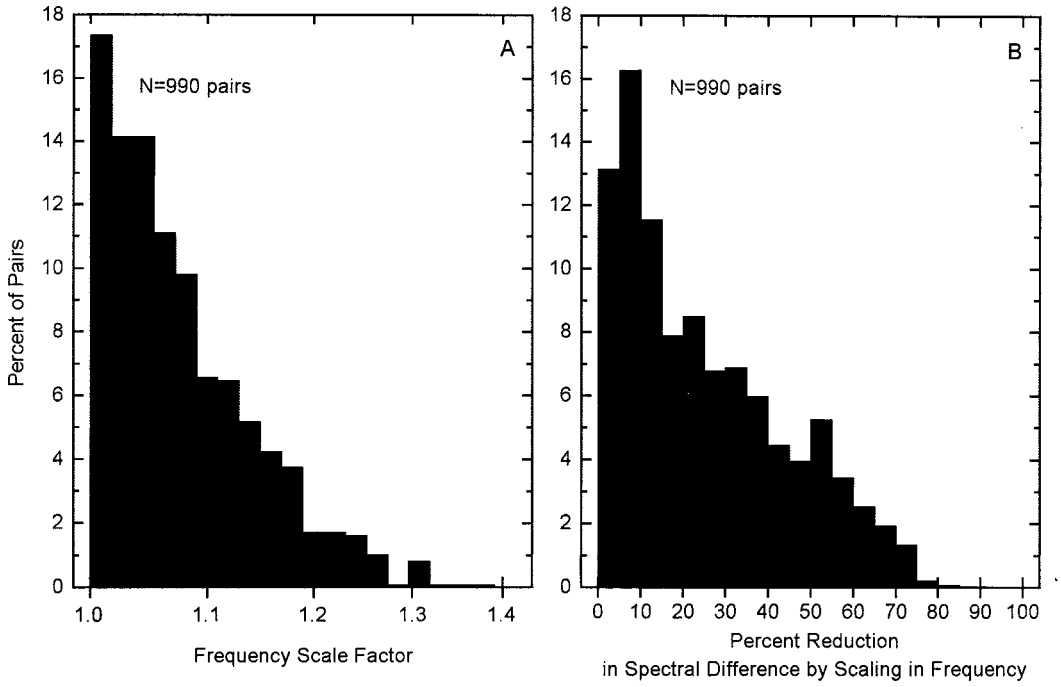


FIG. 7. Distributions of (A) optimal scale factors and (B) percent reduction of spectral difference by frequency scaling. The order of scaling was chosen so that all optimal scale factors were $>=1$. Values were computed for all 990 possible pairwise combinations of 45 subjects.

tions behind the head on the side opposite to the side of the recorded ear where signal-to-noise ratios were low. The spectral difference plot for one of those measurements is shown as a thin line in Fig. 6(A).

The distribution of optimal scale factors across all 990 pairs of subjects is shown in Fig. 7(A). The order of scaling was chosen so that optimal scale factors all were $>=1$, although for every scale factor X that scaled subject A to B, there was a complementary factor $1/X$ that scaled subject B to A. The median and maximum optimal scale factors were 1.061 and 1.38. Figure 7(B) shows the distribution of the percent reduction of the inter-subject spectral difference that was accomplished by scaling DTFs in frequency. Frequency scaling reduced spectral differences by 15.5% or more for half of the subject pairs, and the spectral difference was reduced by more than half for 9.5% of subject pairs.

One can think of an optimal scale factor as the overall frequency disparity between subjects' DTFs. Figure 8 plots

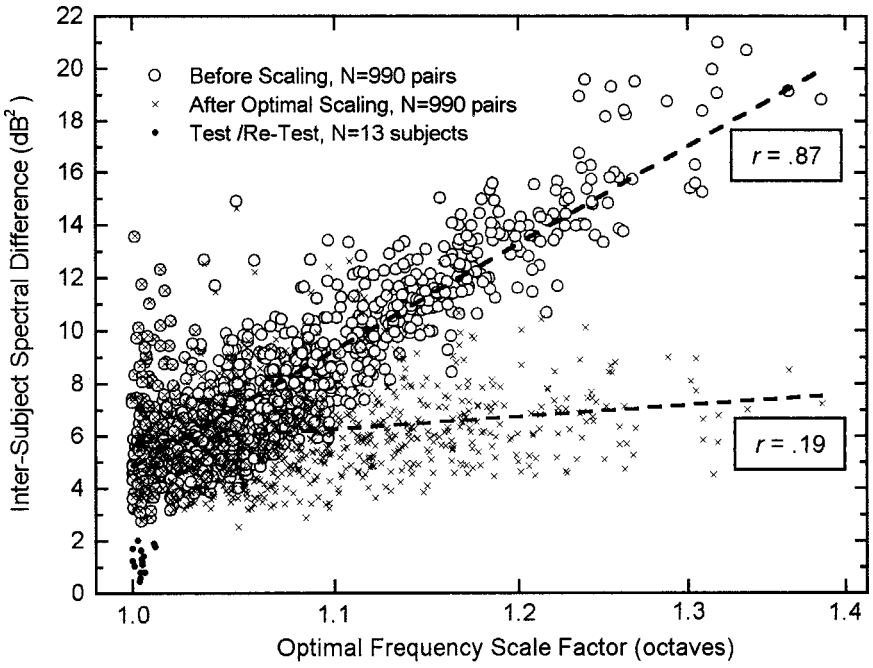


FIG. 8. Inter-subject spectral difference as a function of optimal scale factor. Symbols indicate the inter-subject spectral differences that were measured before (open circles) and after (×'s) scaling DTFs by optimal scale factors. Small filled circles represent spectral differences and optimal scale factors for comparisons of pairs of successive measurements from 13 subjects. Dashed lines are regression lines. $N=990$ pairs.

the inter-subject spectral difference in DTFs, before and after frequency scaling, as a function of the optimal scale factor. Each symbol represents one pair of subjects. Before scaling (open circles), the spectral difference was roughly proportional to the frequency disparity between DTFs. After scaling ($\times$'s), the spectral difference was nearly independent of the magnitude of the optimal scale factor. This indicates that there was a component of the spectral difference that could be eliminated almost entirely by frequency scaling and that remaining components of the spectral difference were largely independent of the overall frequency disparity.

Optimal scale factors computed for subjects' left ears correlated highly with the optimal scale factors computed for their right ears. A correlation coefficient was computed between the base-2 logarithms of optimal frequency scale factors for 990 inter-subject pairs of right ears compared to 990 pairs of left ears. The correlation coefficient was 0.95. Given that observation, one would predict that interaural-level-difference (ILD) spectra would scale in frequency similarly to the DTFs of each ear. To test that prediction, ILD spectra were estimated by filtering right-ear and left-ear DTFs with the bank of band-pass filters, as described above, then subtracting left-ear DTFs in dB from right-ear DTFs in dB. Inter-subject ILD spectral differences for each pair of subjects were computed by subtracting ILD spectra between subjects for each direction, computing the variance of the difference distribution at each direction, then averaging those variances across all directions. The inter-subject ILD spectral differences were measured as a function of inter-subject frequency scale factor. The base-2-logarithms of optimal scale factors computed for ILD spectra correlated closely with those computed for right-ear DTFs ($r=0.96$).

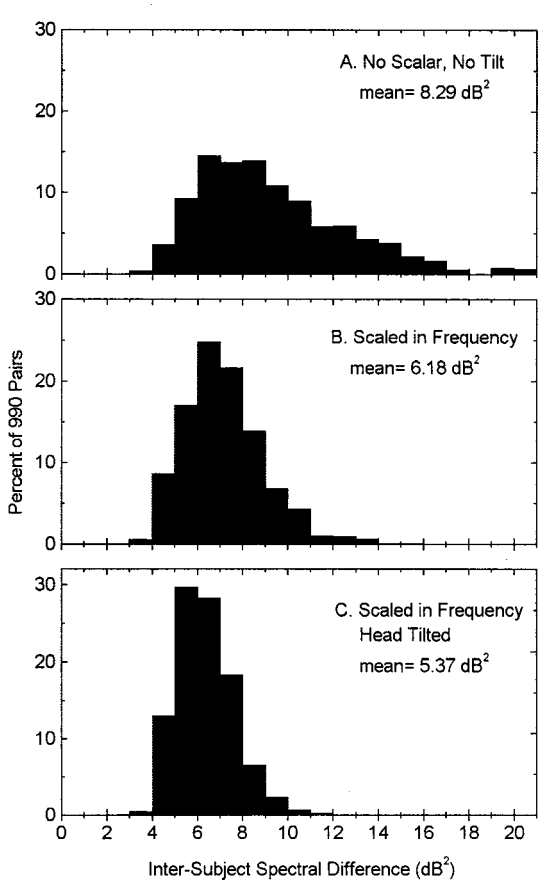


FIG. 9. Distributions of inter-subject spectral differences in DTFs. Distributions of inter-subject spectral differences (A) prior to any manipulation, (B) after scaling in frequency, and (C) after scaling in frequency and introducing relative coordinate-system tilts. $N=990$ subject pairs.

B. Additional components of inter-subject DTF differences

An average of 6.18 dB^2 of inter-subject spectral difference remained after DTFs were scaled in frequency by an optimal factor. Several characteristics of measured DTFs might have accounted for that remaining difference. One possible component was the experimental variability inherent in the recording of DTFs. That variability was estimated by repeating the DTF measurements from 13 of the subjects. In each case, the second set of measurements was made one or more days after the first. The spectral differences and optimal scale factors that were measured for each pair of subjects are shown in Fig. 8 by small filled circles. The average spectral difference was only 1.33 dB^2, which was approximately 20% of the spectral difference between subjects that remained after frequency scaling.

Another possible source of inter-subject differences in DTFs could have resulted from differences in the positions in which subjects held their heads during the DTF measurements and/or in the orientations of subjects' external ears on their heads. The effect of tilting a subject's head about the interaural axis was simulated *post hoc* by rotating the spatial coordinate system that was used to interpolate among DTFs at various locations (see Sec. I D). A relative tilt was introduced between a pair of subjects by rotating one subject's DTFs upward while rotating the other subject's DTFs a symmetrical amount downward. We computed for each pair of subjects the relative tilt that minimized spectral differences in DTFs. Across all 990 pairs of subjects, the unsigned magnitude of the optimal tilt averaged 9.0°, and the head-tilting procedure reduced spectral differences by an average of 13.7%.

Figure 9 shows the distribution of inter-subject spectral differences before manipulating DTFs [Fig. 9(A)], after scaling DTFs in frequency [Fig. 9(B)], and after scaling in frequency and adjusting the tilt [Fig. 9(C)]. The means of the distributions were reduced significantly after each procedure (paired t-test, $N=990$, $p<0.001$). Neither procedure had very much influence on the smallest spectral differences, but both reduced the largest differences. For example, frequency scaling brought the 95th percentile of inter-subject spectral difference from 14.3 to 9.3 dB^2, and subsequent adjustment of relative tilts brought the 95th percentile to 7.6 dB^2.

Visual inspection of DTFs from various subjects gives some insight into the characteristics of the inter-subject differences that remain after frequency scaling and correction for head and ear orientations. Figure 10 shows examples of DTFs from eight subjects for sound-source locations in the frontal midline at elevations of $-60°$ (left column), 0° (middle column), and $+60°$ (right column). In all cases, the coordinate systems were rotated (i.e., the heads were ''tilted'') to minimize spectral differences. The DTFs shown

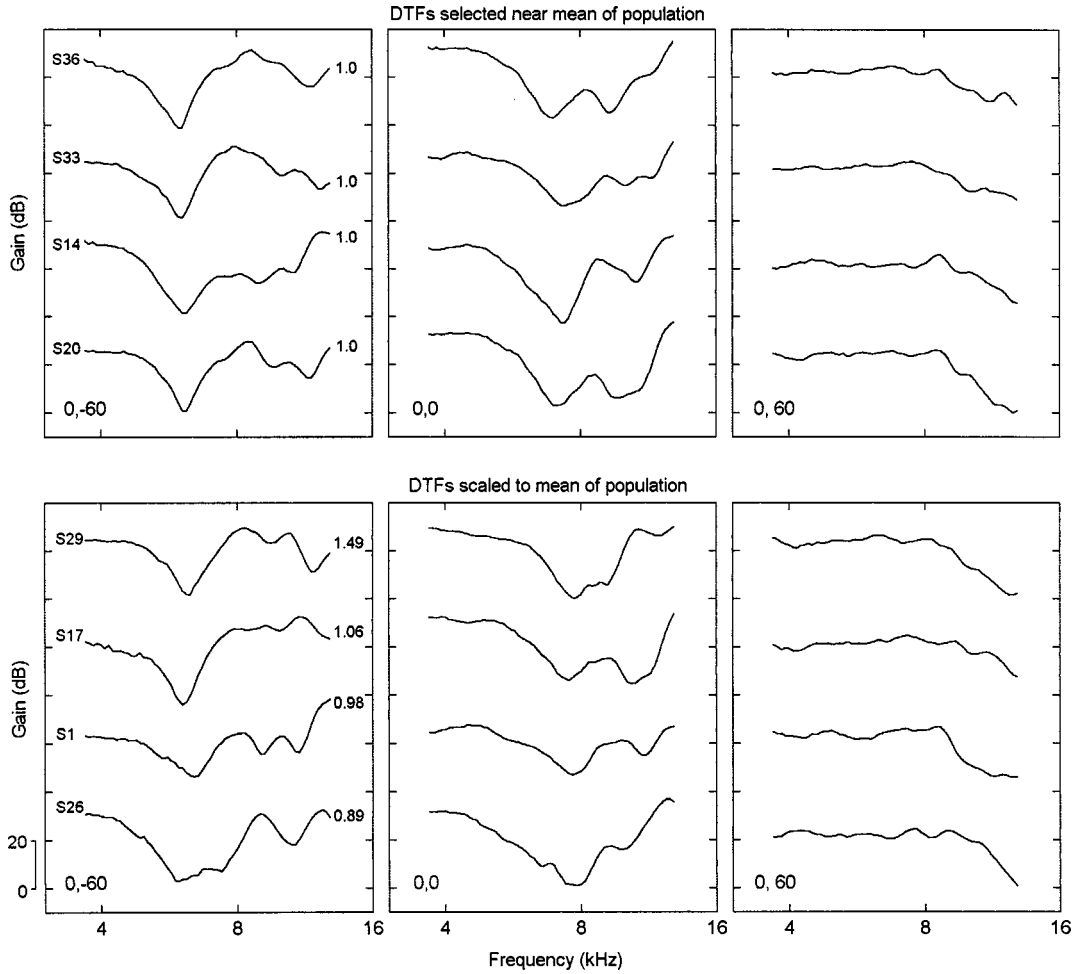


FIG. 10. DTFs from eight subjects. The top row of panels shows DTFs selected from four subjects among whom the optimal scale factors were near one. Those DTFs are illustrated unscaled. The bottom row of panels shows DTFs selected from four subjects that showed a broad range of optimal scale factors. Those DTFs have been scaled to minimize spectral difference with an average subject. The scale factors are shown next to the plots in the lower left panel. The left, middle, and right columns of plots represent DTFs from elevations of −60, 0, and 60° relative to the frontal horizontal plane.

in the upper row of panels were selected because the optimal scale factors among them were near unity. They are shown without any scaling. The DTFs shown in the lower row of panels represent a large range of relative frequency scale factors, and those DTFs are scaled in frequency to minimize the spectral difference among them. The number near each DTF indicates the scale factor that was applied. Although one can see inter-subject differences among the DTFs for each location, it appears that the most prominent spectral features are similar across subjects. In particular, the lowest-frequency spectral notch tends to be well aligned across all subjects. Because the lowest-frequency notch also tended to be the largest spectral feature, one assumes that it tended to dominate the measurements of inter-subject spectral difference and the determination of the optimal scale factor. Higher-frequency spectral features are somewhat less well aligned, and differences among them contribute to residual inter-subject spectral differences. One can see differences in the magnitudes of spectral peaks and notches, which were not influenced directly by any of the manipulations that were employed. Also, there are instances of spectral features that are unique to particular DTFs.

C. Relation of optimal scale factors to interaural delays

The difference in the travel time from a sound source to the two ears introduces an interaural time difference (ITD) that provides an important cue for sound localization. ITDs are produced by some of the same physical structures (i.e., head and external ears) that produce the amplitude component of DTFs. For that reason, one would expect inter-subject differences in ITDs to parallel inter-subject differences in DTFs. ITDs were computed from ''raw'' DTFs that had not been processed with a band-pass filter bank and had not undergone spatial interpolation. For each sound-source direction, a cross spectrum was computed from the complex DTFs from the right and left ears, then an inverse Fourier transform was used to generate the interaural cross- correlogram. The ITD was given by the delay of the peak of the correlogram.

The ITDs for one subject are plotted in Fig. 11(A) as a function of the angle of the sound source relative to the median plane (the *lateral angle*). The ITDs approximate the function that describes ITDs in high-frequency sounds around a spherical head (Woodworth, 1938). The function is

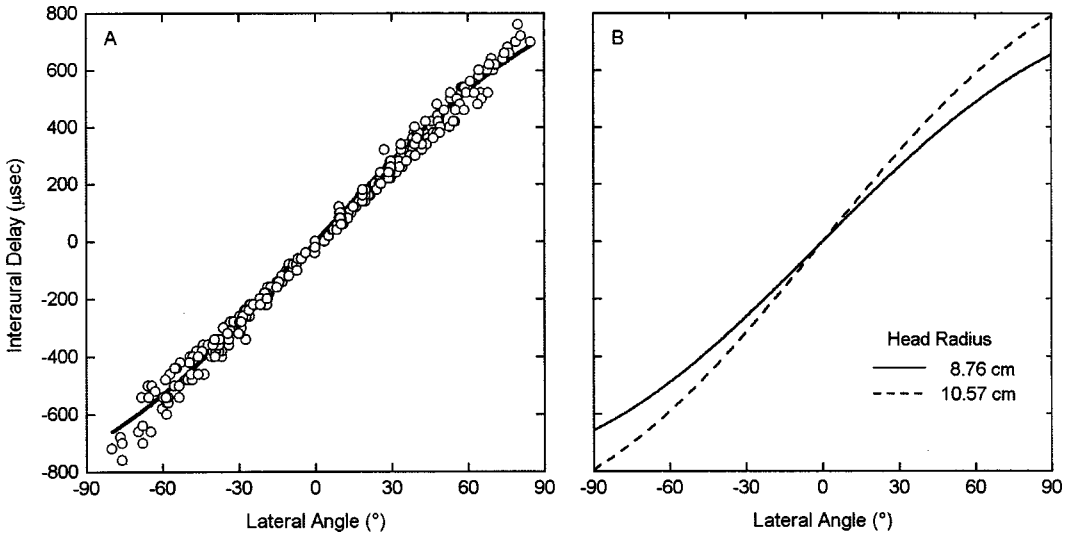


FIG. 11. (A) ITDs from subject S10 are plotted as a function of the angle of the sound source relative to the median plane. Measurements are shown for 400 sound-source directions. The solid line is a fit to the spherical-head ITD expression given in the text. (B) The spherical-head ITD expression is plotted here for two head radii that represent the smallest and largest radii in our sample of subjects.

$\mathrm{ITD} = r/c[\theta + \sin(\theta)]$, where r is the radius of a spherical head, c is the speed of sound, and θ is the lateral angle in radians. The ITDs for each subject were summarized by fitting the data to this spherical-head ITD and adjusting r in the function to minimize the squared error. The function was then evaluated at a lateral angle of 90°. The result is an estimate of the maximum ITD for each subject. Maximum ITDs calculated in the 33 subjects from whom physical dimensions were available (see Sec. II D) ranged from 657 to 792 μs ($709 \pm 32\,\mu$s). Figure 11(B) plots the fitted ITD functions for the subjects who showed the shortest and longest ITDs.

Inter-subject ratios of maximum ITDs were compared with the scale factors that minimized inter-subject differences in DTFs. For the purpose of computing correlation coefficients, inter-subject optimal scale factors and inter-subject ratios of maximum ITDs were expressed as logarithms so that logarithms of ratios and their inverses would be symmetrical on a linear scale. Log ratios of maximum ITD and log optimal scale factors were formed over all 1056 pairwise combinations of 33 subjects, including both subject A/subject B and subject B/subject A. Inter-subject log maximum ITD ratios correlated highly with inter-subject DTF log optimal scale factors ($r = 0.71$).

D. Relation of optimal scale factors and ITDs to physical dimensions

The spectral details of a DTF result from the physical interaction of an incident sound wave with the head and external ears. For that reason, one would expect the frequencies of spectral features to depend on the sizes of the head and ears. A set of physical dimensions was measured from 33 of the subjects; the other 12 subjects were not available for measurements. A micrometer was used to measure distances among the points on the external ear shown in Fig. 12. The measurements included inter-tragal notch to the rim of the helix (*pinna-cavity height*, I to A), *notch to anti-helix* (I to B), *notch to crus* (I to C), *concha depth* (F to G), *tragus to antitragus* (G to H), and *pinna-cavity breadth* (G to E). A caliper was used to measure the *head width* from side to side at the point in front of the tragus that is defined by the condyle of the mandible (J). That point can be identified by palpating in front of the tragus while the subject opens and closes his or her jaw. Finally, as a global measure of each subject's physical size, each subject's *body height* was measured, from bare feet to the top of the head. Table I gives summary statistics of those measurements.

Inter-subject ratios of physical dimensions were compared with inter-subject optimal scale factors. Again, ratios were expressed as logarithms for the purpose of computation of correlation coefficients. Correlations of ratios of physical

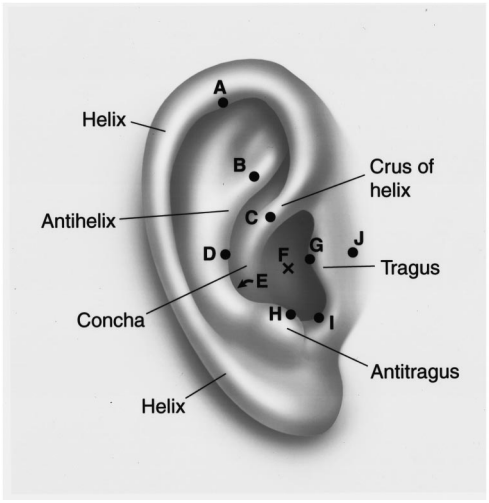


FIG. 12. Drawing of an external ear. Lettered dots indicate the points from which physical measurements were taken. Points A, B, and C are, respectively, the points on the rim of the helix, on the ridge of the anti-helix, and on the ridge of the crus that maximized the distance to the bottom of the inter-tragal notch (point I). Point E is the inside surface of the back of the concha, and point F is directly medial to the tip of the tragus (point G), superficial to the ear canal entrance.

TABLE I. Physical measurements (33 subjects).

	Minimum	Mean	Maximum	Std. Deviation
Pinna-cavity height (mm)	36.1	43.9	56.1	4.8
Notch to anti-helix (mm)	21.6	26.6	32.5	2.8
Notch to crus (mm)	11.5	16.4	21.4	2.4
Concha depth (mm)	6.5	10.4	14.7	2.0
Tragus to anti-tragus (mm)	3.0	6.8	12.9	2.1
Pinna-cavity breadth (mm)	8.2	11.9	16.7	1.9
Head width (mm)	122	134	153	8
Body height (mm)	1520	1690	1930	102
Maximum interaural delay (μs)	657	709	792	32
Optimal scale factor (1056 subject pairs)	0.72		1.38	

dimensions with ratios of optimal scale factors were computed over all 1056 pairs of 33 subjects. Table II shows the correlation coefficients. All of the correlations in the table are significant at least at the $P<0.01$ level. The physical dimensions that correlated most strongly with the inter-subject optimal scale factor were the pinna-cavity height ($r=0.72$) and the head width ($r=0.67$). The correlation of optimal frequency scale factors with all other dimensions that were measured from the pinna was $r<=0.41$. Correlations among physical dimension were computed directly from the dimensions in 33 subjects (i.e., not the logarithms of inter-subject ratios). The greatest correlation in Table I, $r=0.85$, was between head width and maximum ITD. That is consistent with our previous report of a small sample of subjects that showed a correlation of $r=0.91$ between head width and interaural envelope delay (Middlebrooks and Green, 1990). Body height also correlated highly with maximum ITD ($r=0.67$), but that probably reflects the fact that body height correlated with head width itself ($r=0.67$).

The set of optimal scale factors between each subject and every other subject could be summarized by the geometric mean of the optimal scale factors. Geometric means of optimal scale factors ranged from 0.833 to 1.143. Female subjects tended to have higher mean scale factors than males (i.e., spectral features lay at higher frequencies), but there was no indication of a gender effect that could not be accounted for by the tendency of females to be smaller than males. There was no conspicuous trend in scale factors detected across our relatively small sample of various racial groups.

Inter-subject optimal scale factors could be estimated with some accuracy from ratios of certain physical dimensions. The pinna-cavity height and the head width both showed strong correlations with optimal scale factors. Those two measurements showed a correlation of only $r=0.47$ with each other, so a combination of those two measurements should provide a better estimate than either one alone. A linear regression was performed on inter-subject log ratios of the two terms and the resulting regression equation was used to estimate log optimal scale factors. For the present data set, the regression equation to scale subject A to subject B using pinna-cavity height (pinna) and head width (head) was:

$$\log_2(\text{optimal scale factor}) = 0.340^*\log_2(\text{pinna}_A/\text{pinna}_B) + 0.527^*\log_2(\text{head}_A/\text{head}_B). \quad (1)$$

The sign convention was such that if subject A were larger than subject B, subject A's DTFs would be scaled upward in frequency to align with the DTFs of subject B. The scale factors estimated by this equation correlated highly with acoustically measured optimal scale factors ($r=0.82$, $N=1056$ subject pair). The rms difference between log optimal scale factors estimated in this way and logarithms of factors computed from acoustical measurements was 0.0813 octave, which corresponds to a factor of 1.058. For comparison, optimal scale factors in this set of 1056 subject pairs ranged between 0.72 and 1.38. Thus knowledge of the physical dimensions of the head and external ears of a pair of subjects substantially reduced the uncertainty in their optimal scale factor.

III. DISCUSSION

The results confirm, with some reservations, the two working hypotheses of the study. The first hypothesis speculated that DTFs differ systematically among subjects in regard to the position of spectral features along a frequency axis. Scaling of DTFs in frequency more than halved the spectral difference in 9.5% of subject pairs and produced a substantial, though lesser, reduction in spectral difference among the majority of other subject pairs. Some 20% of pairs of subjects showed little or no reduction in inter-subject spectral difference in DTFs, but those pairs involved subjects who showed relatively little spectral difference at the outset. A residue of spectral difference, averaging 6.2 dB^2 across all subject pairs, remained after scaling by an optimal factor. Possible components of that residual spectral difference included variability in experimental measurements and inter-subject differences in the orientations of subjects' heads and ears. The DTFs from individual subjects also contained direction-specific idiosyncrasies that could not be resolved by a global scale factor. Such idiosyncrasies included the

TABLE II. Correlation coefficients (r).

	Optimal scale factor (1056 pairs)	Maximum ITD (33 subjects)	Body height (33 subjects)	Head width (33 subjects)	Pinna-cavity height (33 subjects)
Maximum ITD	0.71				
Body height	0.59	0.67			
Head width	0.67	0.85	0.67		
Pinna-cavity height	0.72	0.54	0.51	0.47	
Pinna height to crus	0.41	0.43	0.58	0.49	0.48

presence or absence of particular spectral features and inter-subject differences in the magnitudes of such features. One's impression from visual inspection of DTF plots, however, is that frequency scaling eliminated the greatest, perhaps the only, systematic difference among individual subjects' DTFs.

The second working hypothesis speculated that the frequency difference between the DTFs from two subjects could be predicted by the relative physical sizes of the two subjects. Physical measurements of subjects identified several dimensions that correlated significantly with optimal scale factors and with maximum ITDs. Given measurements of pinna-cavity height and head width from a pair of subjects, it was possible to estimate the optimal scale factor between those subjects to within an rms error of 0.0813 octave, a factor of 1.058. That is less than three times the smallest increment in frequency scale factors that was tested. Some of the correlations that were observed among physical dimensions and optimal scale factors are likely to reflect a causal relationship. For instance, mathematical models of ITDs produced by a spherical head incorporate head radius as a parameter (Woodworth, 1938; Kuhn, 1977). Similarly, Rabinowitz and colleagues (1993) modeled the DTFs that would be produced by diffraction on the surface of a sphere. In the far-field condition (like that approximated in the present study), the DTF scaled linearly in frequency by $1/\beta$ when the sphere radius was scaled by β. Several investigators have modeled the directionality of the pinna in terms of a system of parallel delays and summations (e.g., Batteau, 1967; Hebrank and Wright, 1974; Watkins, 1978). As the size of the pinna increases, delays within the pinna increase, and notches and peaks in the delay-and-add spectra move downward in frequency. In contrast to these causally related parameters, the relative body heights of subjects probably have no direct influence on their optimal scale factors. Nevertheless, to the degree that the dimensions of subjects' heads and ears are proportional to subjects' overall physical sizes, ratios of subjects' heights might provide a rough estimate of their optimal scale factors.

A. Relation to previous results

Several previous reports have commented on inter-subject differences in the frequencies of spectral features in DTFs. Shaw (1966, 1982) illustrated DTFs from ten listeners. He noted that the ''relative positions of the individual curves [in frequency]... are independent of the angle of incidence'' (Shaw, 1966, p. 469). This observation is in keeping with the notion that DTFs are similar among subjects but tend to be offset in frequency. Shaw (1982) also noted inter-subject differences in the number of spectral ''minima,'' the elevation dependence of the frequencies of minima, and the tendency of minima to cluster along the frequency axis. Those are the types of individual differences that were not influenced by any of the manipulations in the present study. Mehrgardt and Mellert (1977) were concerned with preserving spectral detail when averaging DTFs among 20 subjects. They observed that DTFs from various individual subjects for particular directions overlapped more closely when they were scaled in frequency by as much as 8%. A scale factor of 8% is slightly larger than the median value of optimal scale factors, 6%, in the present study. Mehrgardt and Mellert optimized frequency scale factors on a direction-by-direction basis, whereas the procedure in the present study attempted to find a single scale factor that would minimize inter-subject spectral differences across all source directions.

In a previous study (Middlebrooks *et al.*, 1989), we described the directional sensitivity of amplitude gain of the ear at a several individual frequencies. Directional sensitivity was represented in that report primarily in the form of frequency-specific plots of sound amplitude in the ear canal as a function of two-dimensional sound-source direction. We observed great similarity in directional sensitivity among subjects, but noted that the frequency at which we observed a particular directional pattern tended to vary among subjects. That is, the directional pattern measured at 8 kHz for subject A might correspond more closely to the pattern measured at 8.9 kHz for subject B than to subject B's pattern measured at 8 kHz. We computed, in effect, the optimal scale factor among 5 subjects at 21 reference frequencies across 356 sound-source directions. There was a general tendency, with some variation, for frequency scale factors to be constant across reference frequencies. Moreover, in that study, scale factors tended to rank according to the physical sizes of subjects. In another related study, we analyzed DTFs using principal components analysis (Middlebrooks and Green, 1992). When subjects were ranked according to body height, features of the principal-component basis vectors computed from DTFs of the shorter subjects lay at higher frequencies than those of the taller subjects.

B. Significance of frequency scaling for virtual localization

The DTFs measured from a listener can be used to filter sounds that are presented to the listener through headphones (Wightman and Kistler, 1989b). In the most successful cases, a realistic virtual acoustic environment results, in which acoustic targets appear to occupy well-defined positions in the space surrounding the listener. The main practical motivation of the current study was to develop a means of tailoring one or a few generic sets of DTFs to specific listeners. Scaling DTFs in frequency might accomplish a great part of that goal, and the companion paper reports the results of tests of the perceptual consequences of frequency scaling (Middlebrooks, 1999). Considered here are some of the implications of the magnitudes of optimal scale factors, the significance of head position in measurements of DTFs, and the utility of estimation of optimal scale factors from physical dimensions.

It is instructive to compare the magnitudes of optimal scale factors with perceptual measures of frequency sensitivity. The median optimal scale factor in the present sample was 1.061, and the largest optimal scale factor between any pair of subjects was 1.38. A factor of 1.061 corresponds to 0.086 octave, which is slightly larger than a musical semitone. That value is roughly half the equivalent rectangular bandwidth estimated for an auditory filter around 7 kHz (Moore and Glasberg, 1983). Moore and colleagues (1989)

found that frequency-discrimination thresholds around 8 kHz were of the order of .015 octave for tones or band-pass noise, but that frequency-discrimination thresholds for a 9-dB peak or notch imposed on a broadband base were of the order of 0.08 to 0.10 octave. That means that, if two subjects were to be chosen randomly, there would be a roughly 50% chance that individual spectral features in the DTFs from one subject would fall a detectable frequency increment away from the corresponding features in the other subject's DTFs.

The spectral difference between sets of DTFs from a given pair of subjects often could be reduced by rotating about the interaural axis the spatial coordinate system of one set of DTFs relative to the other. That result suggests either that subjects held their heads at slightly different angles during the measurements or that subjects' external ears are positioned on their heads at somewhat different angles. That variability in head and/or ear position complicates inter-subject comparisons of DTFs. Nevertheless, we are inclined to regard the inter-subject variance in head orientation as a nuisance factor that can be largely ignored for the purposes of virtual localization testing. In a psychophysical localization task there can be considerable variation in nominally upright head position even in a single subject. For instance, when subjects were asked to orient their heads toward a continuous broadband noise source located near the horizontal plane, the head orientations by individual subjects varied with standard deviations of about 3° (Makous and Middlebrooks, 1990; also observed in the companion study). In a situation in which a subject localized using DTFs that were recorded from a head that was tilted downward by 10°, one might expect to see a 10° vertical offset in localization judgements. That offset would be small, however, compared to the front/back confusions that often are observed when a subject listens through another subject's DTFs.

The observation that the optimal scale factor between a pair of subjects can be estimated from physical measurements of their pinna-cavity heights and head widths offers the hope that a generic set of DTFs might be tailored to fit any subject. Estimates of frequency scale factors based on physical measurements [Eq. (1)] differed from scale factors computed from acoustical measurements by an rms error of 5.8%. The magnitude of that error can be appreciated by referring to the plot of inter-subject spectral difference versus frequency scale factor shown in Fig. 5. A scaling up or down by 2 products of 1.06 (scale factors of 0.89 to 1.12) encompassed 95% of errors in estimates (as measured from the error distribution). That range encompassed fewer than six sample points on either side of the minimum value in the figure. In the illustrated example, frequency scaling by any factor in that range reduced the spectral difference by more than 40% or more. It might be the case that a optimal scale factor estimated only from physical dimensions is adequate to produce considerable reduction in inter-subject differences in DTFs. Alternatively, an optimal scale factor estimated from physical dimensions could be used as a starting point for an adaptive psychophysical procedure that would seek an optimal scale factor or that would select from among a library of sets of DTFs. The companion paper explores the perceptual consequences of scaling nonindividualized DTFs and demonstrates the sensitivity of localization performance to changes in the values of frequency scale factors.

ACKNOWLEDGMENTS

I thank Zekiye Onsan for expert technical assistance. Li Xu helped to record the DTFs. Shigeto Furukawa, Ewan Macpherson, Brian Mickey, and Li Xu provided stimulating discussion and commented on an early draft of the manuscript. I thank Professors Edgar Shaw, Frederic Wightman, and a third anonymous reviewer for their constructive criticisms. This research was supported by ONR Grant No. N00014-94-1-0481 and NIH Grant No. DC000420.

Batteau, D. W. (**1967**). ''The role of the pinna in human localization: Theoretical and physiological consequences,'' Proc. R. Soc. London, Ser. B **168**, 158–180.

Blauert, J. (**1969–1970**). ''Sound localization in the median plane,'' Acustica **22**, 205–213.

Bronkhorst, A. W. (**1995**). ''Localization of real and virtual sound sources,'' J. Acoust. Soc. Am. **98**, 2542–2553.

Butler, R. A. (**1987**). ''An analysis of the monaural displacement of sound in space,'' Percept. Psychophys. **41**, 1–7.

Foster, S. (**1986**). ''Impulse response measurement using Golay codes,'' in *IEEE 1986 Conference on Acoustics, Speech, and Signal Processing, Vol. 2* (IEEE, New York), pp. 929–932.

Golay, M. J. E. (**1961**). ''Complementary series,'' IRE Trans. Inf. Theory **7**, 82–87.

Hartmann, W. M., and Wittenberg, A. (**1996**). ''On the externalization of sound images,'' J. Acoust. Soc. Am. **99**, 3678–3688.

Hebrank, J., and Wright, D. (**1974**). ''Spectral cues used in the localization of sound sources on the median plane,'' J. Acoust. Soc. Am. **56**, 1829–1834.

Kuhn, G. F. (**1977**). ''Model for the interaural time differences in the azimuthal plane,'' J. Acoust. Soc. Am. **62**, 157–167.

Makous, J. C., and Middlebrooks, J. C. (**1990**). ''Two-dimensional sound localization by human listeners,'' J. Acoust. Soc. Am. **87**, 2188–2200.

Mehrgardt, S., and Mellert, V. (**1977**). ''Transformation characteristics of the external human ear,'' J. Acoust. Soc. Am. **61**, 1567–1576.

Middlebrooks, J. C. (**1992**). ''Narrow-band sound localization related to external ear acoustics,'' J. Acoust. Soc. Am. **92**, 2607–2624.

Middlebrooks, J. C. (**1999**). ''Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency,'' J. Acoust. Soc. Am. **106**, 1493–1510.

Middlebrooks, J. C., and Green, D. M. (**1990**). ''Directional dependence of interaural envelope delays,'' J. Acoust. Soc. Am. **87**, 2149–2162.

Middlebrooks, J. C., and Green, D. M. (**1991**). ''Sound localization by human listeners,'' Annu. Rev. Psychol. **42**, 135–159.

Middlebrooks, J. C., and Green, D. M. (**1992**). ''Observations on a principal components analysis of head-related transfer functions,'' J. Acoust. Soc. Am. **92**, 597–599.

Middlebrooks, J. C., Makous, J. C., and Green, D. M. (**1989**). ''Directional sensitivity of sound-pressure levels in the human ear canal,'' J. Acoust. Soc. Am. **86**, 89–108.

Møller, H., Sorensen, M. F., Jensen, C. B., and Hammershoi, D. (**1996**). ''Binaural technique: Do we need individual recordings?,'' J. Aud. Eng. Soc. **44**, 451–469.

Moore, B. C. J., and Glasberg, B. R. (**1983**). ''Suggested formulae for calculating auditory-filter bandwidths and excitation patterns,'' J. Acoust. Soc. Am. **74**, 750–753.

Moore, B. C. J., Oldfield, S. R., and Dooley, G. J. (**1989**). ''Detection and discrimination of spectral peaks and notches at 1 and 8 kHz,'' J. Acoust. Soc. Am. **85**, 820–836.

Morimoto, M., and Ando, Y. (**1980**). ''On the simulation of sound localization,'' J. Acoust. Soc. Jpn. (E) **1**, 167–174.

Oppenheim, A. V., and Schafer, R. W. (**1975**). *Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ).

Rabinowitz, W. R. *et al.* (**1993**). ''Sound localization cues for a magnified head: Implications from sound diffraction about a rigid sphere,'' Presence **2**, 125–129.
Shaw, E. A. G. (**1966**). ''Earcanal pressure generated by a free sound field,'' J. Acoust. Soc. Am. **39**, 465–470.
Shaw, E. A. G. (**1974**). ''Transformation of sound pressure level from the free field to the eardrum in the horizontal plane,'' J. Acoust. Soc. Am. **56**, 1848–1861.
Shaw, E. A. G. (**1982**). ''External ear response and sound localization,'' in *Localization of Sound: Theory and Applications* (Amphora, Groton, CT), pp. 30–41.
Watkins, A. J. (**1978**). ''Psychoacoustical aspects of synthesized vertical locale cues,'' J. Acoust. Soc. Am. **63**, 1152–1165.
Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. (**1993**). ''Localization using nonindividualized head-related transfer functions,'' J. Acoust. Soc. Am. **94**, 111–123.
Wightman, F. L., and Kistler, D. J. (**1989a**). ''Headphone simulation of free field listening. I: Stimulus synthesis,'' J. Acoust. Soc. Am. **85**, 858–867.
Wightman, F. L., and Kistler, D. J. (**1989b**). ''Headphone simulation of free-field listening. II: Psychophysical validation,'' J. Acoust. Soc. Am. **85**, 868–878.
Woodworth, R. S. (**1938**). *Experimental Psychology* (Holt, Rinehart, and Winston, New York), pp. 349–361.
Zhou, B., Green, D. M., and Middlebrooks, J. C. (**1992**). ''Characterization of external ear impulse responses using Golay codes,'' J. Acoust. Soc. Am. **92**, 1169–1171.

Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency

John C. Middlebrooks
Kresge Hearing Research Institute, University of Michigan, 1301 East Ann Street, Ann Arbor, Michigan 48109-0506

(Received 1 September 1998; revised 11 March 1999; accepted 26 May 1999)

This study examined virtual sound localization in three conditions that differed according to the directional transfer functions (DTFs) that were used to synthesize the virtual targets. The *own-ear* and *other-ear* conditions used DTFs measured from listeners' own ears and those measured from other subjects, respectively. The *scaled-ear* condition employed other-ear DTFs that were scaled in frequency to minimize the mismatch between spectral features in the listener's and the other subject's DTFs. All measures of localization error typically were lowest in the own-ear condition. In other-ear conditions, all error measures tended to increase in proportion to the inter-subject differences in DTFs. When spectral features in an other-ear set of DTFs fell systematically lower in frequency than in a listener's own DTFs, low frontal targets typically were reported as low in the rear, and high rear targets were reported as high in front. When spectral features in a set of DTFs fell systematically higher in frequency than in a listener's own DTFs, elevation judgements showed an upward bias. In the scaled-ear condition, all measures of performance tended to improve relative to the other-ear condition. In the majority of cases, frequency scaling more than halved the penalty for use of another subject's DTFs.
[S0001-4966(99)01509-X]

PACS numbers: 43.66.Pn, 43.66.Qp [DWG]

INTRODUCTION

One can capture the transfer functions of the head and external ears by presenting a broadband sound in a free field while recording from within a subject's ear canals with miniature microphones. Transfer functions measured in that way have come to be known as ''head-related transfer function'' (HRTFs). When HRTFs are applied to an arbitrary signal and presented to the listener's two ears through headphones, he or she hears a ''virtual target'' that appears to originate from the location of the original sound source (e.g., Wightman and Kistler, 1989b; Bronkhorst, 1995; Møller *et al.*, 1996). In a similar way, one can compile a set of transfer functions for a wide range of sound-source locations and can synthesize a virtual acoustic environment. When listening through HRTFs measured from his or her own ears, a listener reports auditory events that are spatially compact and that appear ''externalized,'' i.e., that seem to arise from sources outside of the listener's head. When listening through HRTFs measured from another subject (i.e., ''nonindividualized'' HRTFs), listeners often complain that auditory events are spatially diffuse, and listeners often make incorrect judgments of the source locations (Wenzel *et al.*, 1993; Møller *et al.*, 1996). The primary goal of the current study was to explore procedures by which a generic set of HRTFs could be customized for use by any arbitrary listener.

The companion report (Middlebrooks, 1999) examined inter-subject differences in the directional components of subjects' HRTFs, which are referred to as *directional transfer functions* (DTFs). DTFs varied systematically among subjects in regard to the positions of spectral features along the frequency axis. Spectral features in the DTFs of one subject could be aligned with those of another subject by scaling his or her DTFs in frequency. To a first approximation, DTFs scaled in proportion to the sizes of subjects' heads and external ears, so that characteristic spectral features of a large subject's DTFs tended to lie at lower frequencies than those of a smaller subject. The present report compares virtual-localization performance under conditions in which virtual targets were synthesized from DTFs measured from the listeners' own ears and from the ears of other subjects. Subjects showed characteristic errors in localization that varied according to frequency differences between their own DTFs and the DTFs through which they listened. Localization performance with other subjects' DTFs improved substantially when inter-subject differences in DTFs were reduced by scaling in frequency.

I. METHODS

A. Subjects

The subject population consisted of 18 paid listeners (11 female and 7 male) drawn from the subject population of the companion study (Middlebrooks, 1999). Ages ranged from 22 to 43 yr (30.4 ± 5.7, mean $\pm$ standard devation), and body heights ranged from 155 to 193 cm (169.7 ± 10.2). Fourteen of the listeners (eight female and six male), including the author, participated in the main blocks of free-field and virtual localization trials. The remaining four listeners participated only as ''naïve'' listeners in tests of virtual localization with varying frequency scale factors. All subjects were screened by conventional tests for hearing thresholds within 20 dB of audiometric zero. The DTFs were obtained in the companion study. They included measurements from 14 of the behavioral subjects plus one additional subject.

B. Apparatus, sound generation, and coordinate system

All behavioral tests were conducted in the double-walled anechoic chamber that was described in the companion paper. Free-field stimulation experiments employed a mechanism that could position two loudspeakers on the surface of an imaginary sphere that was 1.2 m in radius. The loudspeakers were Pioneer model TS-879 two-way coaxials. Virtual-environment stimuli were presented through Sennheiser model 265 circumaural headphones. Audio signals were generated with an Intel-based microcomputer, using a 16-bit analog interface and a digital-signal-processor board (Tucker-Davis-Technologies). Signals were generated with 16-bit precision at a sample rate of 50 kHz.

Free-field stimuli were presented from the moveable loudspeakers. The loudspeakers were calibrated using a $\frac{1}{2}$-in. precision microphone (ACO Pacific) that was positioned in the center of the sphere described by the speaker-movement mechanism. Probe stimuli for the calibration were pairs of 16 384-point Golay codes (Golay, 1961; Foster, 1986; Zhou *et al.*, 1992). An inverse transfer function was computed and stored for each loudspeaker. Virtual stimuli were synthesized using DTFs measured in the companion study. Complex DTFs were transformed to the time domain using an inverse fast Fourier transform (inverse FFT). The resulting impulse responses were truncated to 256 points (i.e., 5.12 ms).

Free-field and virtual stimuli consisted of independent samples of Gaussian noise bursts. The Gaussian samples had Rayleigh-distributed amplitude spectra that were flat on average but differed in detail from trial to trial. Each free-field stimulus was generated by drawing a Gaussian sample, filtering it with the inverse transfer function for the appropriate loudspeaker, and band-pass filtering between 300 and 16 500 Hz. Each virtual stimulus was generated by convolving a Gaussian sample with left- and right-ear impulse responses for the desired target. The headphones were corrected by the manufacturer for a ''diffuse field response.'' No additional headphone correction was applied. Bandwidths of virtual stimuli were limited only by the bandwidth of the headphones, which the manufacturer specified as 10–30 000 Hz. Both types of stimulus were windowed to durations of 250 ms, including 5-ms raised-cosine onset and offset ramps.

One technique that commonly is used for synthesizing virtual targets utilizes HRTFs derived from acoustical measurements at the tympanic membrane (Wightman and Kistler, 1989a). In that technique, one must also measure the transfer function from headphones to the tympanic membranes and deconvolve the headphone-to-tympanic membrane transfer functions from the free-field transfer function. Headphone-to-tympanic-membrane transfer functions vary appreciably among listeners, so individually measured headphone transfer functions should be used (Pralong and Carlile, 1996). The present study used an alternative approach. The transfer functions used in this study were DTFs, which are HRTFs that are processed to isolate the component that is specific to sound-source location (Middlebrooks and Green, 1990; Middlebrooks, 1999). DTFs are obtained from HRTFs by dividing by a listener-specific common component that is computed from an average across the set of all HRTFs. That procedure replaces the separate measurement of the headphone-to-tympanic membrane transfer function.

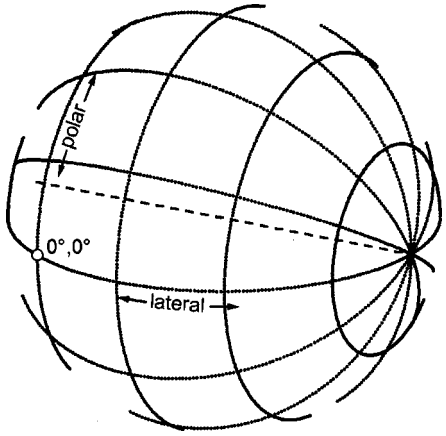


FIG. 1. Horizontal polar coordinate system.

For certain stimulus conditions, DTFs were scaled in frequency. Frequency scaling was accomplished by changing the output rate of the digital-to-analog converter (DAC). The DTFs were collected at a sample rate of 50 kHz, so the base output rate of the DAC was 50 kHz (i.e., a period of 20 μs). The maximum range of scale factors that was used was 0.757–1.320, so the output rate of the DAC was varied from 37 850 to 66 000 Hz. The total number of points in stimulus waveforms was adjusted to maintain a constant overall stimulus duration and rise/fall time.

The locations of free-field and virtual stimuli and the locations of listeners' location judgements are specified in a horizontal polar coordinate system (Fig. 1). This coordinate system is equivalent to one proposed by Morimoto and Aokata (1984). The horizontal position is given by the *lateral angle*, which is the angle formed by the stimulus, the center of the listener's head, and the vertical midline. Lateral angle ranges from left to right 90°, with 0° corresponding to the vertical midline. The vertical and front/back position is given by the *polar angle*. This is the angle around the horizontal pole that coincides with the listener's interaural axis. The polar angle ranges from −90° (beneath the interaural axis), through 0° (on the frontal horizon), through +90° (above the interaural axis), through 180° (on the rear horizon), to 270° (beneath the interaural axis). The advantage of this coordinate system is that the two coordinates correspond to the dimensions signaled primarily by two types of localization cues. The lateral angle is signaled primarily by interaural difference cues. The polar angle unifies front/back and up/down dimensions, both of which appear to be signaled by spectral shape cues. A contour of varying polar angle at a constant lateral angle corresponds to a ''cone of confusion'' (Woodworth, 1938). A disadvantage of the use of polar angle is that the spherical angle that corresponds to a given polar angle is highly compressed at large lateral angles near the poles. For instance, at a lateral angle of 85°, the angular distance between polar angles 0° and 180° is only 10° when measured at the center of the listener's head. For that reason, analysis of the polar dimension of localization judgements was restricted to stimulus locations within 30° of the vertical midline, which corresponded to half of the area of the coor-

dinate sphere. Within that range, spherical angles were compressed by a factor of no more than 0.866 (i.e., the cosine of 30°). By excluding lateral locations from the computation of polar error, front/back confusions probably were undercounted, since localization errors of that type are more prevalent for lateral stimuli (Wightman and Kistler, 1989b; Makous and Middlebrooks, 1990; Carlile *et al.*, 1997). Nevertheless, that undercount should not have influenced the comparison of various virtual localization conditions.

C. Behavioral protocol

The behavioral protocol was similar to that used in our previous studies (Makous and Middlebrooks, 1990; Middlebrooks, 1992). The listener stood on a platform that could be adjusted in height to center his or her head in the imaginary sphere. Trials were conducted in complete darkness. The listener reported the apparent location of a free-field or virtual target by orienting to face the location. Each listener was instructed to ''point with your nose.'' The head orientation was monitored with an electromagnetic device (Polhemous FASTRAK). The sensor for the FASTRAK was attached to a cloth baseball cap from which the visor had been removed.

In free-field trials, the loudspeaker was positioned prior to the beginning of each trial. The beginning of the trial was signaled by the onset of a broadband noise from a loudspeaker that was fixed on the wall at coordinates 0°, 0°. The listener oriented toward the fixed source, then pressed a response key. The broadband noise was terminated when the response key was pressed. The test stimulus was presented after a 600- to 1200-ms delay. The listener oriented toward the apparent position of the test stimulus, then pressed the response key. The computer recorded the listener's head orientation at the time of both key presses. The 250-ms stimulus was sufficiently short that the listener did not move during the presentation of the sound.

The protocol for virtual localization trials was identical to that for the free-field trials with the exceptions that listeners wore headphones and that the initial centering stimulus consisted of a light from a light-emitting diode (LED) rather than a noise. There was concern that some of the virtual targets might seem poorly externalized or spatially diffuse such that the listener could not decide on a single apparent location. For that reason, listeners were provided with an alternate response key that they could press to identify such situations. The use of the alternate key varied widely, but across all listeners and conditions it was used on an average of 9% of trials. Trials on which the alternate key was pressed were eliminated from further analysis.

To quantify stimulus sensation levels, detection thresholds were measured for free-field targets and for each set of DTFs and each frequency scale factor with which a listener was tested. Targets were at free-field or virtual locations 0°, 0°. The sensation level produced by a source of any particular pressure level varied according to the location of the free-field or virtual source. For instance, because of shadowing of sound by the head and external ears, a source at a given level would result in a lower sensation level when presented from behind the listener than when presented from in front. This variation in sensation level is a natural localization cue but is most informative to a listener when the sound-source level is constant. To discourage listeners from relying on absolute level cues for localization, stimulus levels were varied randomly in steps of 1 dB from 40 to 60 dB above the threshold measured at 0°, 0°. Other than varying overall levels, however, no effort was made to specifically invalidate the absolute level cue by, for instance, increasing the level of stimuli presented from the rear.

Fourteen listeners participated in the main blocks of free-field and virtual localization tests. Those listeners were trained as follows: First, they were introduced to the test chamber and the speaker-movement mechanism. They were instructed to orient to targets with full head movements and to overcome the natural tendency to orient with a combination of head and eye movements. They completed two short sessions of free-field localization trials in which the room lights were left on so that the loudspeakers were visible. After each session, they were shown their data and were given verbal feedback on their orienting accuracy. Next, they completed two sessions of trials in which the room lights were turned off and in which feedback was provided immediately after each response by an LED that was illuminated at the center of the loudspeaker cone. Finally, they completed two full sessions of free-field trials in the dark with no immediate feedback. Listeners who persisted in orienting with eye movements rather than head movements received additional instruction and a few more practice sessions. After free-field training, data were collected in three free-field test sessions. Next, listeners received four virtual-localization training sessions in which virtual targets were synthesized from DTFs measured from each listener's own ears. Finally, they completed one or more blocks of seven virtual-localization test sessions. The number of blocks completed by each listener was determined by his or her willingness to continue participation.

The sets of free-field and virtual target locations were drawn from the set of 400 stimulus locations for which DTFs were measured in the companion study. The two locations more than 60° below or more than 79° above the horizontal plane were excluded because it was physically difficult for listeners to orient their heads to those locations. One set of 300 locations was drawn with replacement and was used for the free-field condition for all listeners. A second set of 300 was used for all virtual conditions for all listeners. One hundred forty-four locations fell within 30° of the vertical midline and, thus, were used for analysis of localization in the polar dimension. A block of free-field localization trials, completed in 3 sessions, consisted of 1 trial at each of 300 locations in a set. Listeners completed sessions of 100 free-field trials in 12 to 18 min. In free-field trials, the mechanism that positioned the loudspeaker made an audible noise. That noise always came from the location of the motors, straight overhead, so it did not give a cue to the test stimulus location. Nevertheless, the duration of the motor noise might have given a cue to the distance between successive stimulus locations. For that reason, the test locations were ordered so that the excursion of the speaker mechanism from one position to the next always fell in the range of 60°–120° to the left or right and 30°–70° up or down. This resulted in

speaker excursions that all were around 5 s in duration. Cues to the stimulus location were further confounded by the fact that on any trial the stimulus could be presented from the front or the rear loudspeaker. Because of the special ordering of stimuli, the range of possible locations that could follow any particular stimulus was reduced by 1/3 in azimuth and by 1/2 in elevation. A listener who understood this might have realized that no two stimuli were presented from the same location on successive trials, but aside from that it is difficult to imagine that the ordering of locations had appreciable influence in the free-field localization results.

Virtual localization was tested in three conditions that differed only in the source of the DTFs from which virtual targets were synthesized. In the *own-ear* condition, the DTFs were those recorded from each listener's own ears. In the *other-ear* condition, the DTFs were those recorded from the ears of another subject. In the *scaled-ear* condition, DTFs recorded from the ears of another subject were scaled in frequency to align them with the listener's own DTFs. Trials from three conditions were interleaved within a single block of 900 trials. A block consisted of either 300 trials from each of the three conditions or 300 own-ear trials and 300 other-ear trials using DTFs from each of two other subjects. Each block of 900 trials was broken into 7 sessions of 128 or 129 trials each. Individual listeners completed from 1 to 6 blocks of virtual-localization trials, where blocks differed in the DTFs that were used. Other-ear and scaled-ear performance in each block of trials were evaluated relative to performance on own-ear trials within the same block. Listeners typically completed a session of 129 virtual-localization trials in around 12 min. Listeners usually completed two sessions in a day, separated by a 10-min rest period.

A separate series of trials was conducted to test the sensitivity of virtual-localization performance to the value of a frequency scale factor. Nine listeners participated in that series. DTFs were scaled by a factor that varied randomly between trials in powers of 1.04 (i.e., scale factors of 0.79, 0.82,...,1.22,1.27). In separate blocks, DTFs either were those measured from the listeners' ears or those measured from another subject. Virtual targets were restricted to the midline and had polar locations that ranged from -60 to $+70°$ (in front) and from $+110$ to $+240°$ (in the rear) in increments of 10°. Each combination of location and scale factor was repeated in four trials. Four of the listeners completed these trials after experience in the main blocks of free-field and virtual-localization trials. The additional five listeners received only brief instruction in the free-field condition, then conducted these variable-scale factor trials with no prior virtual-localization experience. One of these naïve listeners later completed full training and the main blocks of free-field and virtual-localization trials.

D. Data analysis

The measurement and computation of DTFs is described in the companion paper. That paper described a procedure that involved filtering with a bank of band-pass filters followed by spatial interpolation. In the present study, that procedure was used only for quantification of optimal frequency scale factors and inter-subject differences between DTFs, not for synthesis of test stimuli. The differences between the DTFs of a given pair of subjects were quantified by the *inter-subject spectral difference*. Computation of inter-subject spectral difference is discussed in the companion paper, but briefly it involved subtraction of dB amplitudes of DTF components between 3.7 and 12.9 kHz for each location, computation of the variance of the resulting difference distributions, and averaging of those variances across all locations. *Optimal frequency scale factors* were computed by systematically scaling in frequency the DTFs of one subject relative to another to find the factor that minimized the inter-subject spectral difference. The companion paper demonstrated that scale factors based on right-ear and left-ear DTFs were highly correlated. In the present study, scale factors were derived from right-ear DTFs, then were used to scale DTFs for both right and left ears.

As described above, stimulus locations and localization judgements are expressed in horizontal polar coordinates. Errors in the lateral and polar dimensions were computed independently by subtracting the target location angle from the response location angle. In some cases, 360° was added or subtracted so that the magnitude of the error was no more than 180°. The precision of localization in the lateral dimension was summarized for each subject and condition by *root-mean-squared* (*rms*) *lateral errors*.

In the polar dimension, computation of summary statistics was complicated by the occurrence of large front/back and up/down errors that appeared to form a distribution that was distinct from the scatter of judgements around the stimulus location. Conventional definitions of front/back and up/down confusions might have included localization judgments that fell within the local scatter of judgements, such as when a target at polar 80° is localized as 91°. For the purpose of distinguishing polar localization judgements that fell well outside the scatter of judgements around the target, *quadrant errors* were defined as polar errors that were larger than 90°. The criterion of 90° was selected by examining histograms of the magnitudes of polar errors across all listeners and conditions. All quadrant errors necessarily involved localization judgements that fell in a different quadrant than the target, whereas some judgments that fell in the wrong quadrant produced polar errors smaller than 90° and were not scored as quadrant errors. The *rms local polar error* was defined as the rms average of polar errors that were less than 90° in magnitude. The purpose of that measure was to characterize localization accuracy in the polar dimension on trials in which quadrant errors did not occur. The distribution of polar errors for any subject and/or condition typically was strongly bimodal, so it was inappropriate to apply a measure of central tendency. For that reason, polar errors are summarized in this report only by the percentage of trials in which quadrant errors occurred (*% quadrant errors*) and by rms local polar errors.

Two signed measures were employed to quantify systematic biases in responses. The *lateral bias* and *elevation bias* were given by the averages of signed lateral and elevation errors, respectively, where elevation is the angle above or below the horizontal plane. Biases measured the tendency of listeners to respond consistently too far lateral to the target

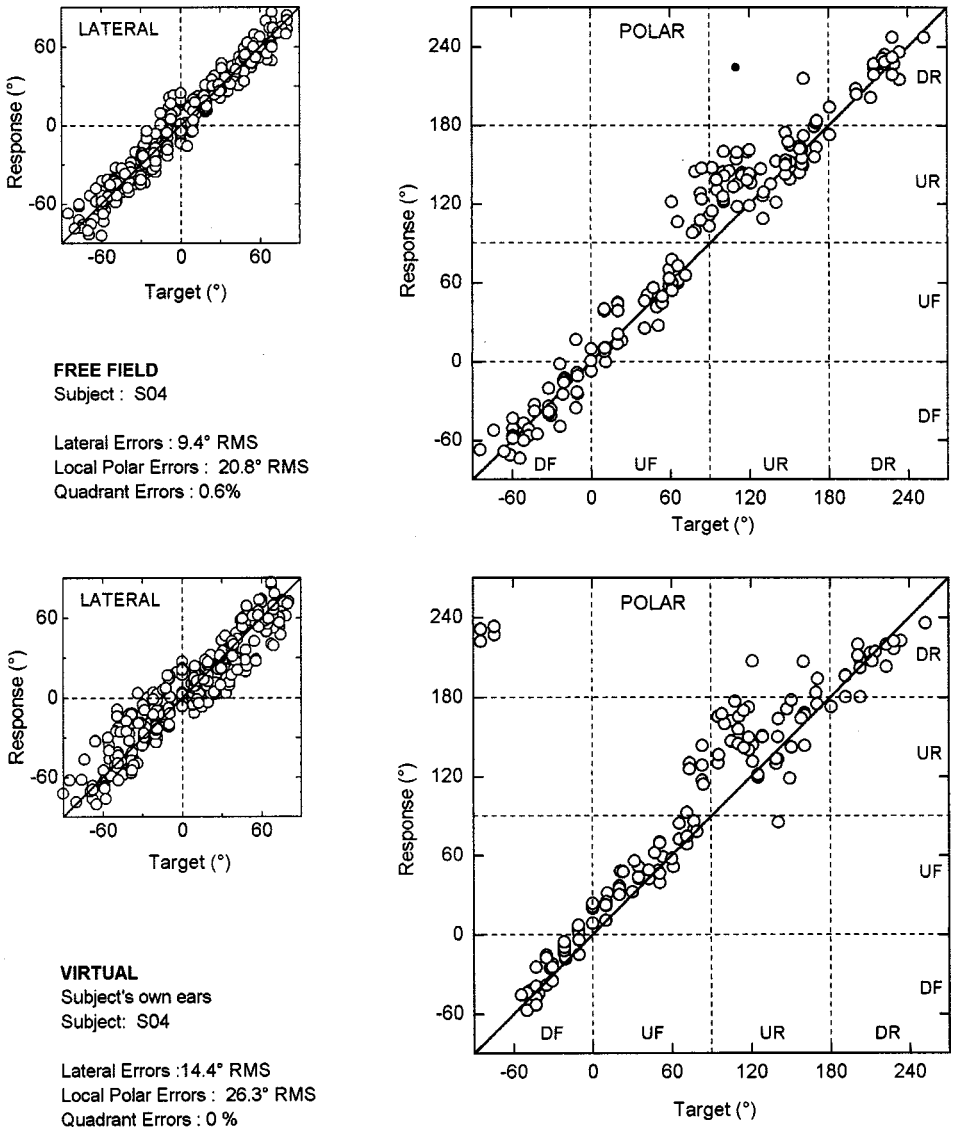


FIG. 2. Localization by listener S04 in the free-field (top panels) and own-ear virtual (bottom panels) conditions. Horizontal and vertical axes show the target locations and the listener's localization judgements, respectively. Localization in the lateral and polar dimensions is plotted in the left and right panels, respectively. Axes for lateral and polar dimensions are scaled so that one degree in either dimension occupies an equal length. Initials DF, UF, UR, and DR indicate down-front, up-front, up-rear, and down-rear quadrants. Filled circles (only one seen in this figure, in the upper right panel) represent quadrant errors. All other localization judgements are represented by open circles.

or too far above the target. Both computations of bias might have been underestimated by a ceiling effect because a response to a stimulus at, for instance, lateral 90° could be no further lateral than 90°. For that reason, biases were computed only for trials in which the stimulus locations were within 60° of the horizontal or midline planes. Two other classes of trials were excluded from the computation of elevation biases: those in which quadrant errors occurred and those in which the stimulus was within the up-rear quadrant. Up-rear quadrant trials were excluded because the responses to those stimuli tended to be so variable that they overwhelmed any systematic variation across listeners or conditions.

II. RESULTS

The results are presented in four sections. The *first* provides baseline measures of performance in the free-field localization condition and in a virtual condition in which targets were synthesized with DTFs measured from listeners' own ears. The *second* characterizes the types of virtual-localization errors that listeners made when they listened through DTFs measured from other subjects' ears. The *third* quantifies improvements in virtual localization that resulted from scaling DTFs in frequency, and it relates those improvements to a descriptive spectrum-matching model. The *final* section demonstrates the sensitivity of localization judgments to the value of the frequency scale factor that was used to scale DTFs.

A. Localization with listeners' own DTFs

Listeners localized in two conditions in which the sounds were filtered with the DTFs of their own ears. In the *free-field* condition, stimuli were presented from a movable loudspeaker and sounds were filtered naturally by the external ears as the sounds traveled from the source to the tympanic membranes. In the *own-ear* virtual localization condition, stimuli presented from headphones were filtered by DTFs that had been recorded previously from each listener's own external ears. Figure 2 shows localization results from one listener (S04) in the two conditions. This listener localized reasonably accurately throughout all lateral and polar angles, with the largest errors occurring in response to stimuli presented in the upper rear quadrant. Performance in

the own-ear virtual condition was nearly as accurate as that in the free-field condition. For the listener represented in Fig. 2, rms lateral errors were 9.4° and 14.4° in the free-field and own-ear conditions, respectively, and rms local polar errors were 20.8° and 26.3°, respectively. As described in Sec. I B, polar errors were computed only for targets located within 30° of the vertical midline.

The rate of occurrence of quadrant errors (defined in Sec. I D) in both free-field and own-ear virtual conditions showed a strongly bimodal distribution across the 14 listeners. In the free-field condition, nine listeners showed quadrant errors in only 0%–2.5% of trials (median = 0.6%), whereas the other five listeners showed quadrant errors in 8.6%–16% of trials (median = 12.3%). In the own-ear-virtual condition, nine listeners made quadrant errors in only 0%–6.3% of trials (median = 2.1%), and the five other subjects made quadrant errors in 10.5%–26.2% of trials (median = 13.2%). The listeners who produced a rate of quadrant errors of 8.6% or more in the free-field condition showed a rate of 4.3%–26.2% in the own-ear virtual condition.

The listener whose localization performance is shown in Fig. 2 is representative of those who produced few quadrant errors. That listener made only one quadrant error in the free-field condition (indicated by a filled circle) and no quadrant errors in the own-ear virtual condition. Figure 3 shows localization performance by one of the five listeners who showed a high rate of quadrant errors in free-field and virtual trials. The majority of that subject's errors consisted of up-rear targets localized to down-front.

Across all listeners who showed quadrant errors, the rate of quadrant errors was highest for targets in the up-rear quadrant. The five listeners who showed a high rate of quadrant errors showed quadrant errors in 17.5%–41.3% (median = 25.4%) of free-field trials and in 32.1%–60.5% (median = 43.2%) of own-ear virtual trials in which targets were in the up-rear quadrant. Examination of DTFs (e.g., Middlebrooks, 1999) indicates that DTFs for up-rear locations tend to be relatively devoid of spectral features. That could explain why discrimination of targets within that quadrant might be imprecise, but it does not explain why targets in that quadrant would be localized to other quadrants.

Table I shows summary statistics for the free-field and own-ear conditions. With few exceptions, mean errors in the own-ear virtual condition were within about one standard deviation of mean errors in the free-field condition.

B. Localization with other subjects' DTFs

In the *other-ear* virtual localization condition, each listener listened to sounds that were filtered with DTFs measured from the ears of a different subject. The difference between each listener's DTFs and the DTFs used in a localization trial was quantified by a global measure of *inter-subject spectral difference*, which is defined in the companion paper. Listeners were matched with sets of DTFs so that cases spanned a range of the smallest to largest spectral differences. Fourteen listeners each localized using DTFs from one to seven other subjects, yielding a total of fifty-eight cases. Figure 4 shows the increase in errors in the other-ear

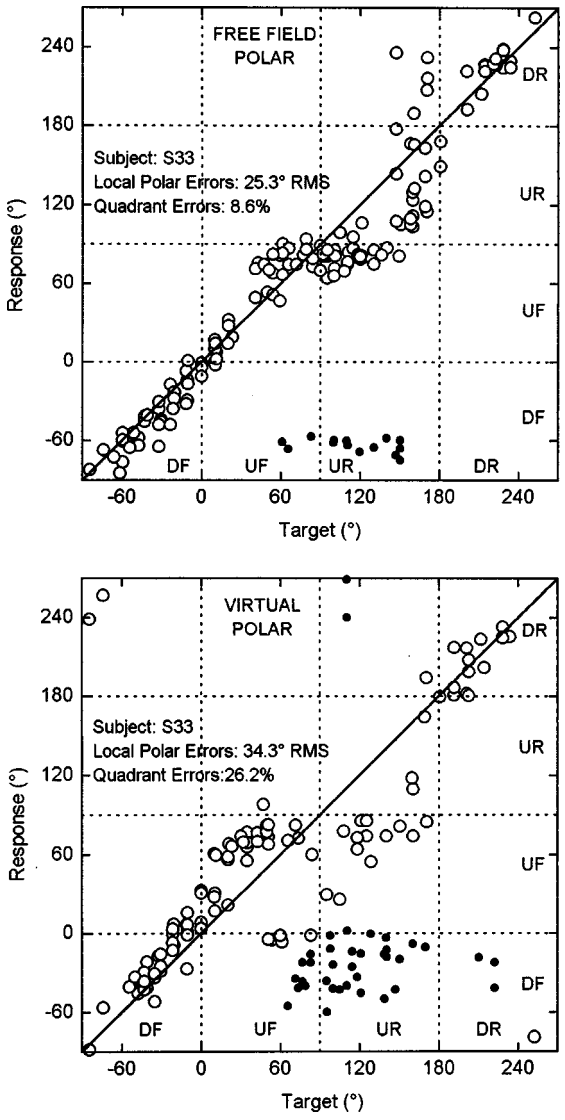


FIG. 3. Localization by listener S33, who showed a high rate of quadrant errors. All other conventions are as in Fig. 2.

condition compared to the own-ear condition as a function of inter-subject spectral difference in DTFs. All error measures of rms errors tended to increase with increases in the spectral difference between the listener's DTFs and the DTFs used in the localization trials. The rate of quadrant errors consistently was larger in the other-ear condition than in the own-ear condition, although that increase was negligible in sev-

TABLE I. Summary measures of free-field and own-ear virtual localization performance (means and standard deviations).

	Free-field	Own-ear virtual
rms lateral error (deg.)	10.6±2.0	14.5±2.2
Magnitude of lateral bias (deg.)	2.9±3.1	3.1±3.9
rms local polar error (deg.)	22.7±5.1	28.7±4.7
Magnitude of elevation bias (deg.)	5.5±4.4	10.2±6.6
Total quadrant errors (% of trials)	4.6±5.9	7.7±8.0
Quadrant error by target quadrant:		
Down-front (% of trials)	0.6±1.1	1.0±2.6
Up-front (% of trials)	1.9±3.3	5.7±6.4
Up-rear (% of trials)	10.6±13.8	21.7±21.1
Down-rear (% of trials)	0±0	1.8±5.0

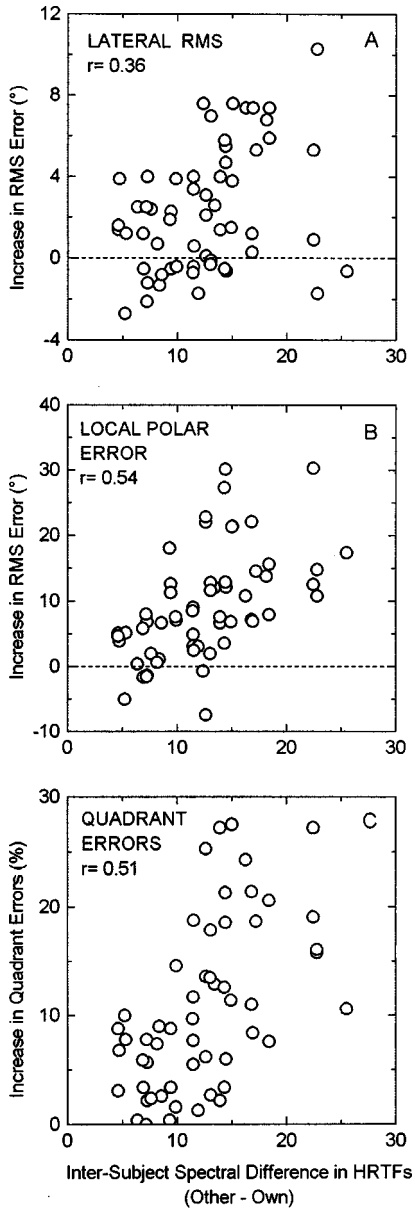


FIG. 4. Changes in localization performance resulting from use of nonindividualized DTFs. Each symbol represents one of 58 cases, each case consisting of one listener using one set of DTFs.

eral instances in which the inter-subject spectral difference in DTFs was smaller than about 10 dB^2. In a few instances, lateral errors and local polar errors were slightly smaller in the other-ear condition, although there was little evidence for a significant improvement in those measures in the other-ear compared to the own-ear condition. As a general rule, listeners' localization performance suffered when they listened through other subjects' DTFs.

Figure 5 shows two examples of virtual localization in the other-ear virtual condition. The results are from the same listener whose free-field and own-ear-virtual results are shown in Fig. 2. In Fig. 5, the upper and lower halves of the figure represent performance in the other-ear virtual condition using DTFs from two different subjects. In the lateral dimension (left two panels) one can see a moderate increase, relative to the own-ear condition, in the scatter of responses but there is still a strong correlation between target and response locations in the lateral dimension. In contrast, responses in the polar dimension (right two panels) showed a considerable increase in quadrant errors, from 0% in the own-ear condition to 27.6% or 12.9% in the other-ear conditions. The effect on local polar errors differed between sets of DTFs. In the upper right panel, local polar errors actually decreased slightly, possibly because the number of responses in the local-error computation was reduced by the large number of quadrant errors. In the lower right panel, the rms local polar error increased by nearly 50% relative to the own-ear condition. The distributions of responses in the polar dimension were different in the two cases. In the case shown in the upper right panel, nearly all down-front targets were localized to the down-rear quadrant. Most responses to down-rear targets were in the correct quadrant, but were below the target (i.e., above the diagonal in the figure). In the case shown in the lower right panel, the majority of down-rear targets were localized to the front. Responses to most targets in the front were systematically too high.

The differences in the distribution of localization judgements in the two other-ear conditions shown in Fig. 5 can be related to a property of DTFs that was described in the companion report (Middlebrooks, 1999). Specifically, that report demonstrated that DTFs tend to differ systematically among subjects in regard to the center frequencies of spectral features (e.g., spectral peaks, notches, and slopes). In addition, inter-subject frequency differences tended to correlate with the physical dimensions of subjects' heads and external ears. Therefore, for ease of presentation, this report will refer to *large* and *small* subjects, where *large* indicates a subject in whose DTFs spectral features tend to lie at lower frequencies than those of a *small* subject. In behavioral results in the other-ear virtual condition, small listeners who listened through large subjects' DTFs tended to produce different patterns of errors than did large listeners who listened through small subjects' DTFs. For instance, the listener represented in Fig. 5 was near the middle of the range of listeners. The localization judgements shown in the upper and lower panels were made when listening through the DTFs of larger and smaller subjects, respectively. Three characteristics of other-ear localization judgements correlated fairly consistently with inter-subject differences in the center frequencies of DTF spectral features or with inter-subject differences in interaural delays. Those characteristics were lateral biases, the distribution of quadrant errors, and elevation biases.

Listeners tended to show positive or negative lateral biases when listening through DTFs from larger or smaller subjects, respectively. Most models of sound localization attribute the lateral component of localization to interaural difference cues (reviewed in Middlebrooks and Green, 1991), and Wightman and Kistler (1992) have shown that interaural time differences (ITDs) dominate the lateral component of localization judgements of sound sources that contain low frequencies. The latter result predicts that lateral biases would correlate with the maximum ITDs that were produced by a given set of DTFs. To test that prediction, the ratio of the maximum ITDs was computed for each combination of behavioral listener and set of other-ear DTFs. Ratios greater

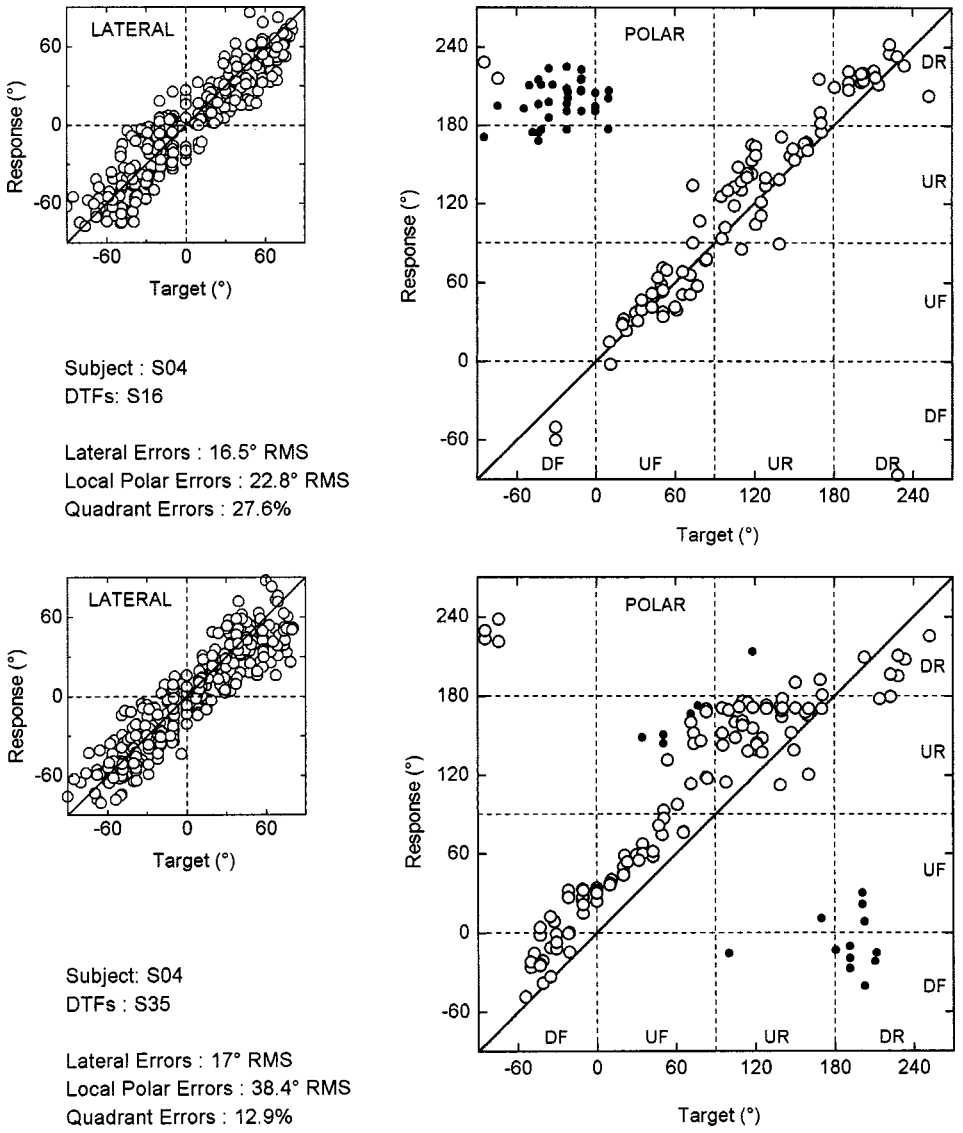


FIG. 5. Localization by listener S04 listening through DTFs from a larger subject (S16, top panels) and those of a smaller subject (S35, bottom panels). All other conventions as in Fig. 2.

than unity corresponded to small listeners localizing with larger subjects' DTFs. The increase in lateral bias was computed by subtracting the lateral bias in the own-ear condition from that in the other-ear condition. Figure 6 shows a strong positive correlation between maximum-ITD ratio and increase in lateral bias. The solid curve is a prediction based on the assumption that when a listener hears an ITD produced by a pair of other-ear DTFs for lateral angle θ_{other}, he or she reports a location θ_{own} at which his or her own DTFs would produce an identical ITD. The relationship between θ_{other} and θ_{own} is given by the following expression, which is derived from the spherical-head model of ITDs (Woodworth, 1938):

$$\text{MaxITD}_{\text{own}}(\theta_{\text{own}} + \sin(\theta_{\text{own}})) = \text{MaxITD}_{\text{other}}(\theta_{\text{other}} + \sin(\theta_{\text{other}})),$$

where MaxITD for each set of DTFs is obtained by fitting the spherical-head model to the complete set of DTFs, then evaluating the model at $\theta = 90°$. The ratio of maximum ITDs was computed for each subject and set of DTFs represented in Fig. 6. Each response θ_{own} was then estimated for every target θ_{other}, and the average of $\theta_{\text{other}} - \theta_{\text{own}}$ was averaged across all targets. The solid curve in Fig. 6 is the predicted

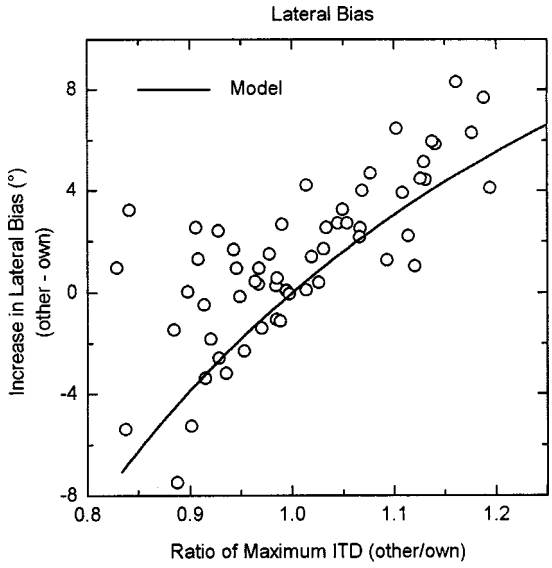


FIG. 6. Increase in lateral bias in the other-ear condition. A positive increase indicates that the listener tended to point farther to the side in the other-ear condition than in the own-ear condition. The abscissa is the ratio of maximum interaural delays, other-ear/own-ear. The solid line is a prediction described in the text ($r = 0.73$, $N = 58$).

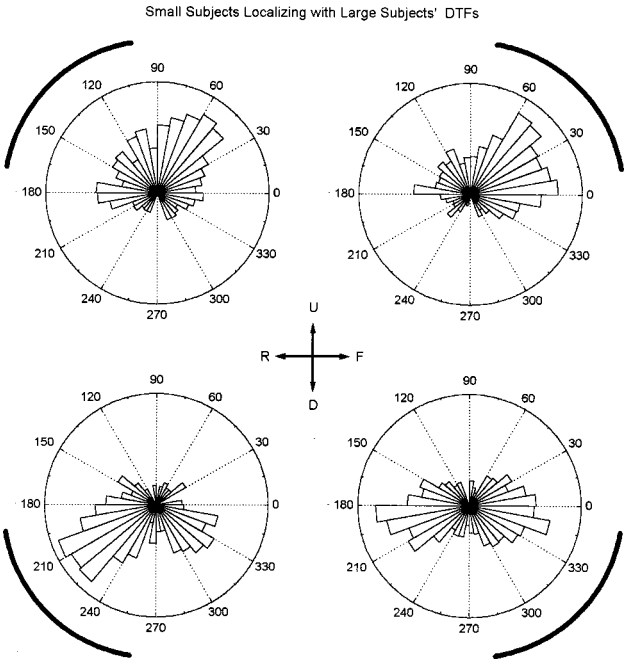


FIG. 7. Distribution of responses in the polar dimension: larger subjects' DTFs. These rose plots are drawn as if viewed from the listeners' right side. Each panel shows the response to targets presented in a particular quadrant, indicated by a thick arc. The *area* of each petal of a rose plot represents the percentage of responses that fell in a particular 10° sector (i.e., the *radius* of each petal is the square root of the percentage). This figure represents the 14 cases that showed the greatest downward frequency disparity between listeners' own DTFs and the DTFs through which they listened; i.e., spectral features in the DTFs tended to lie at lower frequencies than those in the listeners' own DTFs. The listeners' up, down, front, and rear directions are indicated by initials U, D, F, and R.

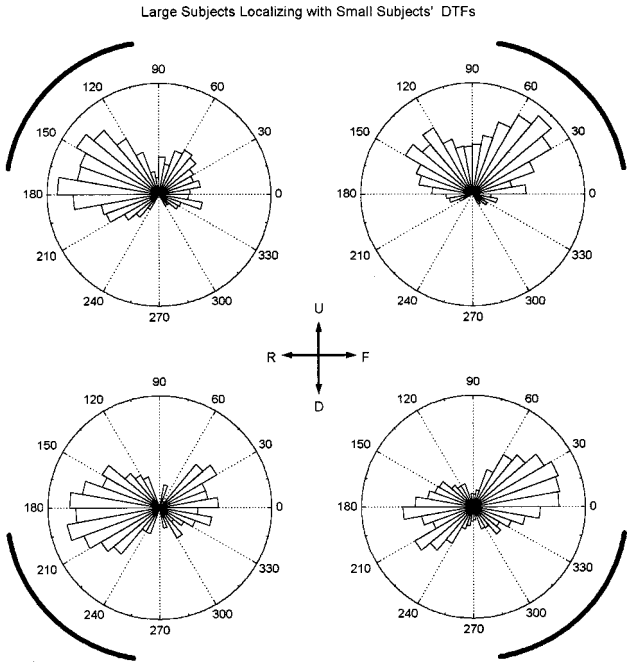


FIG. 8. Distribution of responses in the polar dimension: smaller subjects' DTFs. This figure represents the 14 cases that showed the greatest upward frequency disparity between listeners' own DTFs and the DTFs through which they listened; i.e., spectral features in the DTFs tended to lie at higher frequencies than those in the listeners' own DTFs. All other conventions are as in Fig. 7.

increase in lateral bias for each ratio of head radii. The modeled curve lies within the scatter of data and follows the general trend. The largest deviation of data points from the curve is due to some of the large listeners localizing with small subjects' DTFs (i.e., ITD ratio <1) that did not show as large an undershoot as predicted by the model.

The most conspicuous characteristic of localization errors in the other-ear condition was the quadrant errors. The results shown in Fig. 5 were fairly representative of the behavior of listeners localizing with the DTFs of larger subjects (upper panels) or of listeners localizing through the DTFs of smaller subjects (lower panels). For the purpose of presentation, all the other-ear cases were ranked according to the optimal frequency scale factors between each listener's own DTFs and the DTFs through which he or she listened. The quartile of small listeners/large-subject DTFs is represented in Fig. 7 and the quartile of large listeners/small-subject DTFs is represented in Fig. 8. The distributions of localization judgements are shown in *rose plots*, in which the *area* of each rose petal is proportional to the percentage of responses in a particular sector. Each of the four panels in each figure represents the distribution of judgments of targets presented within the quadrant that is indicated by the bold arc. In the condition of smaller listeners localizing with larger subjects' DTFs (Fig. 7), three characteristics were most common. (1) Down-front targets often were localized to down-rear ($36.3\% \pm 25.2\%$ of trials, mean $\pm$SD). (2) About half of up-rear targets were localized to up-front ($51.6\% \pm 15.0\%$). (3) Front/back confusions were more common than up/down confusions ($44.4\% \pm 7.3\%$ front/back compared to $23.4\% \pm 8.0\%$ up/down). The characteristics were different in the condition of larger listeners localizing with smaller subjects' DTFs (Fig. 8). (1) There was a considerable rate of front/back confusions ($31.2\% \pm 9.7\%$), although the rate was lower than in the small-listener/large-ear condition. (2) More conspicuous was a systematic upward bias of localization judgements from the lower quadrants to on or above the horizontal plane. More than half (58.2 ± 20.3) of targets in lower quadrants were localized to upper quadrants.

Although the quadrant errors tended to dominate the polar component of responses, trends in polar errors were evident even in trials in which quadrant errors did not occur. In particular, listeners tended to show a positive elevation bias (i.e., to point above the target) in conditions in which they listened through smaller subjects' DTFs. The tendency to point above the target is shown clearly in the lower plot of Fig. 5 in which responses above the targets appear above the identity line in the front hemifield and below the identity line in the rear hemifield. A symmetrical tendency to point below targets when listening through larger subjects' DTFs also was present. That tendency was less conspicuous, however, because down-front targets often were localized to incorrect quadrants (as in the upper panel of Fig. 5). Similarly, a clear trend is difficult to see for the up-front quadrant in the example in Fig. 5, although one can see a downward bias in targets in the down-rear quadrant (i.e., symbols fall above the identity line in the rear).

Figure 9 plots the elevation biases of 58 cases as a function of the optimal frequency scale factor that would minimize the spectral difference between the listener's own DTFs

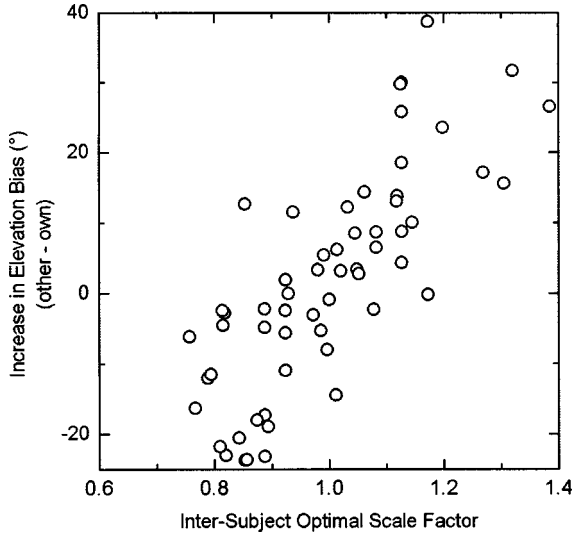


FIG. 9. Increase in elevation bias in the other-ear condition. A positive increase indicates that the listener tended to point to higher elevations in the other-ear condition than in the own-ear condition. The abscissa plots the optimal frequency scale factor that would scale the listeners' DTFs to the DTFs that were used in each case. A scale factor >1 indicates that spectral features in the DTFs tended to lie at higher frequencies than those in the listeners' own DTFs ($r=0.78$, $N=58$).

and the other-ear DTFs. A value greater than 1 on the abscissa indicates that spectral features in the DTFs that were used fell at higher frequencies than in the listener's own DTFs. Biases are expressed as increases relative to the bias measured in the own-ear condition. There is a strong positive correlation. Examination of DTFs (e.g., in the companion paper) indicates that, in general, spectral features tend to shift upward in frequency as sound sources shift upward in elevation. When a listener's localization judgments show an upward bias in response to the DTFs of a smaller subject, that might indicate that the listener hears spectral features at frequencies higher than expected and interprets those features as signaling higher elevations.

C. Localization with frequency-scaled DTFs

The companion paper demonstrated that inter-subject differences in DTFs could be reduced by appropriately scaling one set of DTFs in frequency relative to the other. That result led to the prediction that scaling DTFs in frequency could result in a reduction of errors in other-ear localization. In testing that prediction, optimal frequency scale factors were chosen that minimized the inter-subject spectral difference in DTFs (as described in the companion paper). The condition in which listeners localized with other-ear DTFs scaled in frequency is referred to here as the *scaled-ear* condition. The DTFs were scaled by changing the output rate of the digital-to-analog converter (DAC). When the DAC rate was multiplied by β, DTF spectral features were scaled in frequency by a factor β, interaural difference spectra were scaled in frequency by β, and interaural delays were scaled in time by $1/\beta$.

Figure 10 shows localization performance in the polar dimension in the scaled-ear condition for the same listener who is represented in the own-ear condition in Fig. 2 and in

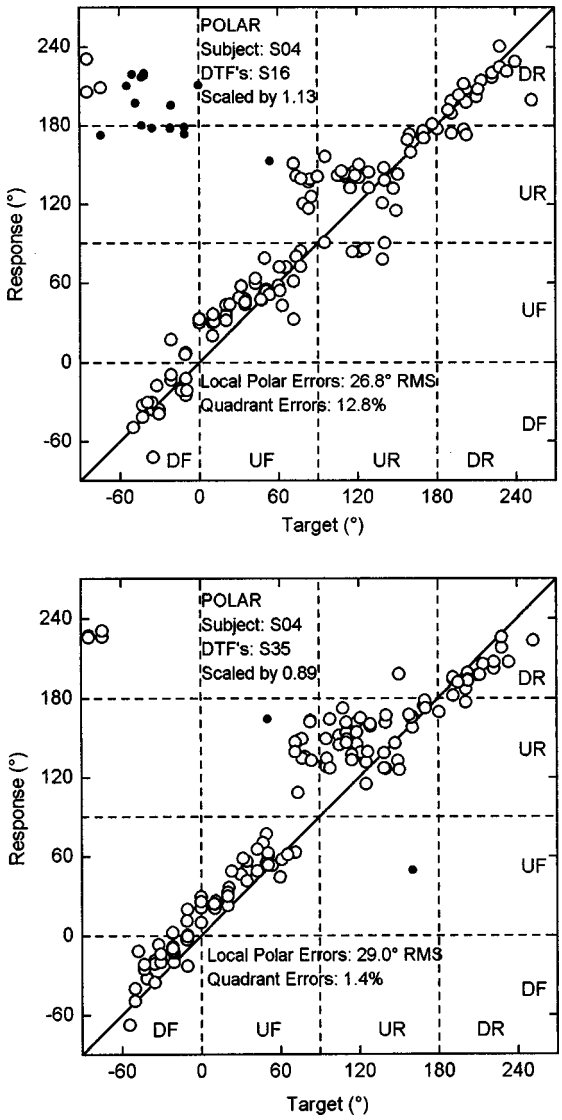


FIG. 10. Localization by listener S04 listening through scaled DTFs from subjects S16 (top panels) and S35 (bottom panels). DTFs were scaled by optimal frequency scale factors of 1.127 (top panel) and 0.888 (bottom panel). All other conventions are as in Fig. 2.

the other-ear virtual condition in Fig. 5. The DTFs were from the same two subjects as in Fig. 5. The top panel of Fig. 10 shows a case in which DTFs from a larger subject were scaled in frequency by a factor of 1.127. Compared to the other-ear judgments in the top panel of Fig. 5, the number of quadrant errors in response to down-front targets was halved and the accuracy of responses in the down-rear quadrant was greatly improved. Local polar errors are slightly increased, but that might be a result of the large number of trials that were excluded from the local-error computation in the other-ear case because quadrant errors occurred. The bottom panel of Fig. 10 represents a case in which DTFs from a smaller subject were in frequency by a factor of 0.888. Compared to the lower panel of Fig. 5, all of the quadrant errors in response to down-rear targets were eliminated and the upward bias in elevation responses was reduced substantially, about to the level seen in this listener's own-ear virtual localization (Fig. 2, lower panel). In this case, frequency scaling reduced

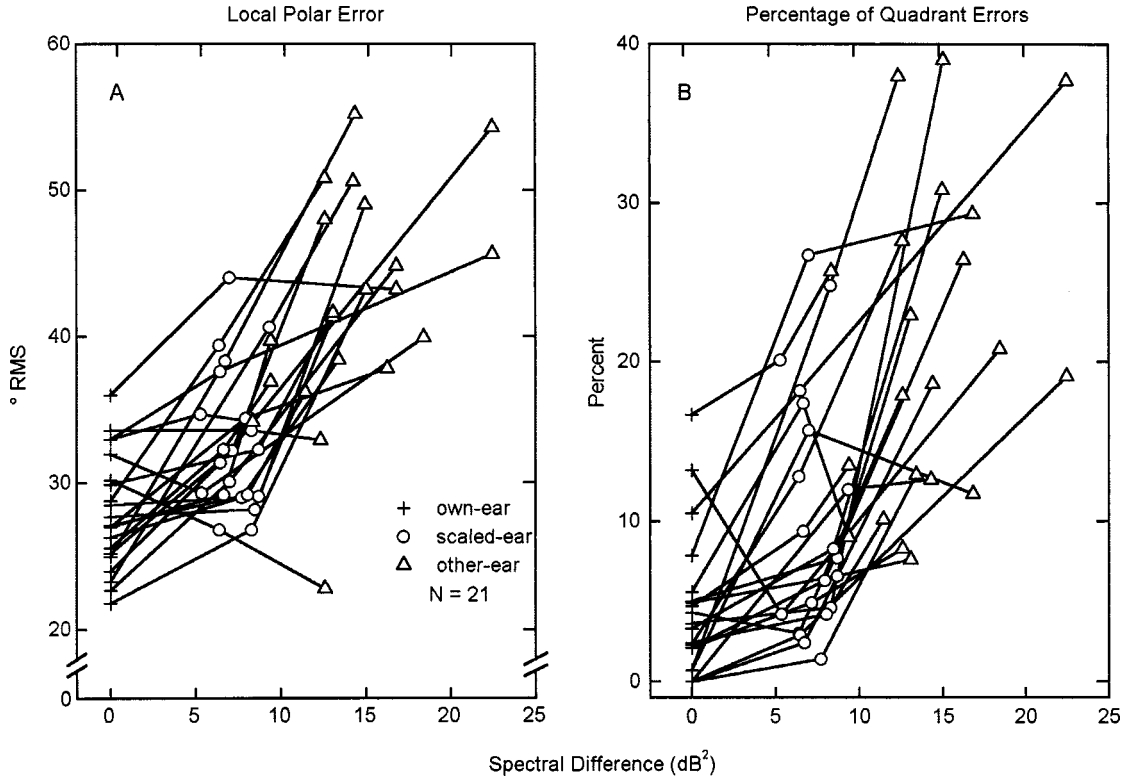


FIG. 11. rms local polar error (A) and percent quadrant errors (B) in three virtual-listening conditions. Each triad of symbols connected by lines represents one listener in the own-ear (plus signs), other-ear (triangles), and scaled-ear (circles) conditions. The abscissa shows the spectral difference between the DTFs that were used in each case and the listeners' own DTFs.

the overall rate of quadrant errors from 12.9% to 1.4%, and local polar errors from 38.4° to 29.0° rms.

The scaled-ear condition was tested in 21 cases (11 listeners tested with scaled DTFs from 1 to 4 other subjects). Figure 11 shows the rms local polar errors [Fig. 11(A)] and the percentage of quadrant errors [Fig. 11(B)] of those cases in the own-ear (plus signs), other-ear (triangles), and scaled-ear (circles) conditions. Lines connect the data points for each listener. Data are plotted against the inter-subject spectral difference between the other-ear DTFs (scaled or unscaled) and each listener's own DTFs. Frequency scaling produced substantial reductions in the rms local polar errors and in the percentage of quadrant errors relative to the other-ear condition in nearly all cases. In 17 of 21 cases, that reduction was proportionately greater than the reduction in the spectral difference between DTFs.

Figure 12 compares quadrant errors and rms local polar errors in the scaled-ear versus other-ear conditions. Errors

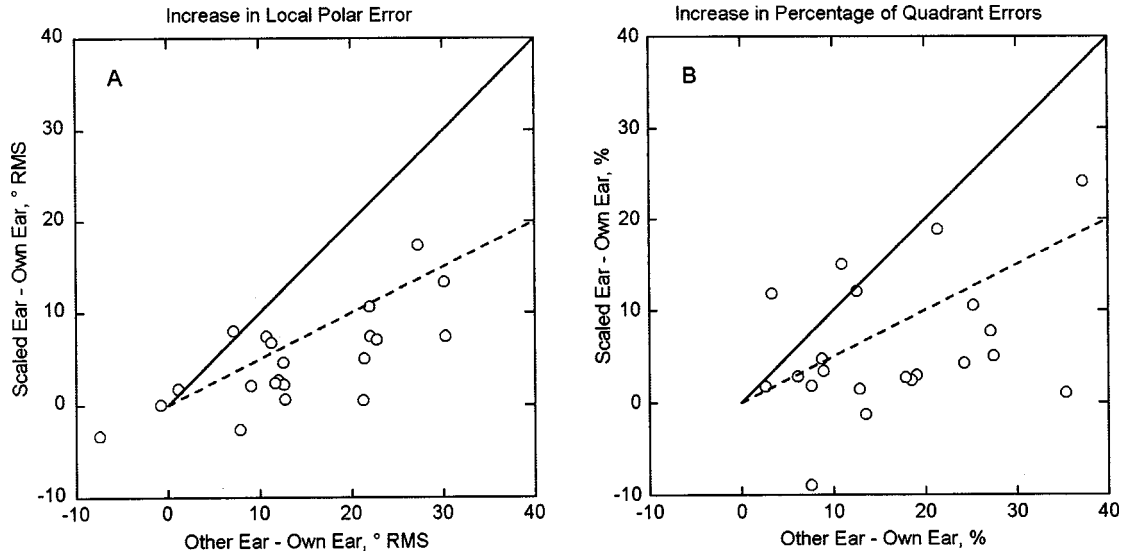


FIG. 12. rms local polar errors (A) and percent quadrant errors (B) in the scaled-ear versus the other-ear condition. Errors are expressed as the increase in rms error or the increase in percent quadrant errors relative to the own-ear condition. The solid line represents equal performance in other- and scaled-ear conditions, and the dashed line indicates 50% reduction in the increase in error.

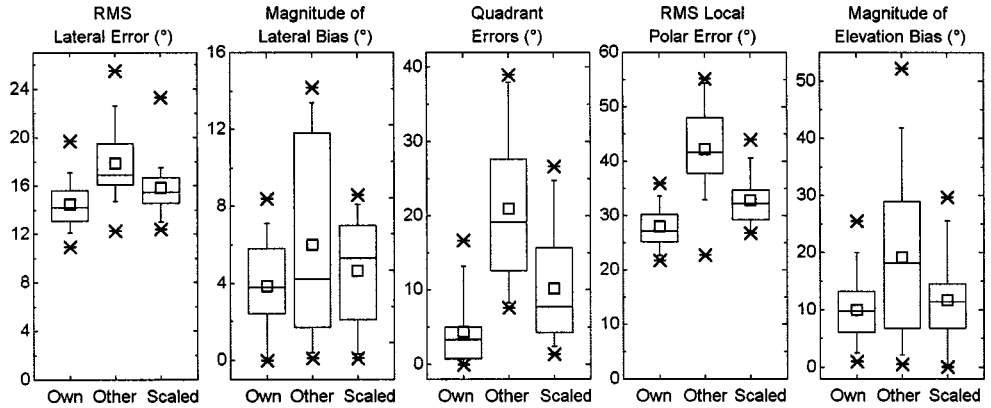


FIG. 13. Summary of performance in three virtual localization conditions. In these box plots, horizontal lines represent 25th, 50th, and 75th percentiles, the extended vertical bars represent 5th and 95th percentiles, and ×'s represent minima and maxima. The small squares indicate the means. Eleven listeners were tested in the *own* condition. The same 11 listeners were tested with 1–4 sets of DTFs from other subjects for a total of 21 cases in the *other* and *scaled* conditions.

are expressed as increases in the magnitude of rms local polar errors and the percentage of quadrant errors in scaled-ear and other-ear conditions relative to the own-ear condition. Two-thirds of points (14/21) fell below the dashed line with slope of $\frac{1}{2}$, indicating that scaling DTFs in frequency reduced by more than half the penalty for listening through another subject's DTFs. Scaling DTFs in frequency resulted in improvements relative to the other-ear condition in every summary measure of localization performance (Fig. 13).

A previous paper proposed a simple model to account for listeners' systematic errors in localization of narrow-band sounds (Middlebrooks, 1992). That model, adapted to the current results, provides some understanding of listeners' localization judgements in the other-ear and scaled-ear conditions. The model assumed that each listener's auditory system maintains in some form a library of templates of the DTFs associated with each sound-source direction. It was speculated that when a stimulus arrives at the tympanic membranes, the listener compares that proximal stimulus with templates of DTFs, and that the listener's localization judgement corresponds to the localization for which the template fits most closely. In the current study, the DTF-matching component of that model was applied to data from the other-ear and scaled-ear conditions. To conform to the procedure in this and the companion paper, comparisons were made among DTFs by computing variances of inter-subject DTF difference spectra instead of by computing correlations as in the previous study. Listeners tended to localize accurately in the lateral dimension regardless of the testing condition. For that reason and to focus on spectral shape recognition, the present analysis neglected interaural disparities and the lateral dimension of localization judgements and simply assumed that listeners evaluate DTF spectra only within the correct lateral plane (i.e., on the correct cone of confusion). The model predicted that listeners would report apparent source locations at points that minimize inter-subject variance in DTFs. Figures 14 and 15 show examples of the application of the DTF-matching model to results from one listener localizing with DTFs from two other subjects.

Figures 14 and 15 both represent localization of a virtual target at a single location: lateral −20°, polar −21° (20° to the left, 21° below the horizon in front). In each figure, the target location is indicated by a dashed vertical line, and the listener's localization judgements are indicated by triangles. The same listener (S25) localized using DTFs from a larger subject (S27, Fig. 14) and from a smaller subject (S35, Fig. 15). The proximal stimulus in each case was synthesized with the other-ear DTF for −20°, −21°, either unscaled (upper panels) or scaled (lower panels). Each plot shows the

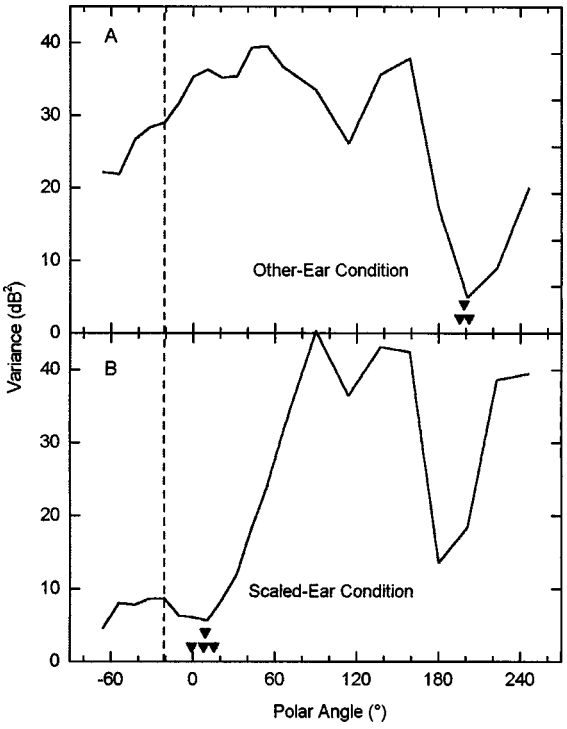


FIG. 14. Polar localization judgements related to variance in DTFs. Localization of a virtual target at left 20°, polar −21° is shown for listener S25 using DTFs from a larger subject, S27. The curve in each plot represents the variance of the difference spectrum between the DTF that was used to synthesize the target (i.e., S27's DTF for −20, −21°) and the listener's own DTFs at each of a range of polar angles at lateral 20°. Vertical dashed lines indicate the target location. Triangles indicate the localization judgements, in some cases displaced upward to avoid overlap. In A, the target DTF was used unscaled. In B, it was scaled upward in frequency by a factor of 1.22.

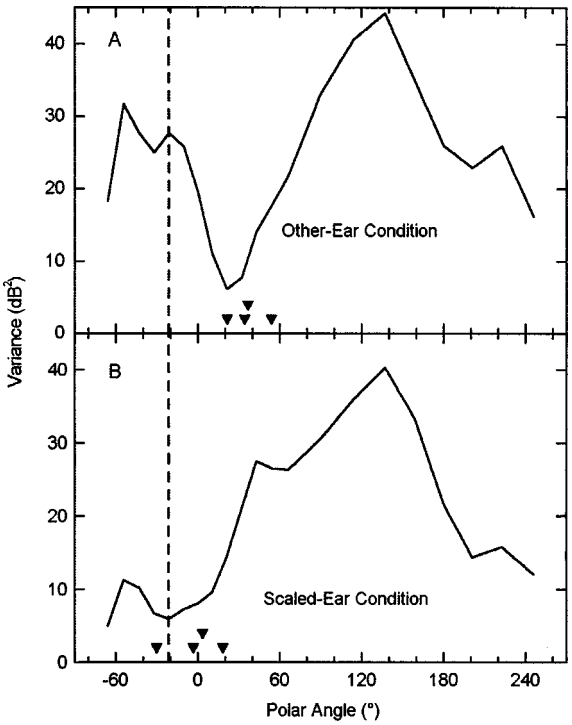


FIG. 15. Polar localization judgements related to variance in DTFs. The format is identical to that in Fig. 14, as are the listener (S25) and the target location. The stimuli in this case were synthesized with DTFs from a smaller subject, S35. In the scaled condition, S35's DTF for −21° was scaled downward by a factor of 0.889.

variance of the difference spectrum between the stimulus spectrum and the listener's own DTFs at all polar angles at lateral −20°. In the other-ear condition in which the DTF came from a larger subject [Fig. 14(A)], the variances between DTFs were fairly high near the virtual-target location but showed a sharp minimum near polar 200° (20° below the horizon in the rear). The listener's three responses to the −21° target fell near that minimum. When the other-ear DTF for polar −21° was scaled in frequency, the variance with the listener's own DTFs showed an entirely different pattern [Fig. 14(B)]. In that condition, local minima in the variance plot appeared at polar −60°, around 0°, and around 180°. The listener's four responses fell in the region of low variance near 0°.

In the other-ear condition in which the DTFs came from a smaller subject [Fig. 15(A)], the variance plot showed a sharp minimum near +20°, 41° higher than the target. The four localization judgments showed an upward elevation bias, falling near the minimum in the variance plot. In the scaled-ear condition [Fig. 15(B)], the minimum in the variance plot shifted downward to lie on top of the target location, and the listener's responses shifted downward by a corresponding amount.

Plots of variance versus polar angle, like those in Figs. 14 and 15, could predict qualitative characteristics of most listeners' localization judgements in the other-ear and scaled-ear conditions. For instance, among the quartile of small listeners localizing through larger subjects' DTFs, the variance plots computed for down-front targets consistently showed a local minimum in the up-rear or down-rear quadrant that corresponded to listeners' responses. Among the quartile of larger listeners localizing through smaller subjects' DTFs, the variance plots computed for down-front targets consistently showed a minimum at frontal locations with an upward elevation bias. Variance plots often showed two local minima, one for a frontal location and one for a rear location, with subjects' localization judgements distributed between those minima. Examples of variance plots containing both front and rear minima are considered in the next section.

D. Sensitivity of localization accuracy to frequency scale factors

A separate series of behavioral measurements tested the influence of the values of frequency scale factors on localization performance. To economize listener time, virtual targets were restricted to the vertical midline. Frequency scale factors were varied in powers of 1.04, and stimuli synthesized with various scale factors were interleaved within blocks of trials. The listener pool consisted of five listeners who had no experience in virtual localization plus four listeners who were experienced virtual localizers. The naïve listeners tended to localize less accurately than the experienced listeners, but the dependence of localization accuracy on frequency scale factor was similar between naïve and experienced listeners.

The accuracy of virtual localization tended to show sharp sensitivity for the value of the frequency scale factor. This is shown in Fig. 16, in which the solid curves and the right ordinates show the percentage of quadrant errors of listener S16 who localized with his own DTFs, scaled by various factors [Fig. 16(A)] or with scaled DTFs measured from listener S04 [Fig. 16(B)]. The sharpness of sensitivity to frequency scale factors was estimated from tuning plots like those in Fig. 16 by computing the width of the plot across which the quadrant error was within 5 percentage points of the minimum. Widths computed in that way are expressed as the maximum scale factor within the 5-percentage point criterion divided by the minimum within that criterion. Widths ranged from 1.13 to 1.22 (median $=1.15$) for the four listeners who were tested with their own DTFs and from 1.04 to 1.22 (median$=1.15$) for nine listeners who were tested with other subjects' DTFs.

The frequency scale factors that resulted in best behavioral performance tended to correspond to the frequency scale factors that minimized the inter-subject spectral difference. The dashed curves and left ordinates in Fig. 16 show the spectral difference between one listener's own DTFs, unscaled, and his own or another subject's DTFs scaled by various factors. Across all 13 cases in which listeners localized with their own or another subject's DTFs scaled by varying factors, the best frequency scale factors obtained from behavioral and acoustical measurements correlated with $r=0.92$. The rms difference between the logarithms of scale factors measured behaviorally and acoustically, expressed as a linear factor, was 1.052, which is little more than the minimum increment in frequency scale factors (1.04) that was tested in the behavioral trials.

When a listener's DTFs are scaled downward or upward in frequency, the DTFs tend to approximate those of a larger

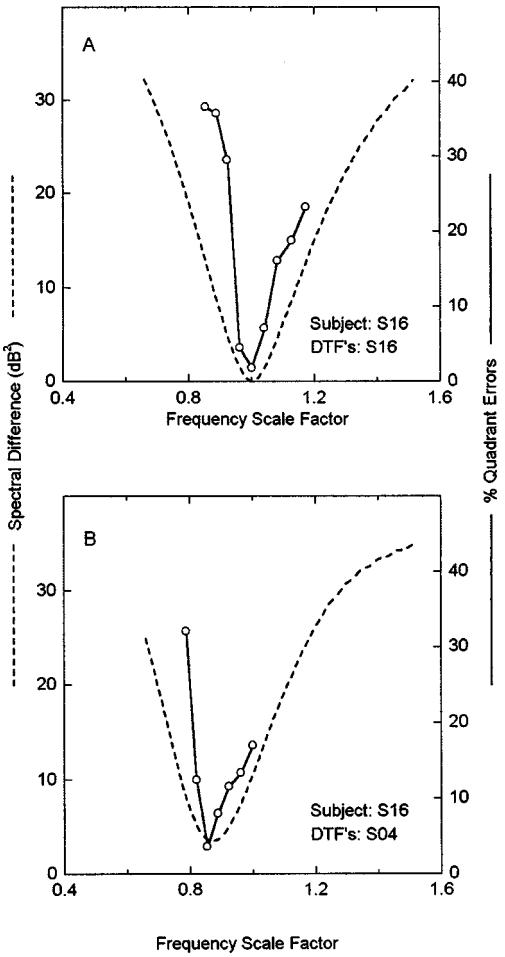


FIG. 16. Sensitivity of the percent of quadrant errors to frequency scale factor. Solid curves and right ordinate show localization performance in the vertical midline by listener S16. The DTFs were from the listener's own ears (A) or from subject S04 (B). DTFs were scaled in frequency by the scale factor that is plotted on the abscissa. Dashed curves and left ordinate show the spectral variance between the listener's own DTFs (unscaled) and either his own DTFs (A) or S16's DTFs (B), scaled by the value plotted on the abscissa.

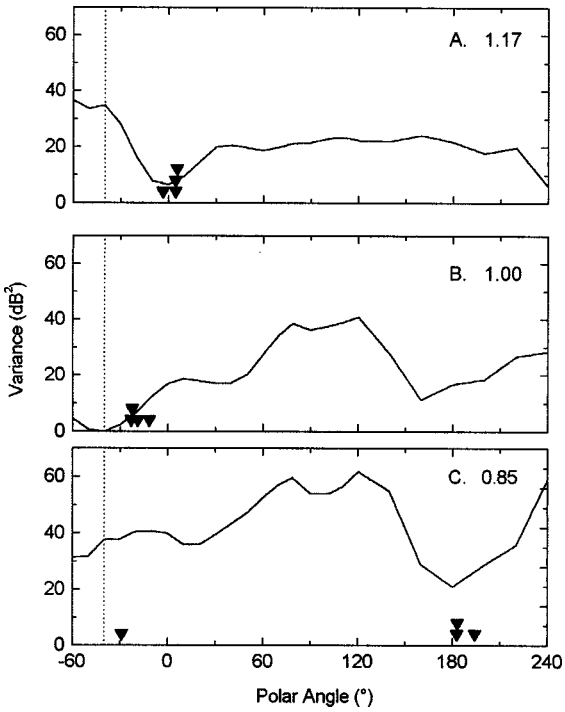


FIG. 17. Influence of frequency scale factors on cues for localization in polar dimension. Data are from trials in which listener S40 localized virtual targets at 0°, −40° using his own DTFs scaled in frequency by 1.17 (A), 1.00 (B), and 0.854 (C). Other conventions as in Fig. 14.

or smaller subject, respectively. For that reason, localization with DTFs scaled by various amounts simulates localization with DTFs from a continuum of other subjects. Figure 17 represents one listener who localized with his own DTFs scaled in frequency by three scale factors. In each case the target was on the frontal midline at −40° elevation. The curves in each panel plot the variance in the difference spectrum between the listener's DTF for −40° elevation, scaled by the stated amount, and the listener's unscaled DTFs for a range of locations in the vertical midline. The vertical dashed line indicates the target location, and the filled triangles indicate the listener's four responses in each condition. When the scale factor was one [Fig. 17(B)], the variance was zero at the target location, and the listener's responses fell slightly above that elevation. In that panel, one can see a second, less deep, minimum in the variance plot located in the rear, around 160°, but none of the listener's responses fell there. As the stimulus DTF was scaled to a lower frequency [Fig. 17(C)] the variance for DTFs at frontal locations grew larger than that in the rear, and the minimum in the variance plot at rear locations became the overall minimum. In that condition, 3/4 of the listener's judgements fell in the rear and one fell near the low-variance locus in front. The variance plot in that condition resembles the variance plot shown for a condition of a listener localizing through larger subject's DTFs (Fig. 14, top panel). When the listener's DTFs were scaled upward in frequency [Fig. 17(A)], the frontal minimum in the variance plot and the listener's responses shifted upward in elevation. The condition of an upward shift in DTFs produced a variance plot similar to that observed when a listener used a smaller subject's DTFs (Fig. 15, top).

III. DISCUSSION

The results that have been presented support two main conclusions. *First*, consistent with previous observations, sound-localization accuracy is degraded when listeners listen to targets synthesized with DTFs measured from ears other than their own. The results extend the previous observations by demonstrating particular patterns of errors that result from use of DTFs from larger or smaller subjects. Listeners showed systematic overshoots or undershoots in the lateral dimension when listening through DTFs from larger or smaller subjects, respectively. In the polar dimension, when listeners localized with DTFs from larger subjects, they tended to localize down-front targets to down-rear and they tended to localize up-rear targets to up-front. When listeners localized through DTFs from smaller subjects, the most conspicuous error was a tendency to displace low targets upward. Possible explanations for those patterns of errors are considered in Sec. III B. *Second*, the results support the working hypothesis that, when a listener localizes through other subjects' DTFs, his or her virtual localization can be

improved by scaling the DTFs in frequency to minimize the spectral difference between those DTFs and the listener's own DTFs. In two-thirds of cases, the penalty for use of nonindividualized DTFs was reduced by more than half. Section III C considers the feasibility of tailoring a generic set of DTFs to individual listeners.

A. Relation to previous results

Localization performance in the free-field condition was similar to that in our previous study (Makous and Middlebrooks, 1990). The rate of quadrant errors was comparable to the rate of ''front/back confusions'' in that study. The previous study did not quantify rates of quadrant errors in specific quadrants, but Fig. 5 in that paper showed that quadrant errors were most common in response to up-rear targets. In the own-ear virtual condition, performance was similar to that in the study by Wightman and Kistler (1989b). Differences in the methods of analysis preclude quantitative comparisons, but performance errors in that study seem to be of roughly the same magnitude as in the current study, and that study showed, as in the current study, that localization errors tended to be largest for targets in the up-rear quadrant.

Several previous studies have examined virtual localization under an other-ear condition. The study most similar to the current one is by Wenzel and colleagues (1993). Those authors studied localization by 16 listeners who all heard targets synthesized from the HRTFs of a 17th subject (subject SDO, from the study by Wightman and Kistler, 1989a). Compared with own-ear data from Wightman and Kistler (1989b), subjects in the study by Wenzel and colleagues showed substantial increases in the rates of front/back and up/down confusions under the other-ear condition. Similar results were reported by Morimoto and Ando (1980), who tested three listeners that were selected for small, medium, and large external ears. Each listener localized most accurately in the own-ear condition. In the other-ear condition, each listener showed reasonable performance in the lateral dimension but an increase in the front/back confusions and in the scatter of responses in the vertical dimension.

Møller and colleagues (1996) tested virtual localization in a condition in which listeners distinguished among a limited number of virtual sources. Listeners showed a significant increase in errors in the vertical dimension in the other-ear compared to own-ear condition. Some of the most common errors resembled those observed in the present study. Specifically, down-front targets often were judged as back, characteristic of small listeners localizing through larger subjects' DTFs in the current study, and sources on or below the horizontal plane often were judged with an upward elevation bias, like the large listeners localizing through smaller subjects' DTFs. The rates of quadrant errors in the study by Møller and colleagues were much lower than in the present study, but the stimuli in that study were substantially longer than in the current study: 5 s of speech sounds, compared to 250 ms of Gaussian noise.

Butler and Belendiuk (1977) and Bronkhorst (1995) employed spatial discrimination tasks in which subjects discriminated among a limited number of virtual-target locations. Bronkhorst's ''confusion'' protocol tested the accuracy of identification of target quadrants. Surprisingly, there was no significant difference between own-ear and other-ear conditions. Bronkhorst speculated that the relatively high confusion rates among his inexperienced listeners might have overwhelmed any difference among conditions. Also, Bronkhorst's subject population might have been more uniform than in the present study, inasmuch as it consisted of seven males and only one female. Thus inter-subject differences among sets of HRTFs in that study might have been lower than in the present study. Butler and Belendiuk (1977) reported a result that is often cited as evidence that a listener can localize more accurately with the HRTFs of an accurate localizer than with his or her own. In that study, four listeners each attempted to localize five targets by listening to tapes of microphone recordings from their own ears and from the ears of the three other listeners. Three of the listeners tended to localize more accurately in the own-ear condition. In contrast, the fourth listener (''P.B.'') localized better when listening to the tapes from other listeners than when listening to his/her own tapes. Likewise, the other three listeners showed their worst performance when listening to P.B.'s tape. One interpretation of that result is that the ears of listener P.B. somehow failed to produce useful cues to elevation and that P.B. could localize better by listening through ears that produced more reliable cues. An alternate interpretation is that there might have been a technical problem with the recording of P.B.'s tape, such as some uncertainty regarding the positions of recording microphones in the ears. Also, stimulus spectra in that study were limited to frequencies below 9 kHz, which might have eliminated or distorted high-frequency spectral cues.

In a preliminary report, Wightman and Kistler (1993) described an analysis of HRTFs from 15 listeners. Listeners localized virtual targets that were synthesized with each other's HRTFs. The differences among subjects' HRTFs were characterized by their Euclidean distances. Euclidean distance is similar to the spectral difference metric used in the present study, except that the metric used here eliminated frequency-invariant offsets in DTFs for individual locations. Consistent with the present results, performance in the Wightman and Kistler study generally was more accurate when listeners used HRTFs that were closer to their own in Euclidean distance than when they used HRTFs that were more distant.

B. Implications for models of sound localization

Most models of sound localization acknowledge that listeners' localization judgements are derived from some combination of interaural difference cues and spectral shape cues (see Middlebrooks and Green, 1991, for review). Our procedure for scaling DTFs in frequency involved changing the output rate of the digital-to-analog converter, which resulted in a concomitant upward scaling in frequency of DTFs and of interaural difference spectra and a shortening of interaural delays (or downward scaling in frequency and lengthening of delays). For that reason, the design of the study did not lend itself to independent manipulation of specific localization cues. In another sense, however, this equivalent scaling of

multiple localization cues was ideal for the purpose of aligning DTFs functions among subjects. Results of the companion study indicated that optimal frequency scale factors for DTFs were very highly correlated with the optimal scale factors for interaural difference spectra ($r=0.96$) and showed substantial correlation with the inter-subject ratios of maximum interaural delays ($r=0.71$). That is, adjustment of a single parameter, the output rate, could simultaneously minimize inter-subject differences in spectral shape cues, interaural level differences, and interaural delays.

In the lateral dimension, listeners tended to overshoot the target (i.e., point too far to the side) when listening through the DTFs of a larger subject and tended to undershoot when listening through the DTFs of a smaller subject. Scaling DTFs in frequency resulted in a significant reduction in lateral biases. Nevertheless, the typical increase in magnitude of the lateral bias that resulted from use of another subject's DTFs, scaled or not, was fairly small compared to the scatter in the lateral component of localization judgements. That is, the mean change in unsigned magnitudes of biases was 2.7°, whereas the rms average of lateral errors in the own-ear condition was 14.7°.

In the polar dimension, interaural differences tend to be ambiguous, so listeners are thought to rely on spectral shape cues for localization in that dimension. Models that attempt to explain the use of spectral shape cues generally assume, either explicitly or implicitly, that a listener has some knowledge of the characteristics of his or her own DTFs that are associated with particular sound-source locations. Some models have emphasized the importance of specific spectral features, such as ''boosted bands'' (Blauert, 1969/70), ''covert peaks'' (Humanski and Butler, 1988), or other spectral peaks, notches, or slopes (e.g., Hebrank and Wright, 1974; Bloom, 1977; Watkins, 1978). When listeners attempted to localize narrow-band noise bursts in a previous study (Middlebrooks, 1992), their localization judgements in the polar dimension were determined more by the center frequency of the noise burst than by the actual location of the source. The location that a listener reported in response to a particular center frequency tended to vary with the physical height of the listener. That paper reported that ''in general, the spatial distributions of responses were most similar among subjects of similar physical size, and a given distribution of responses at a particular frequency observed for a tall subject tended to be found at a higher frequency for a shorter subject (Middlebrooks, 1992, p. 2611).'' Based on the companion study (Middlebrooks, 1999), one would account for that result by noting that the spectral features found in DTFs of smaller subjects tend to be found at higher frequencies relative to those of larger subjects.

As an alternative to examining specific spectral peaks and notches, several investigators have compared HRTFs or DTFs with proximal stimulus spectra across relatively broadbands (Watkins, 1978; Middlebrooks, 1992; Zakarauskas and Cynader, 1993; Hofman and Van Opstal, 1998). Studies that used narrow-band free-field stimuli (Middlebrooks, 1992) and broadband virtual stimuli (the present study) both show that localization judgements tend to fall near locations at which the associated DTF most closely resembles the proximal stimulus. Previous localization studies have reported large ''front/back'' or ''up/down'' confusions, which are referred to here as ''quadrant errors.'' The variance plots in Figs. 14, 15, and 17 demonstrate a likely explanation for the quadrant errors in that the plots often showed two widely separated local minima. A discontinuous distribution of localization judgements, with some judgements near the target and others far away, might reflect the dual minima in variance plots.

In the companion study, a residual inter-subject spectral difference in DTFs of around 6.2 dB^2 remained after scaling DTFs in frequency by an optimal scale factor. That residual spectral difference was largely independent of the initial frequency disparity among the DTFs. Inspection of DTFs indicates that frequency scaling is successful in aligning major spectral features along the frequency axis, but scaling can have no effect on idiosyncratic features such as the detailed shapes or magnitudes of peaks or notches. In the current behavioral results, the percentage improvement in localization performance that resulted from scaling DTFs in frequency often was greater than the percentage reduction in spectral difference. That suggests that the general distribution of spectral features in frequency is more important for virtual localization than is the detailed shapes of those features. That conclusion is supported by the work of Kistler and Wightman (1992), who attempted to reconstruct HRTFs from small numbers of principal components. They found that reconstructions based on as few as three components supported some degree of virtual localization in front/back and up/down dimensions and that reconstructions based on only five components supported virtual localization that was nearly as accurate as that obtained with intact HRTFs. Similarly, Kulkarni and Colburn (1998) found that a veridical virtual synthesis could be obtained with highly smoothed HRTFs.

C. Tailoring generic DTFs to specific users

The present results confirm observations by others that performance in virtual localization is most accurate when listeners use DTFs measured from their own ears. Nevertheless, it is not feasible to make detailed DTF measurements from every listener who might want to utilize a virtual acoustic environment. Situations are considered here in which it might or might not be acceptable to use generic ''nonindividualized'' DTFs, and practical procedures are evaluated for customizing generic sets of DTFs to individual listeners.

In evaluating the potential impact of the use of nonindividualized DTFs for a particular application, one must consider the requirements of the application. In the horizontal dimension, the liability for use of nonindividualized DTFs for horizontal localization is relatively small. In the data, the rms errors in the horizontal dimension averaged 17.1° in the other-ear condition compared to 14.7° in the own ear condition, an increase in error of only 16%. The largest increases in lateral bias in the other-ear condition were around 8°, which is less than 10% of the entire range of lateral angle from 0° to 90°. Moreover, the largest lateral biases were observed only in cases in which the largest listeners localized through DTFs from the smallest subjects, and *vice versa*.

Lateral bias could be minimized by maintaining sets of DTFs from one large subject and from one small subject and choosing the appropriate set for use by large or small listeners. In addition, Shinn-Cunningham and colleagues (1998) have shown that, given feedback, listeners can adapt to modified horizontal localization cues. In applications that require only localization of virtual targets that are known to be restricted to the frontal horizontal plane, nonindividualized DTFs might be entirely adequate.

In contrast to the results in the horizontal dimension, localization in the vertical and front/back dimensions showed, on average, a quadrupling of the rate of quadrant errors and a sizeable increase in rms local polar errors when listeners localized through other subjects' DTFs. It is difficult to imagine an application in which such degradation in performance would be acceptable. Our results from scaling of DTFs suggest two ways in which one might improve virtual localization with nonindividualized DTFs. One possibility would be to maintain a single set of DTFs from one listener and to adapt those DTFs to individual listeners by scaling in frequency. Alternatively, one might maintain sets of DTFs from multiple subjects, indexed by a mean frequency scale factor. An individual listener would select the optimal set of DTFs on the basis of his or her own mean frequency scale factor. The difficulty with either of these schemes is that both require knowledge of a mean frequency scale factor for each listener. That is, if one has DTF measurements from subjects A and B, one can scale A to match B, but one does not know the scale factor needed to scale A to another subject, subject C, from whom there are no acoustical measurements.

One way to obtain a frequency scale factor might be to estimate it from physical measurements of the width of a subject's head and the height of his or her external ear. The companion paper compared scale factors estimated from those two parameters with scale factors measured from full sets of DTFs from pairs of subjects. The rms error in estimation was only 5.8%. Alternatively, one might attempt to identify an optimal scale factor behaviorally by conducting virtual localization trials with a range of scale factors. That approach, when tested with virtual targets restricted to the vertical midline, yielded scale factors that differed from acoustically measured scale factors by an rms value of 5.2%. That accuracy is only slightly better than the estimates obtained from physical measurements of the head and ear. Moreover, the behavioral approach required about 2 h of testing per listener.

It is not yet clear how precisely a frequency scale factor would need to be measured in order to obtain a significant improvement in behavioral performance. When virtual localization was tested for targets restricted to the vertical midline, the rate of quadrant errors increased rapidly as the frequency scale factor was varied away from the optimal value (e.g., Fig. 16), with widths of the tuning curves ranging from 4% to 22% across listeners. This suggests that an error of, say, 6% in estimating a frequency scale factor might produce a small but noticeable degradation in virtual localization performance. Nevertheless, in acoustical measurements from a population of 45 subjects (Middlebrooks, 1998), the scale factors obtained from every pairwise combination of subjects ranged from 0.72 to 1.38. Therefore, an estimate of a scale factor with an rms error of 6% (i.e., a factor of 1.06) would substantially reduce the uncertainty in the scale factor.

One final possibility that we have not evaluated is that, with experience, subjects might adapt to another subject's DTFs. In the present study, own-ear, other-ear, and/or scaled-ear trials normally were interleaved within a test session, so a listener seldom heard more than two or three DTFs in a row from the same set. That issue was addressed to some extent by Møller and colleagues (1996), who compared two other-ear conditions: one in which all the HRTFs were from one subject and the other in which HRTFs from several subjects were interleaved. They observed no significant difference in performance in the two conditions. Another issue is one of feedback. All of the trials in the present study were conducted without trial-by-trial feedback, so listeners had no means of associating particular other-ear HRTFs with correct locations. Shinn-Cunningham and colleagues (1998) reported that, when provided with trial-by-trial feedback, subjects could adjust to modified horizontal localization cues. A recent report by Hofman and colleagues (1998) tested the ability of listeners to adapt to long-term modifications of the acoustics of their own external ears. Subjects wore plastic inserts in their external ears. Initially after placement of the inserts, the listeners showed dramatic errors in free-field localization. After wearing the inserts for a period of 6 weeks, however, the listeners' free-field localization performance improved substantially. The changes in HRTFs introduced by the plastic inserts almost certainly were larger than the abnormalities in HRTFs that would be encountered by a listener who used nonindividualized HRTFs that were scaled imperfectly in frequency.

D. Concluding remarks

The companion study (Middlebrooks, 1999) demonstrated that, for any pair of subjects, scaling in frequency with a single scale factor can reduce substantially the intersubject differences in DTFs for locations throughout the coordinate sphere. The present study exploited that result to attempt to improve virtual localization in conditions in which a listener uses DTFs measured from another subject's ears. The results of that exercise actually were better than expected, in that the percentage improvements in localization performance often were greater than the percentage reduction in spectral difference between sets of DTFs. The sign of the frequency difference between the DTFs through which a listener localized and the listener's own DTFs turned out to be predictive of particular types of localization errors. That is, localization errors differed depending on whether spectral features in DTFs generally lay at higher or lower frequencies than spectral features in a listener's own DTFs. Those localization errors can be accounted for qualitatively by assuming that listeners' localization judgements tend to correspond to locations for which their own DTFs best resembles the DTF that was used to synthesize the virtual target.

The companion study showed that optimal frequency scale factors between pairs of subjects can be predicted with some accuracy from certain physical dimensions of subjects,

and the present behavioral results suggest that those predictions likely are accurate enough to provide appreciable improvements in virtual localization. That result offers encouragement for the use of nonindividualized DTFs. Specifically, in a setting in which one had no facility for measuring an individual listener's DTFs, one could estimate a frequency scale factor from measurements of the width of the listener's head and the height of the external ear. That scale factor then would be used to tailor a generic set of DTFs so that they approximate the listener's own DTFs fairly closely. Initial errors in virtual localization would be corrected with feedback. One would anticipate that this procedure would result in a realistic virtual synthesis of auditory space for the large number of listeners for whom it would not be practical to make individual acoustical measurements of DTFs.

ACKNOWLEDGMENTS

I am pleased to acknowledge the contributions of Zekiye Onsan, who collected the behavioral data and made the illustrations. Shigeto Furukawa, Ewan Macpherson, Brian Mickey, and Li Xu provided stimulating discussion and commented on an early draft of the manuscript. I thank Professor Frederic Wightman and a second anonymous reviewer for constructive comments on an earlier draft. This research was supported by ONR Grant No. N0014-94-1-0481 and NIH Grant No. DC000420.

Blauert, J. (**1969–1970**). ''Sound localization in the median plane,'' Acustica **22**, 205–213.

Bloom, P. J. (**1977**). ''Creating source elevation illusions by spectral manipulation,'' J. Audio Eng. Soc. **25**, 560–565.

Bronkhorst, A. W. (**1995**). ''Localization of real and virtual sound sources,'' J. Acoust. Soc. Am. **98**, 2542–2553.

Butler, R. A., and Belendiuk, K. (**1977**). ''Spectral cues utilized in the localization of sound in the median sagittal plane,'' J. Acoust. Soc. Am. **61**, 1264–1269.

Carlile, S., Leong, P., and Hyams, S. (**1997**). ''The nature and distribution of errors in sound localization by human listeners,'' Hearing Res. **114**, 179–196.

Foster, S. (**1986**). ''Impulse response measurement using Golay codes,'' in *IEEE 1986 Conference on Acoustics, Speech, and Signal Processing, Vol. 2* (IEEE, New York), pp. 929–932.

Golay, M. J. E. (**1961**). ''Complementary series,'' IRE Trans. Inf. Theory **7**, 82–87.

Hebrank, J., and Wright, D. (**1974**). ''Spectral cues used in the localization of sound sources on the median plane,'' J. Acoust. Soc. Am. **56**, 1829–1834.

Hofman, P. M., and Van Opstal, J. A. (**1998**). ''Spectro-temporal factors in two-dimensional human sound localization,'' J. Acoust. Soc. Am. **103**, 2634–2648.

Hofman, P. M., Van Riswick, J. G. A., and Van Opstal, J. A. (**1998**). ''Relearning sound localization with new ears,'' Nature Neurosci. **1**, 417–421.

Humanski, R. A., and Butler, R. A. (**1988**). ''The contribution of the near and far ear toward localization of sound in the sagittal plane,'' J. Acoust. Soc. Am. **83**, 2300–2310.

Kistler, D. J., and Wightman, F. L. (**1992**). ''A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction,'' J. Acoust. Soc. Am. **91**, 1637–1647.

Kulkarni, S., and Colburn, H. S. (**1998**). ''Role of spectral detail in sound-source localization,'' Nature (London) **396**, 747–749.

Makous, J. C., and Middlebrooks, J. C. (**1990**). ''Two-dimensional sound localization by human listeners,'' J. Acoust. Soc. Am. **87**, 2188–2200.

Middlebrooks, J. C. (**1992**). ''Narrow-band sound localization related to external ear acoustics,'' J. Acoust. Soc. Am. **92**, 2607–2624.

Middlebrooks, J. C. (**1998**). ''Individual diffrences in external-ear transfer functions reduced by scaling in frequency,'' J. Acoust. Soc. Am.

Middlebrooks, J. C., and Green, D. M. (**1991**). ''Sound localization by human listeners,'' Annu. Rev. Psychol. **42**, 135–159.

Middlebrooks, J. C. (**1999**). ''Individual differences in external-ear transfer functions reduced by scaling in frequency,'' J. Acoust. Soc. Am. **106**, 1480–1492.

Middlebrooks, J. C., and Green, D. M. (**1990**). ''Directional dependence of interaural envelope delays,'' J. Acoust. Soc. Am. **87**, 2149–2162.

Møller, H., Sorensen, M. F., Jensen, C. B., and Hammershoi, D. (**1996**). ''Binaural technique: Do we need individual recordings?,'' J. Aud. Eng. Soc. **44**, 451–469.

Morimoto, M., and Ando, Y. (**1980**). ''On the simulation of sound localization,'' J. Acoust. Soc. Jpn. (E) **1**, 167–174.

Morimoto, M., and Aokata, H. (**1984**). ''Localization cues of sound sources in the upper hemisphere,'' J. Acoust. Soc. Jpn. (E) **5**, 165–173.

Pralong, D. and Carlile, S. (**1996**). ''The role of individualized headphone calibration for the generation of high fidelity virtual auditory space,'' J. Acoust. Soc. Am. **100**, 3785–3793.

Shinn-Cunningham, B. G., Durlach, N. I., and Held, R. M. (**1998**). ''Adapting to supernormal auditory localization cues. I. Bias and resolution,'' J. Acoust. Soc. Am. **103**, 3656–3666.

Watkins, A. J. (**1978**). ''Psychoacoustical aspects of synthesized vertical locale cues,'' J. Acoust. Soc. Am. **63**, 1152–1165.

Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. (**1993**). ''Localization using nonindividualized head-related transfer functions,'' J. Acoust. Soc. Am. **94**, 111–123.

Wightman, F. L., and Kistler, D. J. (**1989a**). ''Headphone simulation of free-field listening. I: Stimulus synthesis,'' J. Acoust. Soc. Am. **85**, 858–867.

Wightman, F. L., and Kistler, D. J. (**1989b**). ''Headphone simulation of free-field listening. II: Psychophysical validation,'' J. Acoust. Soc. Am. **85**, 868–878.

Wightman, F. L., and Kistler, D. J. (**1992**). ''The dominant role of low-frequency interaural time differences in sound localization,'' J. Acoust. Soc. Am. **91**, 1648–1661.

Wightman, F. L., and Kistler, D. J. (**1993**). ''Multidimensional scaling analysis of head-related transfer functions,'' Proc. ASSP (IEEE) Workshop on Applications of Signal Processing to Audio and Acoustics.

Woodworth, R. S. (**1938**). *Experimental Psychology* (Holt, Rinehart, and Winston, New York), pp. 349–361.

Zakarauskas, P., and Cynader, M. S. (**1993**). ''A computational theory of spectral cue localization,'' J. Acoust. Soc. Am. **94**, 1323–1331.

Zhou, B., Green, D. M., and Middlebrooks, J. C. (**1992**). ''Characterization of external ear impulse responses using Golay codes,'' J. Acoust. Soc. Am. **92**, 1169–1171.

Morphology and development of the human vocal tract: A study using magnetic resonance imaging

W. Tecumseh Fitch[a)]
Speech and Hearing Sciences Program, Harvard/MIT, 33 Kirkland Street, Room 982, Cambridge, Massachusetts 02138

Jay Giedd
National Institute of Mental Health, Child Psychiatry Branch, Building 10, Room 6N240, Bethesda, Maryland 20892-1600

(Received 8 January 1999; accepted for publication 23 April 1999)

Magnetic resonance imaging was used to quantify the vocal tract morphology of 129 normal humans, aged 2–25 years. Morphometric data, including midsagittal vocal tract length, shape, and proportions, were collected using computer graphic techniques. There was a significant positive correlation between vocal tract length and body size (either height or weight). The data also reveal clear differences in male and female vocal tract morphology, including changes in overall vocal tract length and the relative proportions of the oral and pharyngeal cavity. These sex differences are not evident in children, but arise at puberty, suggesting that they are part of the vocal remodeling process that occurs during puberty in males. These findings have implications for speech recognition, speech forensics, and the evolution of the human speech production system, and provide a normative standard for future studies of human vocal tract morphology and development. [S0001-4966(99)02008-1]

PACS numbers: 43.70.Aj, 43.70.Ep [AL]

INTRODUCTION

The morphology of the supralaryngeal vocal tract (hereafter ''vocal tract'') is of fundamental importance in human speech production, because the shape of the vocal tract determines the articulatory possibilities, and thus possible formant patterns in speech (Chiba and Kajiyama, 1941; Stevens and House, 1955; Fant, 1960; Lieberman and Blumstein, 1988; Titze, 1994). Traditionally, the interaction between vocal tract form and speech acoustics has been studied with radiographic techniques (e.g., Fant, 1960; Perkell, 1969; Holbrook and Carmody, 1937), but the potential health hazards of the ionizing radiation required has typically limited these studies to small sample sizes both of subjects and articulatory positions. With the advent in the last decade of commercially available magnetic resonance imaging (MRI), which uses reversals in strong magnetic fields to provide high-quality anatomical images with no known health risks, a number of studies using MRI to investigate vocal tract morphology have appeared (e.g., Baer *et al.*, 1991; Story *et al.*, 1996; Sulter *et al.*, 1992; Moore, 1992). Unlike radiographic studies which have generally focused on bony structures (vertebral column, skull base, hyoid bone, etc.), MRI can clearly image the soft tissues relevant to speech production (glottis, velum, lips, etc.). The results of these studies have essentially validated older radiographic findings, indicating a close correspondence between vocal tract configuration and speech acoustics, as predicted by the acoustic theory of speech (Sulter *et al.*, 1992; Chiba and Kajiyama, 1941; Fant, 1960). Due to the expense of, and limited time available on, MRI scanners, these studies have focused on investigating the detailed relationship between acoustics and morphology in a small number of subjects (one to five).

However, a variety of important questions remain to be answered that require larger subject populations. For instance, it has long been predicted on the basis of comparisons of the acoustic vowel spaces of adults and children that there should be a correlation between body size and vocal tract length (Peterson and Barney, 1952; Fant, 1966; Mattingly, 1966; Lieberman, 1984), but this prediction has never been explicitly tested. Similarly, acoustic differences and scattered radiographic data suggest a significant difference between male and female vocal tracts: Relative to women, males are predicted to have longer vocal tracts overall, as well as longer pharyngeal cavities (Fant, 1966; Peterson and Barney, 1952; Sachs *et al.*, 1973). Resolving these questions requires a larger number of subjects than the few available in past studies.

The current study results from a unique opportunity at the National Institutes of Health to scan a large number of normal children and young adults as part of a broader ongoing study investigating brain development in diverse clinical populations. Details of the methods and aims of this study can be found elsewhere (Giedd *et al.*, 1996). Here we report data on vocal tract morphology from midsagittal scans of 129 normal children and adults, ages 2 to 25. Besides providing normative data for future studies of vocal tract abnormalities, this relatively large data set of normals allows us to focus on the changes in vocal tract anatomy that occur during growth and maturation, and in particular, on sex differences and on the relation between vocal tract length and body size.

a)To whom correspondence should be addressed.

Literature review and hypotheses

One of the abiding problems in speech research is the significant variation in formant patterns observed between different speakers, first clearly documented by Peterson and Barney (1952) and recently verified by Hillenbrand *et al.* (1995). This acoustical variability has ramifications for theories of speech perception, language acquisition, automatic speech recognition, speech forensics, and other fields. It has been customary since Peterson and Barney (1952) to assume that vocal tract anatomy, and particular vocal tract length, play a crucial role in this variability, following the assumption that vocal tract length is correlated with body size (Ladefoged and Broadbent, 1957; Fant, 1960, 1966; Nordström, 1977; Nearey, 1978; Lieberman, 1984). However, the nature and strength of the putative body size/vocal tract length relationship has never been empirically examined [although Fitch (1997) reported a strong positive correlation between vocal tract length and body size in rhesus macaques]. The first goal of the current study is thus to evaluate the hypothesis that vocal tract length is correlated with body size (height and/or weight) in humans.

A second long-standing problem in speech science was first addressed by Fant (1966), who suggested that a uniform scaling of vocal tract length to body size is not enough to account for the observed formant differences in adult men and women. Based on unpublished radiographic data, Fant suggested that men have a disproportionately long pharynx relative to women and children, and that this anatomical difference can adequately account for the remaining acoustic differences between the sexes. This claim engendered a debate concerning the relative importance of anatomical versus cultural factors in explaining male and female ''dialects'' (Mattingly, 1966; Sachs *et al.*, 1973; Nordström, 1975; Fant, 1975; Nordström, 1977), which ended provisionally with most parties conceding that both factors play a role. Unfortunately, a more precise resolution of this issue has been prevented by a lack of data documenting any disproportionate anatomical differences between males and females. Thus a second major goal of the research reported here is to analyze sex differences in vocal tract anatomy, controlling for differences in overall body size.

Finally, if a disproportionate sex difference does exist, when does it originate? Human infants have a larynx in the standard mammalian position, where it can be inserted into the nasal passages to allow simultaneous suckling and breathing (Negus, 1949; Lieberman, 1984; Laitman and Crelin, 1976; Crelin, 1987). Between three months and three years of age, the larynx recedes from this high intranarial location down into the throat. There is little information about the time course and magnitude of this ontogenetic ''descent of the larynx'' from the neonatal to adult position (Sasaki *et al.*, 1977). Radiographic studies made for the purposes of assisting respiratory intubation of patients suggest that vocal tract lengthening occurs shortly after birth and perhaps during puberty as well (Westhorpe, 1987). However, another study (Roche and Barkla, 1964) of eight boys and eight girls found a steady (presumably growth related) descent throughout childhood, and reported no pubertal change. Thus it is currently unclear when putative adult differences originate. Because raters can discriminate the speech of prepubescent boys from girls (Sachs *et al.*, 1973), and boys have been reported to have slightly lower formants than girls (Sachs *et al.*, 1973; Bennet, 1980; Lee *et al.*, 1999), a reasonable hypothesis would be that there is a larger neonatal descent in boys than girls, leading to sex differences carried forward from early childhood throughout life.

An alternative possibility is suggested by the longitudinal radiographic study of pharynx dimensions of King (1952). King took lateral x-rays of the heads of 24 males and 26 females from the age of three months to 16 yrs. His data indicate a gradual lengthening of the pharynx throughout the entire age period and additional slight peripubertal growth spurts in both males and females. However, other researchers have disputed the number, timing, and existence of such growth spurts, with some researchers finding evidence for them and others failing to [see King (1952) for a review]. Goldstein (1980) reanalyzed King's data, along with other scattered data from the medical literature, in the context of vocal production, and pointed out (p. 76, p. 185) that the data available were inadequate to resolve the issue firmly. No statistical analyses were performed by King (1952) or Goldstein (1980) to determine if this ''slight'' pubertal growth spurt differed between males and females, or indeed if the pharyngeal spurt was any greater than expected based simply on the spurt in overall body growth at puberty. King had no data on the pubertal stage of his subjects, and thus based his growth curves on subject age. But because of the high variability in the age at pubertal onset, along with normal variation in body size at different ages, age-based analysis tends to obscure changes tied directly to puberty (see Results). Thus the third hypothesis evaluated in the current study is that there is a secondary ''descent of the larynx'' that occurs at puberty, and that sex differences in this descent could account for any differences observed in adult male and female vocal tract proportions. To overcome the difficulties associated with age-based analysis, we used a standardized rating system of pubertal stage developed by Tanner (1962) and subsequent workers (e.g., Petersen *et al.*, 1988).

I. METHODS

A more detailed description of the methods used to recruit normal volunteers, and the scanning protocols, is given in Giedd *et al.* (1996). Briefly, subjects are recruited from the community and undergo a three part screening process (telephone, parent and teacher questionnaires, and face-to-face interview/examination) in which only one of six initial volunteers is accepted into the study. Exclusion criteria relevant to this report are congenital anomalies, or history of speech delay or language impairment, in the subject or first degree relatives of the subject. All attempts were made to ensure that our subjects were normal in every way. Height (cm) and weight (g) were measured using a Model 595KL electronic scale (Healthometer, Bridgeview, Illinois). Unlike the situation in many older adult populations where excess body fat would be a complicating factor in evaluating the relationship between body size and vocal tract morphology, the current sample contained very few overweight individuals (see Results). Pubertal status was quantified using a self-

administered questionnaire, yielding Tanner ratings of pubertal stage from 1 (pre-pubescent), 2 to 4 (intermediate stages), or 5 (fully mature) based on the development of secondary sexual characteristics (Duke *et al.*, 1980; Morris and Udry, 1980; Petersen *et al.*, 1988). All MRI scans were acquired on the same GE 1.5 Tesla scanner located at the clinical center of the NIH. The imaging sequence was a sagittal spin echo series with time to echo 14 ms, time to repeat 400 ms, field of view 30 cm, 5-mm slice thickness with 1.5-mm gap between slices, acquisition matrix 256×128, and number of excitations $=0.75$. The protocol was approved by the Institutional Review Board of the National Institute of Mental Health.

Previous studies which have sought to correlate acoustic output with vocal tract shape have, by necessity, performed MRIs of subjects attempting to maintain a certain articulatory position without movement for the duration of the scan, a difficult task. These studies document substantial changes in vocal tract length during production of different vowels or voice types (Sulter *et al.*, 1992). In the present study, subjects were asked to lie motionless and to breathe quietly while being scanned. No sedation was used. Virtually all scans showed subjects in a nasal breathing posture, with the tongue dorsum in full contact with the palate (five subjects lacked such contact, but in all these cases the curve of the tongue closely approximated that of the palate). Measuring vocal tract length during quiet respiration avoids errors that would be introduced by differences in articulatory position between subjects. This nonphonetic posture also minimizes the possible confounding effects of cultural factors, as well as the requirement of a special task for subjects (an important concern when scanning young children). Nonetheless, some of the younger subjects found remaining motionless difficult and had to be excluded due to motion artifacts (see below).

Three to five para- and midsagittal images for each patient were transferred from the GE Scanner to a Sun workstation using GINX software, and then transferred via FTP to a Macintosh computer for further processing using NIH Image 1.61 software (developed at the National Institutes of Health and available free on the Internet at http://rsb.info.nih.gov/nih-image/). After importing the images into this program, the single midsagittal slice that most clearly showed the anatomical structures of interest (lips, tongue, velum, glottis) was selected for further analysis. Images in which these structures were not visible ($N=3$), or which showed distortion caused by orthodontic devices ($N=3$) or movements ($N=4$) were excluded. No other preselection was performed, leaving a total of 129 midsagittal images, one per subject, for further analysis.

Measurement of vocal tract dimensions was accomplished in two different ways. Both measures are reported in mm, and both originate at the level of the glottis, at the anterior point of the vocal processes of the arytenoid cartilages (hereafter simply ''glottis''). The glottal level was indicated by one or more of the following: the vocal fold visible beneath the ventricular fold, a supraglottal notch marking the vestibule and anterior commisure, the caudal delimitation of the pre-epiglottic fat, or the border between the cricoid and arytenoid cartilages [see Curtin (1996) and Tucker (1987) for a more detailed introduction to the radiological anatomy of the larynx]. With the ''curved line'' technique, vocal tract length was measured from the glottis, traveling up through the pharyngeal midline and between the tongue and palate, and terminating at a plane touching the upper and lower external borders of the lips (Fig. 1). This curvilinear measurement is acoustically motivated, closely approximating the path taken by longitudinal pressure waves generated at the larynx and emanating from the lips (see, e.g., Fant, 1960). Due to the position of the tongue against the palate, oral vocal tract length could be unambiguously measured (using the ''freehand line'' tool in the NIH Image program). Measurement accuracy was determined by making ten repeated measurements on ten randomly selected subjects.

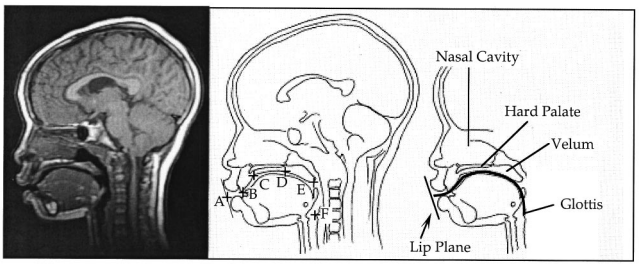


FIG. 1. Left: Typical midsaggital MRI image (this is a 5 year old girl). Middle: Illustration of landmarks used in the ''line segment'' method: A. Bilabial plane; B. Tongue tip; C. Alveolar ridge; D. Hard palate/velar junction; E. Uvula; F. Glottis. The solid line indicates the ''line segment'' method of vocal tract measurement. See Fig. 3 for labels of the line segments themselves. Right: Illustration of major landmarks, with the ''curved line'' technique of vocal tract length measurement illustrated by the solid line.

However, most previous reports have used a series of line segments to measure vocal tract length (e.g., Sulter *et al.*, 1992; Dmitriev and Kiselev, 1979; Goldstein, 1980), or even simple linear measures of the oral or pharyngeal cavities (e.g., King, 1952). In keeping with this tradition, we developed a second ''line segment'' technique which compartmentalized the vocal tract into a series of anatomically defined subsections, thus providing more detail about vocal tract morphology (although less accurate total length information). Five line segments were defined via six cusp points, as described in Fig. 1. The cusp points were recorded using the Object–Image extension to NIH Image (available on the Internet via ftp: simon.bio.uva.nl). Each of the points is intended to provide meaningful information from the point of view of speech production, each tube being controlled primarily by independent articulators. For example, the lip tube is lengthened by lip protrusion, corresponding to the phonetic feature of rounding seen in some vowels. Similarly, the pharyngeal tube is shortened by laryngeal raising. While three of these cusp points [lips, glottis, and uvula (free edge of velum)] were always clearly defined; others (tongue tip, alveolar ridge, and hard palate/velar junction) were less clear in some scans; thus the data for these midvocal tract landmarks may be less reliable. The length of the segments connecting these points, and the angles between them, were recorded and saved in a spreadsheet for further processing.

Statistical analyses were carried out using Statview 4.5.

TABLE I. Ages (years), weights (kg), heights (cm), and vocal tract lengths (mm), measured by the curved line technique, for all subjects, broken down by age group, and by sex above age 14 yrs. ''All ages'' provides the means for the entire subject pool combined.

		Wt. (kg)	Ht (cm)	VTL (mm)				
	N	Mean	Mean	Mean	Std. Dev.	Std. Error	Minimum	Maximum
All ages	129	41.8	145.8	126.1	17.6	1.5	91.4	166.5
2 to 4 (yrs)	9	17.9	104.8	99.2	6.5	2.2	91.4	110.2
5 to 6	18	21.1	115.9	105.4	6.1	1.4	96.6	118.0
7 to 8	15	28.1	132.0	115.9	6.3	1.6	104.7	129.2
9 to 10	24	35.0	142.8	120.9	7.0	1.4	111.5	139.4
11 to 12	19	42.9	151.1	130.0	5.7	1.3	122.3	140.2
13 to 14	15	55.6	165.4	139.2	8.7	2.2	126.3	157.0
15 to 16	10	60.1	168.3	141.4	7.0	2.2	131.1	150.9
17 to 18	9	71.0	172.9	151.7	9.0	3.0	140.9	166.5
19 to 25	10	70.3	178.4	150.4	8.9	2.8	135.2	160.0
Separated by sex:								
15 to 16 yrs Female	5	55.5	162.2	136.7	5.8	2.6	131.1	144.0
15 to 16 yrs Male	5	64.8	174.5	146.2	4.5	2.0	139.6	150.9
17 to 18 Female	3	74.1	170.2	143.7	3.5	2.0	140.9	147.6
17 to 18 Male	6	69.4	174.2	155.7	8.1	3.3	142.1	166.5
19 to 25 Female	3	61.4	173.1	138.8	3.4	1.9	135.2	141.9
19 to 25 Male	7	74.1	181.6	155.4	4.4	1.7	146.9	160.0

(Abacus Concepts, Berkeley, CA). Both least squares and reduced major axis (RMA) regression slopes are given; the former is appropriate for evaluating predictive relationships (e.g., predicting vocal tract length from body size), while the latter is more appropriate for evaluating ontogenetic scaling relationships [see LaBarbera (1989) for a review].

II. RESULTS

A total of 129 scans, one per subject, were analyzed in this study. These included scans of 53 females and 76 males, with an age range from 2.8 years to 25 years. Mean age was 11.5 years for males and 11.6 years for females. The accuracy of our measurement techniques was high: standard errors for 10 repeated measures of 10 randomly chosen subjects ranged from 0.23 to 0.46 mm (mean 0.32 mm), with the maximal deviations between measurements for any one individual no more than 3% of the total length.

A linear regression analysis comparing vocal tract length as measured by the ''curved line'' versus ''line segment'' techniques showed that the two techniques gave highly correlated results ($r=0.983$, $p<0.0001$). The slope of the regression line was 1.02, not differing significantly from the expected slope of 1.0 (95% confidence intervals: 0.98 to 1.05). The intercept of the regression line on the ''line segment'' axis was 3.6 mm, indicating that the ''line segment'' technique produces slightly higher estimates of VTL (by 4–8 mm). All further analyses give equivalent results regardless of which measure is analyzed, so hereafter we will report the curved line vocal tract length results (''VTL'') unless otherwise noted.

Table I gives the ages, weights, heights, and ''curved line'' vocal tract lengths for all subjects, broken down by age group. The groups were chosen to provide an approximately equal number of subjects per group. Unpaired *t*-tests at each age revealed no significant sex differences before age 15 years, so these data are not separated in the table. However, after 15 years, average vocal tract length diverged significantly between males and females ($p<0.05$), so data are presented separately for these age classes in Table I. Table II gives results of a similar analysis of the individual vocal tract segments.

A. Vocal tract length and body size

A linear regression analysis revealed a strong correlation between VTL and total body length (=''height'') ($r=0.926$, adj. $r^2=0.86$, $p<0.0001$), shown graphically in Fig. 2(a). Due to missing height data, only 121 subjects were available for this analysis. The 0.68 slope of this regression line indicates that VTL does not increase as rapidly as body size, as expected given the disproportionately large size of a child's head relative to an adult's.

Because volume is proportional to the cube of length, it is customary to use logarithmically transformed data in allometric comparisons of length and mass, thus converting the cubic relation back to a linear one. Linear regression analysis revealed a strong relationship between $\log_{10}$ body mass and $\log_{10}$ VTL ($N=129$, $r=0.941$, adj. $r^2=0.89$, $p<0.0001$), shown graphically in Fig. 2(b). This is comparable to the strong relationship between VTL and height. This similarity is unsurprising given the strong relationship between height and $\log_{10}$ body weight in our sample ($N=121$, $r=0.952$, adj. $r^2=0.91$, $p<0.0001$), which verifies our impression that our subject pool contained few overweight individuals (which would have weakened this correlation). VTL was also correlated with age (adj. $r^2=0.80$), although this correlation was weaker than the size correlations.

Polynomial regression of non-log-transformed VTL versus body weight using either second- or third-order polynomials gave comparable results ($N=129$, second order: adj. $r^2=0.878$; third order: adj. $r^2=0.877$) indicating that the

TABLE II. Age category breakdown of individual vocal tract segment lengths. ''All ages'' provides the means for the entire subject pool combined.

Age (yrs)	Lip	Blade	Dorsum	Velum	Pharynx	Total (seg)
All ages:	14.0	21.3	24.4	29.9	42.3	131.9
2 to 4	12.2	17.6	22.2	25.4	26.3	103.6
5 to 6	13.3	18.0	21.6	25.7	32.1	110.8
7 to 8	13.6	20.7	22.2	28.4	35.9	120.9
9 to 10	13.7	21.3	24.1	28.5	39.2	126.8
11 to 12	13.8	22.0	25.1	29.9	43.6	134.5
13 to 14	15.3	22.0	27.6	33.4	48.8	147.2
15 to 16	14.5	22.0	25.6	33.1	52.5	147.8
17 to 18	15.5	24.1	26.9	33.9	55.3	155.8
19 to 25	14.0	25.3	26.1	35.0	56.4	156.7
Separated by sex:						
F, 15–16	13.2	20.8	26.5	32.0	49.1	141.5
M, 15–16	15.7	23.3	24.8	34.3	56.0	154.0
F, 17–18	14.4	24.0	28.0	33.3	53.2	152.9
M, 17–18	16.7	25.2	26.4	34.7	57.8	160.8
F, 19–25	12.8	24.8	24.6	37.2	46.9	146.4
M, 19–25	14.5	25.5	26.8	34.0	60.4	161.2

linear regression of the log-transformed data adequately accounts for the variance in this data. Accordingly no further polynomial regression results will be reported in this study.

Finally, we used an analysis of covariance (ANCOVA) to examine the effects of other variables on VTL with the effect of body size removed. An ANCOVA using body height, age, and sex as independent variables indicated that most variance in VTL was explained by height ($F=71.7$, $p<0.001$), but that significant additional variance was explained by age ($F=16.9$, $p<0.001$) and sex ($F=14$, $p=0.0003$). No significant interactions among these variables were found. Thus as suggested by Fant (1966), there are sex- and age-related differences in vocal anatomy that go beyond those explainable simply by body size. Accordingly, we now turn to an analysis of the relationship among VTL, age, and sex.

B. Development of vocal tract morphology

A preliminary analysis of the data indicated that substantial changes in vocal tract morphology occur at puberty. Because of variation in age at puberty, these changes would be obscured by age-based analyses like those in Tables I and II. We thus used our data on the subjects' pubertal stage (Tanner ratings) to divide the dataset into three groups: pre-, peri-, and postpubertal. The transition values from our data fit well with recent national averages (Herman-Giddens *et al.*, 1997; Biro *et al.*, 1995). The onset of puberty (Tanner stage 2) occurred at a mean age of 10.3 years (s.d.=1.7 years). For most analyses, all children under this age were classed as ''average prepubertal'' ($N=54$). However, some younger children reported being at Tanner stage 2 (the youngest was 8.1 years). Although we suspected that these reports were likely the least reliable in our dataset, we also performed analyses using an ''exclusive prepubertal'' category consisting of only children reporting Tanner stage 1 ($N=31$, age 2.8–8.1 years). The mean age of the final adolescent stage (Tanner stage 4, 14.5 years) was used to define the upper bound for the ''peripubertal'' category ($N=39$, age 10.3–14.54 years). All remaining subjects were considered ''postpubertal'' and sexually mature ($N=33$, age 14.7–25.1 years, mean age 17.9 years).

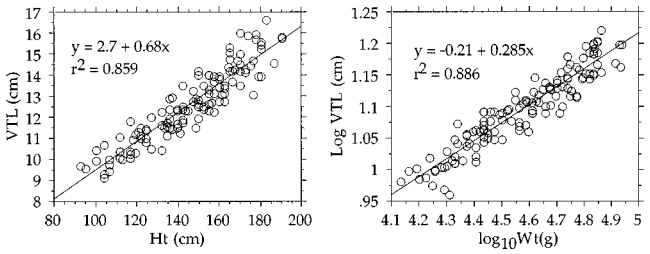


FIG. 2. Correlations between (a) (left) vocal tract length (cm, measured with the ''curved line'' method) and body length [''height (cm)''] and (b) (right) log vocal tract length with log body mass (g).

Linear regression analyses revealed no significant differences in the regression equations for the three groups. In particular, no differences were observed between the slope or intercept of the ''exclusive prepubertal'' versus the ''average prepubertal'' category; hence we will report the average prepubertal values in the analyses below. A factorial one-way ANOVA with pubertal category as the independent variable revealed that VTL varied significantly between the groups ($N=126$, $F=143$, $p<0.001$); Post hoc tests showed significant differences between all three pubertal categories (Fisher's PLSD, all $p<0.001$).

To determine the specific morphological locus of the observed maturational increase in VTL, we analyzed each segment of the vocal tract as measured using the ''line segment'' technique. Mean values and results of one-way factorial ANOVAs for each segment are given in Table III. A graphic representation of the mean segment lengths and the mean angles between each segment is provided in Fig. 3, which also makes clear that there are no appreciable changes in segmental angles in this dataset. These data show that,

TABLE III. Mean values and ANOVA results showing changes in lengths of individual segments of the vocal tract across the three age groups. ''Pre-'' pubertal children are less than 10.3 years in age, ''Post-'' pubertals are older than 14.5, and ''Peri-'' pubertals are in between.

	Mean Lengths (mm)			ANOVA		Fisher's PLSD p values		
Segment	Pre-	Peri-	Post-	F	p	Pre/Peri	Pre/Post	Peri/Post
Lip	13.4	14.3	14.7	5.1	0.0077	0.04	0.003	0.39 NS
Blade	19.2	22.3	23.5	21.3	0.0001	0.0001	0.0001	0.14 NS
Dorsum	22.6	25.7	26.2	15.1	0.0001	0.0001	0.0001	0.50 NS
Velum	27.0	30.9	33.7	36.5	0.0001	0.0001	0.0001	0.0011
Pharynx	33.9	43.8	54.9	137.0	0.0001	0.0001	0.0001	0.0001

although significant length increases occur across childhood and puberty in all portions of the vocal tract, only the velum and pharynx enlarge disproportionately during early adulthood. There is a disproportionate increase in pharyngeal length in both transitions. When each segments' change in length is scaled by its average length, the lip, blade, dorsum, and velum segments enlarge by an average of 12% (range 6%–14%) between childhood and puberty, while pharynx length increases by 22%. Similarly, between puberty and adulthood, the upper portions of the vocal tract grows by an average of only 5% (range 3%–9%) while the pharynx increases its length by 25%. This disproportionate change in pharynx length is most pronounced in males (see below).

C. Sex differences in vocal morphology

We used unpaired t-tests to compare male and female vocal tract lengths. When subjects of all ages were included in the analysis, we found no significant difference in VTL ($t=1.22$, $p=0.23$) (as reported by Lieberman and McCarthy, in press). However, an analysis of each of the age groups defined above revealed that, while prepubertal boys and girls have no significant difference in vocal tract length ($t=1.37$, $p=0.18$), peripubertal subjects showed a small but significant sex difference ($t=2.54$, $p=0.015$), with males 7.5 mm longer on average. Finally, postpubertal subjects showed a highly significant sex difference ($t=5.5$, $p<0.0001$), with males averaging 12.9 mm longer than females. These data indicate that the greater length of the male vocal tract, frequently noted by earlier investigators, has its origin at puberty.

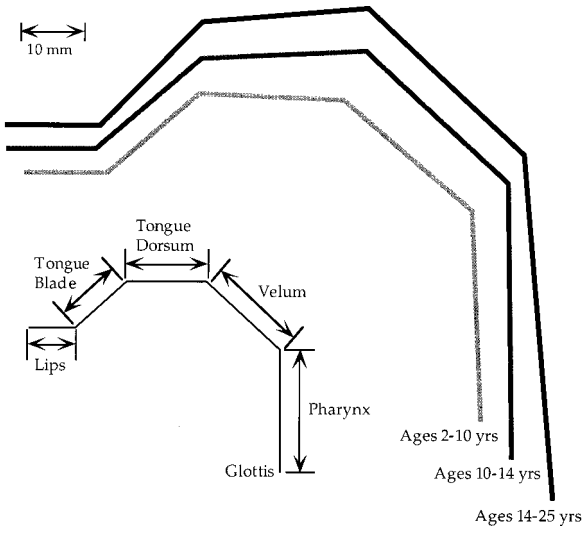


FIG. 3. Averaged vocal tract morphology (as measured by the ''line segment'' method) for pre-, peri-, and postpubescent children. Both the segment lengths and their angles have been averaged to create these figures.

To extend and refine this analysis, we first performed a two-way factorial ANOVA with vocal tract length as the dependent variable and sex and age group as independents. Unsurprisingly, there was a strongly significant effect of age group ($F=142$, $p<0.0001$) as well as a significant main effect for sex, with males having longer tracts, once age was factored out ($F=9.8$, $p=0.002$). Finally, there was a significant interaction ($F=8.4$, $p=0.004$) between the factors. To help further understand the morphological changes underlying these results, we performed a MANOVA with all of the individual segment lengths, again using sex and age group as independent variables. The results, given in Table IV, show that while all segments increase length significantly with age, there are significant sex differences in only the lip and pharynx segments. However, the pattern exhibited by these two segments over time is quite different: there was a significant interaction between sex and age group only for pharynx length (this was also true in a three-factor MANCOVA including height, sex and age group, $F=7.7$, $p=0.0001$). This is because the pharyngeal segment showed no significant difference before puberty ($p>0.05$), but there was a significant sex difference for peripubertal ($t=2.20$, $p=0.03$) and postpubertal subjects ($t=3.8$, $p=0.0006$) (see Fig. 4). In contrast, the lip difference, with the male lip segment slightly longer, existed at all ages. This, combined with the fact that the mean difference in lip length was much smaller (0.7–1.6 mm, depending on age) than the postpubescent pharynx length difference (6.8 mm), indicates that the disproportionate difference observed in adult male and female vocal tract lengths is primarily due to an increase in pharynx length. Figure 4 illustrates this finding graphically.

Another perspective on this male-specific change in vocal tract morphology was obtained by analyzing the ratio of oral length (sum of lips, tongue blade, and tongue dorsum) to pharyngeal length for each subject. We submitted the ratios of males versus females to an unpaired t-test. For adult men and women there was a significant difference in this ratio (Tanner Stage 5, $N=10$ F, 8 M, $t=3.26$, $p=0.004$). There was no significant difference in the prepubertal sample ($N=15$ F, 33 M, $t=1.04$, $p=0.31$) or the peripubertal sample ($N=18$ F, 21 M, $t=1.08$, $p=0.29$).

Finally, we analyzed the correlations between body size and vocal tract length separately for prepubertal children and

TABLE IV. Results of two-factor MANOVA with vocal tract segment lengths as dependent variables, and age group and sex as independent factors. ''Age Group'' is pre-, peri-, or postpubertal. ''Age Group*Sex'' indicates the MANOVA interaction term.

	Age group		Sex		Age group*sex	
Segment	*F*	*p*	*F*	*p*	*F*	*p*
Lip	4.23	0.017	7.91	0.006	0.57	0.57
Blade	16.87	0.0001	0.218	0.64	1.68	0.19
Dorsum	13.17	0.0001	0.155	0.69	2.21	0.11
Velum	33.95	0.0001	1.37	0.24	2.15	0.12
Pharynx	132.16	0.0001	8.42	0.004	7.96	0.0006

sexually mature males and females (peri- and postpubertal stages). These correlations are presented in Table V, which also presents a comparable analysis of the data collected for rhesus monkeys by Fitch (1997) for comparison. The human regression data for VTL versus height are presented graphically in Fig. 5.

III. DISCUSSION

The current data conclusively document a long-suspected (Peterson and Barney, 1952; Fant, 1966; Mattingly, 1966; Lieberman, 1984) positive correlation between body size and vocal tract length. This correlation is remarkably strong, and closely parallels that found in nonhuman primates (Fitch, 1997). Our data also clearly document a sex difference in vocal tract length that goes beyond sex differences in size: as suggested by Fant (1966), adult males have disproportionately longer vocal tracts than females, and this difference is mostly accounted for by the greater length of the pharynx. This sex difference is not present during childhood: We found no evidence for appreciable sex differences in children, suggesting that the clearly discriminable differences in girls' and boys' voices (Sachs *et al.*, 1973; Bennet, 1980) are primarily due to behavioral, not anatomical, differences. Finally, our data indicate that the adult difference in vocal tract length is caused by a secondary ''descent of the larynx'' which occurs in males at puberty. We suggest that this secondary descent represents a sexually dimorphic morphological adaptation to give adult males a more imposing and resonant voice relative to females or prepubescent males. Below, we discuss these points, along with their acoustic implications and some methodological issues. Finally, we conclude with some speculations about the evolutionary significance of these findings.

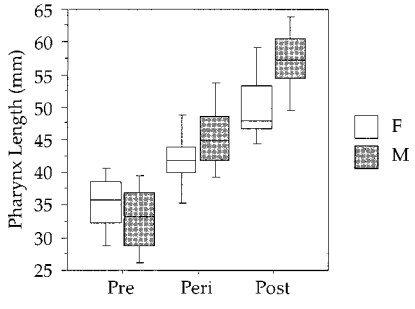


FIG. 4. Boxplot showing the distribution of female (white) and male (shaded) pharynx lengths (measured using the ''line segment'' method) pre-, peri-, and postpuberty. Center lines indicate medians, box edges correspond to 25th and 75th percentiles, and the error bars indicate the 10th and 90th percentiles of pharynx lengths.

A. Accuracy of the data and comparison with previous studies

The results of this study demonstrate a strong positive correlation between vocal tract length (VTL) and body size, regardless of whether height or weight was used to measure body size, and regardless of whether the ''curved'' or ''line segment'' measurements of VTL were used. However, the current sample, with its large variation in body size and young, healthy individuals, may overestimate the strength of the correlation in a randomly chosen sample of adults. A number of factors could potentially obscure the relationship between body size and VTL. The most prominent factor is overweight subjects: Drastic variation in body weight, uncorrelated with height, is common among adults in Western society, and such random weight variation would of course decrease the strength of the correlations found here. We predict that the prevalence of overweight subjects in adult populations will make the correlation of VTL with weight weaker than the correlation with height (despite the fact that the weight correlation was stronger in the current study). This may weaken, but is unlikely to eliminate, the strong correlations documented by the current data.

The adult vocal tract lengths in our study appear shorter in general than those measured in earlier x-ray studies based on upright seated subjects (e.g., Fant, 1960). For Fant's adult

TABLE V. Summary of regressions of body size (height, cm, and $\log_{10}$ weight, g) versus vocal tract length (mm) in prepubertal children, and peri- or postpubertal males and females. Values given are least-squares regression slope, reduced-major-axis slope, intercept and *r* value. Values of the regression for monkeys are given for comparison.

	LS slope	RMA slope	Intercept	*r*
Ht (cm) vs VTL (mm)				
Children	0.521	0.589	45.7	0.884
Females	0.584	0.702	40.2	0.832
Males	0.687	0.804	29.8	0.855
Log Wt (g) vs Log VTL (mm)				
Children	0.274	0.318	−0.163	0.862
Females	0.215	0.246	0.115	0.875
Males	0.277	0.307	−0.160	0.903
Monkeys	0.299	0.318	0.753	0.939

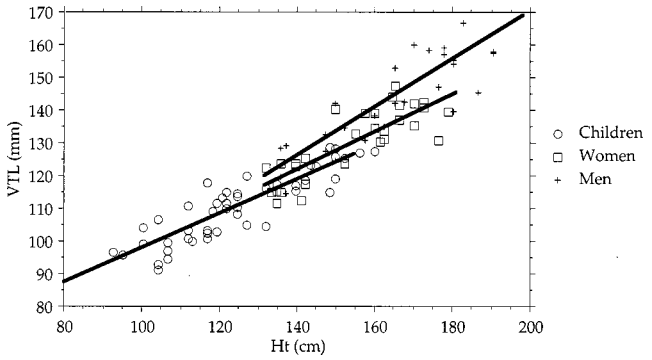


FIG. 5. Height (cm) versus vocal tract length (mm), with separate regression lines illustrating the difference between sexually mature male vocal tract allometry and that of women and children.

male subject, total VTL varied from 16.5 to 19.5 cm, depending on vowel (p. 116, Table 2.33); in the current study, the longest VTL measured was 16.7 cm. One possible explanation for this difference is that our measure of VTL originated at the glottis, while Fant's extended ''to the bottom of the larynx cavity'' (p. 98). If this indicates an origin at the inferior margin of the cricoid, this could add 1 cm or more to the total vocal tract length as measured here (Westhorpe, 1987). Another potential source of discrepancy is movements of the lips and/or larynx during speech. The VTL measurements performed here were during quiet respiration, with a stable laryngeal position. Although Dmitriev and Kiselev (1979) found that larynx position was held fixed in singing, during speech VTL can vary due to movements of both the lips (e.g., Fant, 1960) and larynx (e.g., Sulter *et al.*, 1992). Besides accounting for the difference in our VTL measurements and Fant's, this factor may also decrease the strength of the VTL/body size correlation during running speech. However, the mean value of VTL, around which such perturbations would occur, should remain a relatively stable correlate of body size.

Another difference between Fant's (1960) study and the MRI measurements reported here is that our subjects lay on their backs during the scans, which may lead to a raising of the hyolaryngeal complex (see, e.g., Brock, 1946, p. 29). If this is true, the vocal tract lengths reported here may be systematically shorter than would be the case if the subjects were upright. No studies we are aware of have quantified the actual amount of laryngeal rising caused by recumbency, so we are unable at present to further evaluate this possibility. Nevertheless, the degree of laryngeal shifting due to dorsal recumbency would presumably be consistent between subjects, and thus unlikely to effect the conclusions of this study.

B. Acoustic implications of the vocal tract length/ body size correlation

The anatomical correlation between VTL and body size documented here suggests that formant frequencies, which are closely linked to vocal tract morphology, could provide an acoustic cue to body size. A host of prior studies have documented a close correspondence between vocal morphology and speech acoustics in humans. Sulter *et al.* (1992) looked at differences in VTL for different vocal postures in a single subject, and found a close correspondence between observed and predicted formant values. Moore (1992) also found good agreement between observed and predicted formant values for his five adult male subjects. However, there appeared to be little variation in VTL in his subjects, and the effect of VTL on formants, or its correlation with body size, was not discussed. In an abstract, Bennet (1980) reported significant negative correlations between children's body size and formant frequencies. Similarly, Baer *et al.* (1991), Story (1996) and other workers have found close correlations between vocal tract morphology and speech acoustics. Thus a considerable body of careful work has documented the fit between MRI-derived vocal tract measures and vocal acoustics.

The anatomical data presented here also accord very well with acoustical data presented in Lee *et al.* (1999). These researchers analyzed formant frequencies for 436 normal children aged 5–17 years, finding the expected steady decrease in frequency with age. They also documented a divergence in formant values between males and females starting at age 11 and progressing until about age 15, after which formant values held relatively steady for both sexes. Based on these acoustic data, Lee *et al.* suggested that a vocal tract growth spurt occurs in males between ages 10 and 15 in males. This prediction is clearly borne out by the current data. Although Lee *et al.* did not analyze their acoustic data in relation to body size, the steady drop in formant frequencies with age in both males and females is clearly consistent with an inverse relationship between formants and body size.

The current study complements this earlier work. The addition of children greatly extends the size range of human vocal tracts which have been investigated with MRI, and strongly supports the notion that acoustic correlates of vocal tract length (formant frequencies) could provide an indication of a human speaker's body size. There are a number of practical implications of these findings. For speech forensics, it may be possible to form a reliable estimate of body size and/or age based on acoustic correlates of vocal tract length. Because the spacing between formants is predicted to be correlated with vocal tract length on both theoretical (Fant, 1960; Titze, 1994) and empirical (Fitch, 1997) grounds, formant spacing would represent a promising starting point for investigations of the acoustic correlates of vocal tract length. However, other acoustic variables such as average formant frequency (Cleveland, 1977), the lowest frequency of the first formant, or formant bandwidths, may also be worthy of investigation for such applications. A second potential application is in the field of automatic speech recognition. Children's voices are notoriously difficult for computer speech recognition systems (Lee *et al.*, 1999; Palethorpe *et al.*, 1996), and the wide variety of vocal tract lengths seen in children of different ages and sizes probably contributes to this problem. A system which used information about the talker's size to normalize to vocal tract length, thus duplicating the vocal tract normalization abilities of humans (Ladefoged and Broadbent, 1957; Broadbent and Ladefoged, 1960;

see Lieberman, 1984 for a review) might ameliorate this problem.

C. Sex differences and the ontogeny of the vocal tract

The current results indicate that the development of vocal tract morphology must be discussed together with sex, because vocal development at puberty differs significantly between males and females. Throughout childhood, vocal tract length maintains a positive and linear relationship to both height and log body mass, indicating that there is a steady gradual lengthening of the vocal tract as the child grows. This finding is consistent with previous work (King, 1952; Goldstein, 1980; Westhorpe, 1987). The data reveal no difference in total vocal tract length in prepubertal boys and girls. They are thus consistent with the hypothesis that the discriminable acoustic differences in the speech of boys and girls are due to behavioral rather than anatomical differences (Mattingly, 1966; Sachs *et al.*, 1973; Bennet, 1980). However, we did find a small but significant difference in lip tube length alone, with boys averaging about 1 mm longer than girls at all ages. This finding is consistent with the proposal of Sachs *et al.* (1973) that one facultative way for boys to achieve lower formants is to protrude their lips, thus behaviorally imitating the anatomical difference in total vocal tract length that characterizes adult males.

For females, the pattern of vocal tract growth continues essentially unaltered through puberty and into adulthood. However, males show an additional disproportionate vocal tract lengthening during puberty, which is caused mainly by a descent of the larynx and consequent lengthening of the pharynx. This results in a significant difference in both overall vocal tract length and shape (ratio of oral to pharyngeal cavities) between adult men and women. Although many researchers have noted that men's vocal tract are longer than women's (e.g., Fant, 1960; Peterson and Barney, 1952; Senecail, 1979), to our knowledge, the developmental time course of this difference had not previously been conclusively documented. Combining the results of the present MRI data with earlier radiographic studies, we can thus distinguish two ''descents of the larynx'' in humans: one early in life which occurs for both sexes and is primarily responsible for the morphological uniqueness of the human vocal tract (Sasaki *et al.*, 1977), and a second one at puberty, which is restricted to males.

The phenomenon of laryngeal descent during late infancy in humans is well known (Negus, 1949; Laitman and Crelin, 1976; Sasaki *et al.*, 1977; Lieberman, 1984; Crelin, 1987): We appear to be the only mammalian species in which the epiglottis completely and permanently loses contact with the velum. Our unique anatomy is not without cost: The low position of the larynx is implicated in the high rates of food choking in humans, a major cause of accidental death (Lieberman, 1984; Heimlich, 1975). Most mammals can maintain a patent nasal airway while they swallow, because the velum forms a seal with the larynx, preventing the entry of fluids into the respiratory tract (Harrison, 1995; Crompton *et al.*, 1997). This also is true of human newborns, who can simultaneously suckle and breathe nasally (although neonates are not obligate nose breathers: Miller *et al.*, 1984; Rodenstein *et al.*, 1985), and several researchers (see Bosma and Showacre, 1975) have posited a causal relationship between postpartum descent of the larynx in infants and Sudden Infant Death Syndrome (''crib death''). Although the precise timing of this perinatal descent of the larynx remains obscure (Sasaki *et al.*, 1977; Westhorpe, 1987; Harrison, 1995; Laitman and Crelin, 1976; Laitman and Reidenberg, 1988; Flügel and Rohen, 1991), it occurs between three months and three years of age, and thus prior to the scans performed in the current study.

The uniquely human vocal tract morphology that results from the initial descent of the larynx is widely thought to subserve the production of human speech (Lieberman *et al.*, 1969; Lieberman, 1984), by providing humans with a ''two-tube'' vocal tract capable of creating a much wider variety of formant patterns than that of a typical mammal (Lieberman, 1975, 1984). None of the data presented here are inconsistent with this hypothesis. However, the male-specific descent of the larynx at puberty documented by our data cannot be explained as an adaptation for speech per se, since there is no evidence that adult males speak more intelligibly or are otherwise superior in speech abilities to adult women; Indeed, available evidence suggests the contrary (McCarthy, 1954; Koenigsknecht and Friedman, 1976; Kimura, 1983; Henton, 1992). Why then does this relatively dramatic change in male vocal anatomy occur at puberty?

Could the increase observed be a side effect of some other morphological change at puberty? For example, radiographic data indicate that the ramus of the mandible is longer in adult males. However, this change in jaw size is not yet present in 17 year olds, apparently appearing well after puberty (Hunter and Garn, 1972), and thus not explaining the peripubertal change documented here. Another possibility is that the apparent descent of the larynx is secondary to the increase in overall size of the male larynx at puberty. The increase in circulating testosterone at puberty leads to a near doubling in the anteroposterior length of the glottis in males, which in addition to lowering voice pitch (Hollien *et al.*, 1994), decreases airway resistance (Venn *et al.*, 1998). A measure of vocal tract length which took its origin at the base of the larynx (inferior margin of the cricoid cartilage) would be influenced by this overall increase in larynx size (see, e.g., Westhorpe, 1987). However, our VTL measure originated near the top of the larynx, so it would not be so influenced. Thus the pubertal descent of the larynx is anatomically distinct from the well-known enlargement of the larynx that occurs with male puberty. Finally, although one might suppose that the pubertal descent confers some physiological advantage, available evidence suggests that a descended larynx impairs rather than aiding respiratory and digestive function [see Lieberman (1984) for a review].

A prominent effect of laryngeal lowering is to lengthen the vocal tract and thus lower formants. The acoustic effects of vocal tract elongation are independent from those of the vocal fold elongation. The latter results from laryngeal growth, and accounts for the male-specific drop in fundamental frequency (F_0) at puberty (Titze, 1989, 1994; Hollien *et al.*, 1994). However, both of these changes appear to result

in a more imposing ''deep'' voice relative to prepubescent males, where ''deep'' is understood to indicate not just low F_0 but also possessing lower and more closely spaced formant frequencies, and thus a more resonant, baritone voice quality or timbre. Two previous studies (Cleveland, 1977; Dmitriev and Kiselev, 1979) addressed timbral differences due to vocal tract length in professional singers, and found that baritones and basses have longer vocal tracts than tenors. The fact that vocal tract elongation occurs at puberty suggests that, like facial hair growth and vocal fold lengthening, the pubertal descent of the larynx is a male-specific secondary sexual characteristic in our species.

D. Evolutionary significance of vocal tract elongation

We speculate that our data showing a correlation between vocal tract length and body size provide a plausible evolutionary explanation for the lengthening of the male vocal tract at puberty: that vocal tract elongation serves to exaggerate the impression of size conveyed by the male voice. In general, since body size is correlated with vocal tract length, any acoustic correlates thereof should provide an indication of a speaker's overall body size. Fitch (1994) used computer-synthesized vowels of variable vocal tract lengths to show that human listeners use vocal tract length cues to judge the body size of a speaker. Since the acoustic correlates of vocal tract length are used by listeners to gauge body size, an individual which could lengthen its vocal tract, thereby matching the mean vocal tract length of a larger individual, could exaggerate the impression of size conveyed by its voice. This explanation is similar to the acoustic explanation of smiling (which shortens the vocal tract) suggested by Ohala (1980, 1984). Thus we suggest that the permanent vocal tract lengthening documented in the current study might represent a morphological means of exaggerating vocally projected body size, and its restriction to males suggests that it is perhaps the acoustic equivalent of size-exaggerating secondary sexual characteristics seen in males of many other species, such as lions' manes, bisons' humps, or orangutan cheek pads. Interestingly, human males may pay a price for this adaptation: death rate from suffocation among boys aged 10–14 far exceeds that of any other age group except infants, and the ratio of male to female death rates is 14 to 1 in this peripubertal age group, versus about 3 to 1 in other age groups (Harris *et al.*, 1984; Baker *et al.*, 1992).

In order for vocal tract elongation to have effectively exaggerated size in our ancestors, perceivers must have used VTL as a cue to body size. Data from other species are consistent with this prediction. The strengths of the VTL/body size relationships in this study are nearly identical to that found in rhesus macaques by Fitch (1997) using the ''curved line'' method on lateral radiographs, who found a correlation of $r^2=0.90$ for height/VTL (compare with $r^2=0.854$ for the current human data), and a correlation of $r^2=0.89$ for body mass/VTL ($r^2=0.885$ for humans). Furthermore, the actual scaling coefficient (the RMA slope of the log weight versus log VTL regression line) is identical between human children and monkeys (0.318 in both cases), though the intercept varies (0.75 for monkeys, -0.16 for humans). Thus, despite the differences in human and macaque vocal anatomy (Negus, 1949; Lieberman *et al.*, 1969), both show a similar strong positive correlation between VTL and body size, suggesting that this basic correlation is phylogenetically primitive. However, the descent of the larynx is not present in macaques: the CAT scans of macaques performed by Flügel and Rohen (1991) revealed no laryngeal descent, either in infancy or at puberty. Fitch (1997) also studied his monkey subjects' vocalizations and found that VTL had a strong inverse relationship to formant frequency spacing, as predicted by acoustic theory. These data suggest that formant frequencies could have provided a cue to body size for our primate ancestors, and are thus consistent with the ''size-exaggeration'' hypothesis.

E. Future directions

The data presented here suggest a number of directions for further research. Most obviously, an investigation of the acoustic correlates of vocal tract length and morphology in humans is necessary. Additionally, more detailed volumetric analyses than the relatively simple vocal tract length measures used here, such as those performed by previous MRI researchers (Baer *et al.*, 1991; Sulter *et al.*, 1992; Moore, 1992; Story *et al.*, 1996), might uncover other interesting sex or age differences. Finally, it would be extremely interesting to gain an understanding of the causal mechanism underlying these changes in vocal morphology. A likely first guess is that growth or steroid hormones play a role; testosterone receptors have been discovered in the larynx (Aufdemorte *et al.*, 1983; Tuohimaa *et al.*, 1981), and experimentally administered testosterone leads to increase in thyroid cartilage size in gelded sheep (Beckford *et al.*, 1985). However, as discussed earlier, simple enlargement of the larynx will not result in increased vocal tract length, which requires other changes in the system of ligaments and muscles by which the hyolaryngeal apparatus is suspended. A more detailed understanding of the timing and extent of the morphological changes accompanying puberty, and their acoustic correlates, requires a longitudinal approach, with repeated scans of individuals passing through puberty (e.g., Hollien *et al.*, 1994). Such investigations are currently underway at the NIH.

ACKNOWLEDGMENTS

This work was supported by NIH/NIDCD Postdoctoral Fellowship T32 DC00038 to W.T.F. We gratefully acknowledge the comments of Asif Ghazanfar, Marc Hauser, Daniel Lieberman, Philip Lieberman, Anders Löfqvist, Ignatius Mattingly, Michael Owren, Brad Story, and an anonymous reviewer on earlier versions of this manuscript.

Aufdemorte, T. B., Sheridan, P. J., and Holt, G. R. (**1983**). ''Autoradiographic evidence of sex steroid receptors in laryngeal tissues of the baboon (*Papio cynocephalus*),'' Laryngoscope **93**, 1607–1611.

Baer, T., Gore, J. C., Gracco, L. C., and Nye, R. W. (**1991**). ''Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels,'' J. Acoust. Soc. Am. **90**, 799–828.

Baker, S. P., O'Neill, B., Ginsburg, M. J., and Li, G. (**1992**). *The Injury Fact Book* (Oxford University Press, New York).

Beckford, N. S., Rood, S. R., and Schaid, D. (**1985**). ''Androgen stimulation and laryngeal development,'' Ann. Otol. Rhinol. Laryngol. **94**, 634–640.

Bennet, S. (**1980**). ''Vowel formant frequency characteristics of preadolescent males and females,'' J. Acoust. Soc. Am. **67**, S25.
Biro, F. M., Lucky, A. W., Huster, G. A., and Morrison, J. A. (**1995**). ''Pubertal staging in boys,'' J. Pediatr. (St. Louis) **127**, 100–102.
Bosma, J. F., and Showacre, J. (**1975**). ''Development of upper respiratory anatomy and function'' (U.S. Government Printing Office, Washington, DC).
Broadbent, D. E., and Ladefoged, P. (**1960**). ''Vowel judgments and adaptation level,'' Proc. R. Soc. London **151**, 384–399.
Brock, R. C. (**1946**). *The Anatomy of the Bronchial Tree* (Oxford University Press, New York).
Chiba, T., and Kajiyama, M. (**1941**). *The Vowel: Its Nature and Structure* (Tokyo-Kaiseikan, Tokyo, Japan).
Cleveland, T. F. (**1977**). ''Acoustic properties of voice timbre types and their influence on voice classification,'' J. Acoust. Soc. Am. **61**, 1622–1629.
Crelin, E. (**1987**). *The Human Vocal Tract* (Vantage, New York).
Crompton, A. W., German, R. Z., and Thexton, A. J. (**1997**). ''Mechanisms of swallowing and airway protection in infant mammals (*Sus domesticus* and *Macaca fasicularis*),'' J. Zool. **241**, 89–102.
Curtin, H. D. (**1996**). ''Larynx,'' in *Head and Neck Imaging*, edited by P. M. Som and H. D. Curtin (Mosby, New York), pp. 612–709.
Dmitriev, L., and Kiselev, A. (**1979**). ''Relationship between the formant structure of different types of singing voices and the dimension of supraglottal cavities,'' Folia Phoniatr. **31**, 238–241.
Duke, P. M., Litt, I. F., and Gross, R. T. (**1980**). ''Adolescents' self-assessment of sexual maturation,'' Pediatrics **66**, 918–920.
Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague).
Fant, G. (**1966**). ''A note on vocal tract size factors and nonuniform *F*-pattern scalings,'' Speech Transactions Laboratory Quarterly Progress and Status Report **4**, 22–30.
Fant, G. (**1975**). ''Nonuniform vowel normalization,'' Speech Transactions Laboratory Quarterly Progress and Status Report **2–3**, 1–19.
Fitch, W. T. (**1994**). ''Vocal tract length perception and the evolution of language,'' Ph. D. thesis, Brown University.
Fitch, W. T. (**1997**). ''Vocal tract length and format frequency dispersion correlate with body size in rhesus macaques,'' J. Acoust. Soc. Am. **102**, 1213–1222.
Flügel, C., and Rohen, J. W. (**1991**). ''The craniofacial proportions and laryngeal position in monkeys and man of different ages. (A morphometric study based on CT-scans and radiographs),'' Mechanisms of Aging & Development **61**, 65–83.
Giedd, J. N., Snell, J. W., Lange, N., Rajapakse, J. C., Casey, B. J., Kozuch, P. L., Vaituzis, A. C., Vauss, Y. C., Hamburger, S. D., Kaysen, D., and Rapoport, J. (**1996**). ''Quantitative magnetic resonance imaging of human brain development: Ages 4–18,'' Cerebral Cortex **6**, 551–560.
Goldstein, U. G. (**1980**). ''An articulatory model for the vocal tracts of growing children,'' D. Sc. thesis, Massachusetts Institute of Technology.
Harris, C. S., Baker, S. P., Smith, G. A., and Harris, R. M. (**1984**). ''Childhood asphyxiation by food: A national analysis and overview,'' J. Am. Med. Assoc. **251**, 2231–2235.
Harrison, D. F. N. (**1995**). *The Anatomy and Physiology of the Mammalian Larynx* (Cambridge University Press, Cambridge; New York).
Heimlich, H. J. (**1975**). ''A life-saving maneuver to prevent food-choking,'' J. Am. Med. Assoc. **234**, 398–401.
Henton, C. (**1992**). ''The abnormality of male speech,'' in *New Departures in Linguistics*, edited by G. Wolf (Garland, New York).
Herman-Giddens, M., Slora, E. J., Wasserman, R. C., Bourdony, C. J., Bhapkar, M. V., Koch, G. G., and Hasemeier, C. M. (**1997**). ''Secondary sexual characteristics and menses in young girls seen in office practice: A study from the Pediatric Research in Office Settings network,'' Pediatrics **99**, 505–512.
Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). ''Acoustic characteristics of American English vowels,'' J. Acoust. Soc. Am. **97**, 3099–3111.
Holbrook, R. T., and Carmody, F. J. (**1937**). ''X-ray studies of speech articulations,'' University of California Publications in Modern Philology **20**, 187–238.
Hollien, H., Green, R., and Massey, K. (**1994**). ''Longitudinal research on adolescent voice change in males,'' J. Acoust. Soc. Am. **96**, 2646–2653.
Hunter, W. S., and Garn, S. M. (**1972**). ''Disproportionate sexual dimorphism in the human face,'' Am. J. Physical Anthropol. **36**, 133–138.
Kimura, D. (**1983**). ''Sex differences in cerebral organization for speech and praxic functions,'' Can. J. Psychol. **37**, 19–35.
King, E. W. (**1952**). ''A roentgenographic study of pharyngeal growth,'' Angle Orthodont. **22**, 23–37.
Koenigsknecht, R. A., and Friedman, P. (**1976**). ''Syntax development in boys and girls,'' Child Dev. **47**, 1109–1115.
LaBarbera, M. (**1989**). ''Analyzing body size as a factor in ecology and evolution,'' Annu. Rev. Ecol. Syst. **20**, 97–117.
Ladefoged, P., and Broadbent, D. E. (**1957**). ''Information conveyed by vowels,'' J. Acoust. Soc. Am. **39**, 98–104.
Laitman, J. T., and Crelin, E. S. (**1976**). ''Postnatal development of the basicranium and vocal tract region in man.'' in *Symposium on Development of the Basicranium*, edited by J. F. Bosma (U.S. Government Printing Office, Washington, D.C.).
Laitman, J. T., and Reidenberg, J. S. (**1988**). ''Advances in understanding the relationship between the skull base and larynx with comments on the origins of speech,'' J. Human Evol. **3**, 99–109.
Lee, S., Potamianos, A., and Narayanan, S. (**1999**). ''Acoustics of children's speech: Developmental changes of temporal and spectral parameters,'' J. Acoust. Soc. Am. **105**, 1455–1468.
Lieberman, D. E., and McCarthy, R. C. (**1999**). ''The ontogeny of cranial base angulation in humans and chimpanzees and its implications for reconstructing pharyngeal dimensions,'' J. Human Evol. **36**, 487–517.
Lieberman, P. (**1975**). *On the Origins of Language* (Macmillan, New York).
Lieberman, P. (**1984**). *The Biology and Evolution of Language* (Harvard University Press, Cambridge, MA).
Lieberman, P., and Blumstein, S. E. (**1988**). *Speech Physiology, Speech Perception, and Acoustic Phonetics* (Cambridge University Press, Cambridge, U.K.).
Lieberman, P., Klatt, D. H., and Wilson, W. H. (**1969**). ''Vocal tract limitations on the vowel repertoires of rhesus monkeys and other nonhuman primates,'' Science **164**, 1185–1187.
Mattingly, I. G. (**1966**). ''Speaker variation and vocal tract size,'' J. Acoust. Soc. Am. **39**, 1219.
McCarthy, D. (**1954**). ''Language development in children,'' in *Manual of Child Psychology*, edited by L. Carmichael (Wiley, New York).
Miller, M. J., Martin, R. J., Carlo, W. A., Fouke, J. M., and Strohl, K. P. (**1984**). ''Oral ventilation occurs in term neonates during spontaneous breathing or nasal obstruction,'' Am. Rev. Respir. Dis. **129**, A207.
Moore, C. A. (**1992**). ''The correspondence of vocal tract resonance with volumes obtained from magnetic resonance images,'' J. Speech Hear. Res. **35**, 1009–1023.
Morris, N. M., and Udry, J. R. (**1980**). ''Validation of a self-administered instrument to assess stage of adolescent development,'' J. Youth and Adolescence **9**, 271–280.
Nearey, T. (**1978**). *Phonetic Features for Vowels* (Indiana University Linguistics Club, Bloomington).
Negus, V. E. (**1949**). *The Comparative Anatomy and Physiology of the Larynx* (Hafner, New York).
Nordström, P.-E. (**1975**). ''Attempts to simulate female and infant vocal tracts from male area functions,'' Speech Transactions Laboratory Quarterly Progress and Status Report **2–3**, 20–33.
Nordström, P.-E. (**1977**). ''Female and infant vocal tracts simulated from male area functions,'' J. Phonetics **5**, 81–92.
Ohala, J. J. (**1980**). ''The acoustic origin of the smile,'' J. Acoust. Soc. Am. **68**, 33.
Ohala, J. J. (**1984**). ''An ethological perspective on common cross-language utilization of Fø of voice,'' Phonetica **41**, 1–16.
Palethorpe, S., Wales, R., Clark, J. E., and Senserrick, T. (**1996**). ''Vowel classification in children,'' J. Acoust. Soc. Am. **100**, 3843–3851.
Perkell, J. S. (**1969**). *Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study* (MIT Press, Cambridge, MA).
Petersen, A. C., Crockett, L., Richards, M., and Boxer, A. (**1988**). ''A self-report measure of pubertal status: Reliability, validity, and initial norms,'' J. Youth Adolescence **17**, 117–133.
Peterson, G. E., and Barney, H. L. (**1952**). ''Control methods used in a study of vowels,'' J. Acoust. Soc. Am. **24**, 175–184.
Roche, A. F., and Barkla, D. H. (**1965**). ''The level of the larynx during childhood,'' Ann. Otol. Rhinol. Laryngol. **74**, 645–654.
Rodenstein, D. O., Perlmutter, N., and Stanescu, D. C. (**1985**). ''Infants are not obligatory nasal breathers,'' Am. Rev. Respir. Dis. **131**, 343–347.
Sachs, J., Lieberman, P., and Erickson, D. (**1973**). ''Anatomical and cultural determinants of male and female speech,'' in *Language Attitudes: Current*

Trends and Prospects, edited by R. W. Shuy and R. W. Fasold (Georgetown University Press, Washington, D.C.).
Sasaki, C. T., Levine, P. A., Laitman, J. T., and Crelin, E. S. (**1977**). ''Postnatal descent of the epiglottis in man,'' Arch. Otolaryngol. **103**, 169–171.
Senecail, B. (**1979**). *L'Os Hyoide: Introduction Anatomique a l'Etude de certains Mecanismes de la Phonation* (Faculté de Médicine de Paris, Paris).
Stevens, K. N., and House, A. S. (**1955**). ''Development of a quantitative description of vowel articulation,'' J. Acoust. Soc. Am. **27**, 484–493.
Story, B. H., Titze, I. R., and Hoffman, E. A. (**1996**). ''Vocal tract area functions from magnetic resonance imaging,'' J. Acoust. Soc. Am. **100**, 537–554.
Sulter, A. M., Miller, D. G., Wolf, R. F., Schutte, H. K., Wit, H. P., and Mooyaart, E. L. (**1992**). ''On the relation between the dimensions and resonance characteristics of the vocal tract: A study with MRI,'' Magn. Reson. Imaging **10**, 365–373.
Tanner, J. M. (**1962**). *Growth at Adolescence*, 2nd ed. (Blackwell Scientific, Oxford).
Titze, I. R. (**1994**). *Principles of Voice Production* (Prentice–Hall, Englewood Cliffs, NJ).
Titze, I. R. (**1989**). ''Physiologic and acoustic differences between male and female voices,'' J. Acoust. Soc. Am. **85**, 1699–1707.
Tucker, H. M. (**1987**). *The Larynx* (Thieme Medical, New York).
Tuohimaa, P. T., Kallio, S., and Heinijoki, J. (**1981**). ''Androgen receptors in laryngeal cancer,'' Acta Oto-Laryngol. **91**, 159–64.
Venn, A., Lewis, S., Cooper, M., Hill, J., and Britton, J. (**1998**). ''Questionnaire study of effect of sex and age on the prevalence of wheeze and asthma in adolescence,'' British Medical Journal **316**, 1945–1946.
Westhorpe, R. N. (**1987**). ''The position of the larynx in children and its relationship to the ease of intubation,'' Anaesthesia and Intensive Care **15**, 384–388.

On pitch jumps between chest and falsetto registers in voice: Data from living and excised human larynges

Jan G. Švec[a)]
Centre for Communication Disorders, Medical Healthcom, Ltd., Řešovská 10/491, 181 00 Prague 8, The Czech Republic

Harm K. Schutte and Donald G. Miller
Groningen Voice Research Lab, Department of Biomedical Engineering, Faculty of Medical Sciences, University of Groningen, Bloemsingel 10, 9712 KZ Groningen, The Netherlands

(Received 22 April 1998; revised 12 December 1998; accepted 18 May 1999)

The paper offers a new concept of studying abrupt chest–falsetto register transitions (jumps) based on the theory of nonlinear dynamics. The jumps were studied in an excised human larynx and in three living subjects (one female and two male). Data from the excised larynx revealed that a small and gradual change in tension of the vocal folds can cause an abrupt change of register and pitch. This gives evidence that the register jumps are manifestations of bifurcations in the vocal-fold vibratory mechanism. A hysteresis was observed; the upward register jump occurred at higher pitches and tensions than the downward jump. Due to the hysteresis, the chest and falsetto registers can be produced with practically identical laryngeal adjustments within a certain range of longitudinal tensions. The magnitude of the frequency jump was measured as the ''leap ratio'' F_{0F}:F_{0C} (fundamental frequency of the falsetto related to that of the chest register) and alternatively expressed as a corresponding musical interval, termed the ''leap interval.'' Ranges of this leap interval were found to be different for the three living subjects (0–5 semitones for the female, 5–10 and 10–17 for the two males, respectively). These differences are considered to reflect different biomechanical properties of the vocal folds of the examined subjects. A small magnitude of the leap interval was associated with a smooth chest–falsetto transition in the female subject. [S0001-4966(99)06308-0]

PACS numbers: 43.70.Bk, 43.25.Ts, 43.75.Rs, 43.40.At [AL]

INTRODUCTION

Registers in the human voice have long been a subject of interest for singers and voice teachers, as well as for voice scientists, yet there remains much about registers which is not fully understood. The existence of two basic vibratory patterns of the vocal folds related to voice registers is not considered controversial. The ''heavier'' of the two registers is characterized by more pronounced vertical phase differences in vibration of the vocal folds, larger contact area, larger closed quotient, and lower fundamental frequencies, although there is a considerable range of fundamental frequencies where the two registers overlap in most voices. The assignment of the names ''chest'' and ''falsetto'' to the ''heavier'' and ''lighter'' of these registers is also generally accepted, even if the terminology problem is far from resolved. Valuable overviews on registers can be found, e.g., in publications of van den Berg (1963, 1968b); Large (1972); Hollien (1974); van Deinse (1981); Hirano (1982); and Titze (1994).

In spite of the general agreement on the descriptions of chest and falsetto registers, there has been little investigation of the abrupt transitions, or jumps, between these registers. As used in the present study, a register jump is distinct from a ''smooth register transition,'' the ideal in the western operatic tradition. While this singing tradition generally considers the chest–falsetto jump an accident to be avoided, some other singing traditions—for example, yodeling and country and western—deliberately exploit the chest–falsetto jumps for artistic purposes. In clinical practice, the sudden chest–falsetto jumps are well-known as phenomena occurring in patients with mutational disorders as ''pitch jumps,'' appearing spontaneously and out of control.

Two different mechanisms of creation of chest–falsetto jumps might be considered. The prevailing opinion has been that the sudden change of vibratory pattern and pitch is evoked by a sudden change in activity of the laryngeal muscles (van den Berg, 1963, 1968b; Hirano *et al.*, 1970; Hirano, 1974; Titze, 1988, 1994). An alternative hypothesis is that the sudden change of the vibration of the vocal folds can arise spontaneously without any necessary sudden change in adjustment of the laryngeal muscles. This hypothesis is based on the theory of nonlinear dynamics, which has been successfully applied to vocal-fold vibration in this decade, showing that the voice source is a nonlinear dynamic system (Titze *et al.*,1993; Herzel *et al.*, 1994; Baken, 1995). Under such a hypothesis, a chest–falsetto jump might be seen as a manifestation of ''bifurcation'' in the vocal-fold vibratory system. The term bifurcation designates a sudden qualitative change in the vibration of the vocal folds while the phonatory adjustment is changed gradually (for a general introduction to the topic of bifurcations, see, e.g., Bergé *et al.*, 1984). The bifurcation hypothesis, implying important consequences for understanding not only the voice registers

a)Electronic mail: svecjan@mbox.vol.cz

but also the overall control of the voice mechanism, will be the main subject of interest in our study.

Bifurcations associated with register transitions have been found in vocal-fold models (e.g., Herzel, 1993; Lucero, 1996), as well as in excised canine larynges (Berry *et al.*, 1996). In the latter study, the authors investigated vibration of the canine vocal folds under various phonatory adjustments (varied asymmetry in adduction or elongation of the vocal folds as well as varied subglottal pressure). These authors found distinct patterns of vibrations of the vocal folds in different ranges (''regions'') of these phonatory adjustments. The distinct patterns of vibration (''vibratory regimes'') of the vocal folds were associated with various voice registers (e.g., chest, falsetto, vocal fry, whistle). The changes between neighboring vibratory regimes were discontinuous; furthermore, the regions of phonatory adjustments for the different vibratory regimes were observed to overlap, forming ''regions of coexistence.'' In such regions, more than one vibration pattern resulted from the same phonatory adjustment, and spontaneous jumps between the vibratory patterns were observed here.

The present paper restricts its attention to the chest and falsetto registers. Whereas in the study of Berry *et al.* (1996) excised canine larynges were employed, the present study explores the chest–falsetto jumps in an excised human larynx and in living human subjects. The paper is considered introductory and focuses mainly on the changes in the fundamental frequency and on phonatory adjustments during chest–falsetto jumps. Certain other related questions, such as the relation between the magnitude of the chest–falsetto jumps and the chosen pitch (Švec and Pešák, 1994), the selected vowel, or intensity of phonation, are not addressed here.

I. MATERIALS AND METHOD

Three singers, two male and one female, were used as subjects. Subject 1 was a male bass-baritone opera singer and voice teacher, age 62; subject 2 was a baritone, singing student, age 25; and subject 3 was a mezzo-soprano and singing teacher, age 35. In the preliminary phase of the experiment, the singers were asked to imitate yodeling in an informal protocol to assure that they are able to produce the chest–falsetto jumps. None of these subjects had previous experience with yodeling or similar techniques employing chest–falsetto jumps in singing. The inexperience of the singers in producing chest–falsetto jumps was not considered disturbing, since the interest of the present study was to investigate pitch changes which occurred without full voluntary control. In the experiment proper, the subjects were asked to sustain a tone on an arbitrary pitch in chest register and to produce chest–falsetto leaps away from and back to this tone, without aiming at any specific interval. An ongoing rough perceptual evaluation of pitch-change magnitudes (''leap intervals,'' as defined in Sec. II A), as well as of abruptness of transition, was done by the authors in order to encourage the singers to produce jumps that sounded maximally abrupt. The subjects were allowed to choose an arbitrary vowel, which was kept constant during any one of the phonation tasks. At least 30 of the chest–falsetto leaps were recorded for each subject.

Microphone and electroglottograph (Laryngograph) signals were recorded on a two-channel DAT recorder (Denon DTR-2000) and then analyzed with the CSL (Kay Elemetrics Corp, model 4300B) analysis package. For the analysis, all the data were downsampled to 10 000 Hz, except the data from the female subject, which were sampled also at 51 200 Hz (in order to improve the accuracy of the CSL pitch extraction as done in Fig. 6). The spectrogram of the audio signal (70 Hz resolution) was used to document the changes in the acoustic properties of voice during register transitions. Narrow-band spectrograms of the electroglottographic (EGG) signals (14 Hz resolution) were used to measure changes in frequency of the glottal vibration during the leaps. (The audio and EGG signals show the identical fundamental frequency, but the EGG spectrogram was preferred in this study over the audio spectrogram for illustrative purposes, because of a clearer display of subharmonic frequencies of the vocal folds during the chest–falsetto transitions).

In order to address the question of change in vocal-fold tension for the jumps, data from an experiment with an excised male human larynx were analyzed. These data were taken from a VHS video copy of the instructional film *The Vibrating Larynx*, showing classical experiments on human excised larynges carried out by Jw. van den Berg and colleagues at the University of Groningen in the late 1950s (van den Berg *et al.*, 1960). The longitudinal tension of the vocal folds was slowly varied here by applying an external force on the larynx via a rod connected to the anterior part of the thyroid cartilage. More information on the experimental setup can be found elsewhere (van den Berg and Tan, 1959; Tan, 1960; van den Berg, 1960, 1968a). The data on the change of the vocal-fold vibration between chest and falsetto registers are found in the film in section 49, ''Intermediate adjustments.'' The sound, produced by the larynx and recorded on a soundtrack of the film, was digitized by means of CSL analysis package (sampling frequency of 10 000 Hz) and analyzed by means of narrow-band (14 Hz resolution) spectrograms. F_0 values before and after the chest–falsetto register jump were determined from the 10th harmonic frequency of the sound depicted in the spectrogram and dividing it by 10 (accuracy of 1.4 Hz).

In all the spectrograms the fast Fourier transform (FFT) frame length was set to 1024 samples; the standard Blackman window weighting and 0.8 pre-emphasis factor were used. The FFT frame advance was set for each spectrogram automatically by the CSL package (CSL instruction manual, 1994), the values used (which were subsequently calculated using a rule prescribed in the instruction manual) are given in the figure legends.

II. RESULTS

A. Excised human larynx

The data on the chest–falsetto change in the excised larynx were particularly illuminating for understanding the chest–falsetto jumps; therefore we begin with the excised larynx. Figure 1 displays sound spectrograms from an experiment monitoring the change of the fundamental fre-

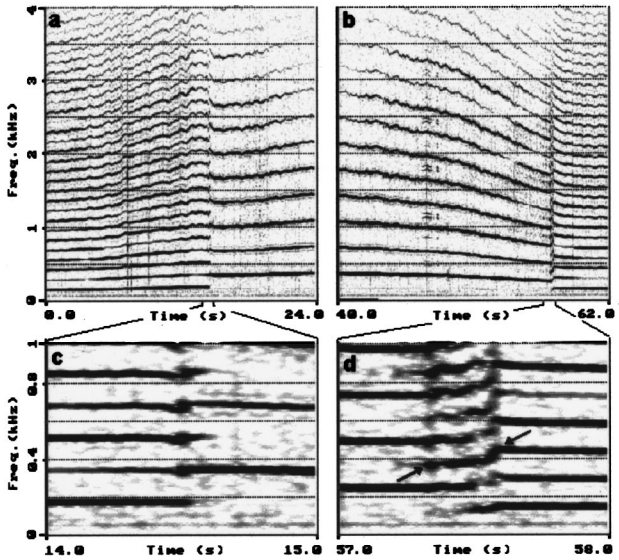


FIG. 1. A narrow-band spectrogram showing change of fundamental frequency in an excised human larynx in response to a smoothly increasing and decreasing longitudinal tension, with all other factors constant (''spectral bifurcation diagram''). A contact microphone attached to the excised larynx was used to record the sound. (a) Applying longitudinal tension to the vocal folds, starting from chest register, causes smooth rising of fundamental frequency, which is interrupted by an upward frequency jump during which the vocal folds change their vibratory pattern from chest to falsetto register. Then, the fundamental frequency rises smoothly again. (b) Subsequent smooth decrease of longitudinal tension causes a smooth declination of the fundamental frequency until the fundamental frequency suddenly drops down. At that moment, the vocal folds abruptly change their vibratory pattern from the falsetto back to the chest register. With further reduction of longitudinal tension, the fundamental frequency declines smoothly again. Note that the upward jump occurs at higher frequencies than the downward jump [(c) versus (d)]. (c) Time detail of the upward chest–falsetto transition. (d). Time detail of the downward falsetto–chest transition. Note the rapid gliding change in the frequency visible in the harmonics (see the two arrows): the frequency is ascending here despite the fact that the tension of the vocal folds was slowly decreasing. [Spectrogram parameters: FFT frame length 102.4 ms, frame advance 100 ms (a) and (b) or 4.5 ms (c) and (d), respectively.]

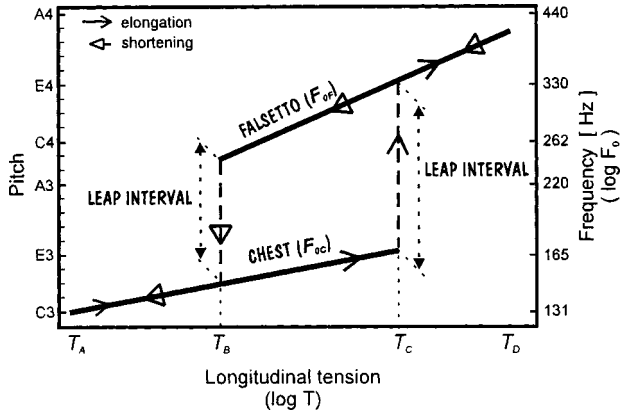


FIG. 2. Schematic illustration of the results from Fig. 1 showing the fundamental frequency of the vocal folds as a function of relative longitudinal tension. A hysteresis loop is apparent here: the upward jump occurs at higher tension than the downward jump. Two vibratory regimes with two different fundamental frequencies F_{0C} and F_{0F}, corresponding to the chest and falsetto registers, respectively, are distinguished. In the range of tensions between T_B (downward-jump value) and T_C (upward-jump value), the two registers can be produced with practically identical laryngeal adjustments. The magnitude of the jump $F_{0F}:F_{0C}$ is defined as ''the leap ratio'', and the corresponding musical interval is termed ''the leap interval.''

quency of the vocal folds when their longitudinal tension is smoothly increased [Fig. 1(a)] and decreased [Fig. 1(b)]. All the other factors were kept constant.

In the beginning of Fig. 1(a), the vocal folds vibrate in chest register at a fundamental frequency 134 Hz (pitch close to $C3$). With increasing longitudinal tension, the fundamental frequency of the chest register rises continuously up to a point where the vocal folds suddenly jump into falsetto. At this point the frequency of the vocal folds abruptly changes from 168 Hz (pitch ca. $E3$) to 332 Hz (pitch ca. $E4$). With further increase of longitudinal tension of the vocal folds, the falsetto frequency rises smoothly again until the frequency of 358 Hz (pitch ca. $F4$) is attained. Next, the external force was smoothly reduced, gradually shortening the vocal folds. The change of frequency in this part of the experiment is shown in Fig. 1(b): it decreases continuously until a point of transition to the chest register is reached, where it suddenly drops from 243 Hz (pitch ca. $B3$) to 146 Hz (pitch ca. $D3$). Thus, the downward jump occurs at a lower fundamental frequency than the upward jump [see Fig. 1(c) versus (d)]. With further relaxation of the vocal folds, these continue vibrating in chest register, and the frequency decreases smoothly until the lowest pitch is reached.

The vibration of the vocal folds (observed from above the larynx by a movie camera under stroboscopic light) showed larger amplitudes and vertical phase differences, as well as a considerable closed phase in the chest register. Smaller amplitudes and vertical phase differences, and a slightly shorter (but clearly present) closed phase were visible in the falsetto register. The film did not reveal any noticeable change in the length of the vocal folds at the moments of the upward and downward register jumps.

The results of this experiment are schematically illustrated in Fig. 2 (for a more detailed explanation of how the scheme was derived, see the Appendix). The figure presents two important features: (1) a continuous change of vocal-fold tension induces at a given point an abrupt change (a bifurcation) in the vocal-fold frequency; (2) there is an apparent hysteresis effect: the upward jump happens at higher longitudinal tension than the downward jump; within the range of tensions from T_B to T_C (''hysteresis interval'') there exist two possible vibratory regimes of the vocal folds, with different fundamental frequencies F_{0C} and F_{0F} corresponding to the chest and falsetto vibratory patterns of the vocal folds.

The magnitude of the frequency jump, $F_{0F}:F_{0C}$, is defined here as the ''leap ratio.'' This ratio reflects the difference between the fundamental frequencies of chest and falsetto registers at a given vocal-fold adjustment. The musical interval corresponding to the leap ratio is called here the ''leap interval.'' The spectrogram in Fig. 1(a) reveals that the upward leap ratio is close to 2:1 (the second harmonic frequency of the chest register immediately before the jump links with the fundamental frequency of the falsetto after the jump). This leap ratio corresponds to a leap interval of an octave or 12 semitones. The downward leap ratio is close to 5:3 [in Fig. 1(b), the third harmonic of the falsetto links

TABLE I. Results of evaluation of the chest–falsetto leap intervals produced by three living subjects.

Subject (initials, age, sex, voice type)	Number of produced chest–falsetto transitions	Range of identified leap intervals (semitones)
1 (DM, 62, male, bass-baritone)	more than 100	10–17
2 (MD, 25, male, baritone)	34	5–10
3 (HvL, 35, female, mezzo-soprano)	62	0–5

approximately with the fifth harmonic of the chest after the jump], corresponding to a leap interval close to a major sixth or 9 semitones.

Figure 1(c) and (d) provide time details of spectrograms from Fig. 1(a) and (b), focusing on the moment of transition. A remarkable phenomenon is particularly apparent in the downward falsetto–chest transition in Fig. 1(d). During the transition, there occurs an $F_{0F}/2$ subharmonic frequency. Before the vibratory regime of the chest register is established, the subharmonic frequency exhibits a gliding change, lasting about 0.3 s, which is distinctly visible in the overtones [see the two arrows in Fig. 1(d)]. This gliding change of frequency is clearly much faster than the F_0 changes before and after the jump and, furthermore, rises paradoxically upwards while the tension was slowly decreasing throughout this part of the experiment. Since this gliding F_0 change was apparently not caused by a sudden change of the vocal-fold tension, it suggests that such an effect is rather of nonlinear origin (see Sec. IIIB).

B. Living subjects

Investigation of the chest–falsetto leap intervals revealed differences among the three living subjects (Table I). In total, the leap intervals varied between zero (indicating no frequency jump during a smooth register transition) and 17 semitones. Each of the subjects, however, had a distinct range of intervals, within which there was some variation. Subject 1 produced intervals in the range from 10 to 17 semitones. Subject 2 produced leap intervals in the range of 5–10 semitones. The female subject (no. 3) showed the smallest intervals, generally less than five semitones. The chest–falsetto leaps were sometimes accompanied by more complex vibratory patterns, such as period doubling, period tripling, and irregular vibrations (these effects will be treated in more detail in a forthcoming paper). The differences among the subjects are represented here by a single example of the chest–falsetto leap for each subject.

1. Subject 1: "2:1" chest–falsetto leap

Figure 3 presents an example of a chest–falsetto leap produced by subject 1, who produced the largest leap intervals of all three subjects, usually exceeding an octave. The narrow-band spectrogram in the upper part of the figure reveals that around the jump the fundamental of the falsetto is linked with the second harmonic frequency of the chest register, indicating that the magnitude of the leap ratio is close to 2:1 or an octave (12 semitones). But, there is an evident upward glide in pitch during the time segment marked between the dashed lines, making the final leap ratio larger, ca. 7:3, corresponding to the interval of minor tenth or 15 semitones.

At this point, it is appropriate to distinguish between the ''transient leap ratio'' and the ''final leap ratio.'' The transient leap ratio is determined by the fundamental frequencies of the two registers immediately before and after the jump, as observed in the narrow-band spectrogram (here 2:1), while the final leap ratio is determined by the *stable* fundamental frequencies of the two registers before and after the whole transition (its magnitude is given by the magnitude of the transient leap ratio corrected by the gliding pitch changes, here 7:3). The gliding pitch change may be brought about by some change in the vocal-fold tension (such a contribution cannot be excluded in the living larynges); however, the occurrence of a similar pitch shift in the excised larynx [Fig. 1(d)] suggests that this effect may, alternatively, be evidence of a nonlinear phenomenon (see Sec. IIIB).

The bottom of Fig. 3 shows the change of the vibratory pattern of the vocal folds as reflected in the electroglottographic (EGG) signal.[1] The chest vibratory pattern is inter-

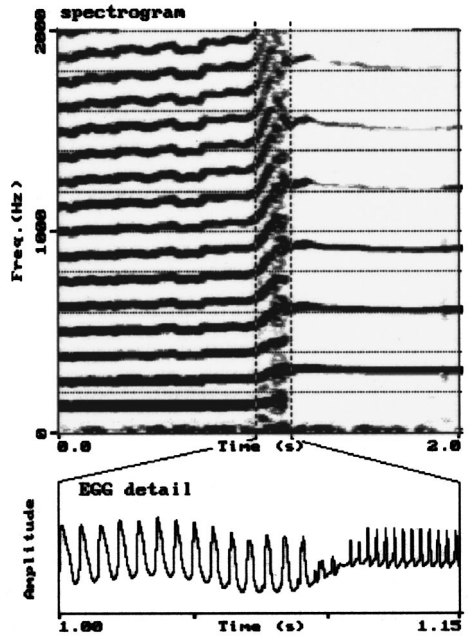


FIG. 3. Chest–falsetto leap observed in male subject 1 (vowel/i/). Top: a narrow-band spectrogram of the EGG signal reflecting the change of the fundamental frequency of the vocal folds (FFT frame length 102.4 ms, frame advance 7 ms). The second harmonic of the chest register lines up with the fundamental of the falsetto, which shows that the transient leap ratio is 2:1. The EGG data at the bottom of the figure are taken from the interval between the dashed lines marked in the spectrogram. (Orientation of the EGG signal: upward indicates increased vocal-fold contact.)

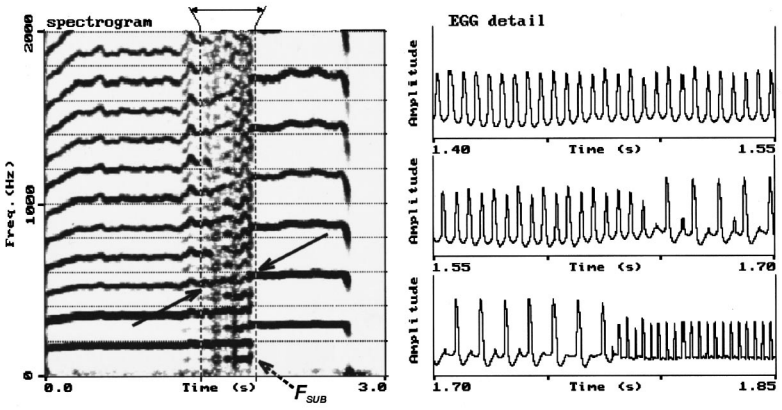


FIG. 4. Chest–falsetto leap observed in male subject 2 (sound /m/). Left: a narrow-band spectrogram of the EGG signal reflecting the change of the fundamental frequency (FFT frame length 102.4 ms, frame advance 10 ms). The third harmonic of the chest register links with the second harmonic of the falsetto (marked by two solid arrows), revealing that the transient leap ratio is 3:2. A subharmonic frequency F_{SUB} occurs during the transition (dashed arrow). Right: the change of the EGG vibratory pattern from chest via subharmonic to falsetto. The EGG data are taken from the interval marked in the spectrogram by vertical dashed lines. A clear period-doubling pattern can be seen in the time interval from 1.65 to 1.77 s.

rupted here by an episode (ca. 30 ms) of greatly diminished vocal-fold contact, after which the EGG curves exhibit a morphology typical of falsetto.

2. Subject 2: "3:2" chest–falsetto leap

Figure 4 shows an example of a sudden chest–falsetto transition recorded from subject 2. The narrow-band spectrogram shows that the third harmonic of the chest register links approximately with the second harmonic of the falsetto (see the two black arrows), indicating that the transient leap ratio is close to 3:2 or a perfect fifth (seven semitones). However, careful examination of the course of the harmonics during the moment of transition (between the dashed lines) reveals, as in Fig. 3, a slight rise in pitch (the two black arrows) which makes the final leap ratio slightly larger, ca. 5:3, corresponding to the interval of major sixth or nine semitones. A striking feature in the spectrogram is the production of the $F_{0C}/2$ subharmonic frequency (an octave lower than the pitch of the original chest register, see the dashed arrow) at the moment of transition.

The EGG signal at the bottom of the figure reveals the change of the vibration of the vocal folds within the interval between the dashed lines in the spectrogram. It shows the transformation of a chest vibratory pattern into a subharmonic pattern (clearly visible between 1.65 and 1.78 s), after which the falsetto pattern is established.

3. Subject 3: Small chest–falsetto leap interval, smooth chest–falsetto transition

The leap intervals of the female subject were the smallest of the three subjects. The subject herself reported finding the production of chest–falsetto jumps to be difficult: sometimes her attempts resulted in a perceptually rather smooth register transition. Figure 5 presents an example of three successive upward and downward register transitions illustrating the change in the acoustic properties of the voice. The change from chest to falsetto register is evident in the audio spectrogram in the diminished intensity of the upper harmonics. All the transitions show signs of abruptness (i.e., sudden change in the intensity of the upper harmonics at the moment of the pitch jump), except the last falsetto–chest transition, which appears smooth (the change in the upper harmonics does not accompany the last downward pitch change, is ca. 0.4 s delayed after the jump, and happens rather smoothly while the pitch is already held constant).

Figure 6 focuses on the pitch changes during the first upward and downward transition (Fig. 5, between the two dashed lines). In the upward transition, the 6th harmonic frequency of the chest register is linked with the 5th harmonic of the falsetto (''middle'') register;[2] thus, the transient leap ratio is almost exactly 6:5, corresponding to a minor third (four semitones) leap interval. The final leap ratio can be determined from the spectrogram as slightly smaller, close to 8:7; after the jump, the frequency of the falsetto descends before becoming stabilized. At the bottom of Fig. 6, there is detail of the EGG signal at the upward jump, together with the values of fundamental frequency extracted from this signal. Note that the upward change of frequency is accomplished very rapidly, within ca. five vibratory periods of the vocal folds, a total of ca. 15 ms duration (time between 1.62 and 1.64 s).

In contrast to the upward frequency change, which is abrupt (as revealed by ''broken frequency lines''), the spectrogram in Fig. 6 does not reveal any abrupt frequency change during the downward transition. This transition is considered here as smooth with respect to pitch.

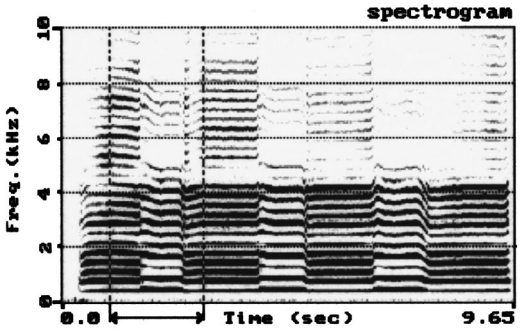


FIG. 5. Three successive upward and downward register transitions produced by female subject 3. Narrow-band spectrogram of the audio signal (vowel /ɛ/) showing the change of the sound quality (FFT frame length 20 ms, frame advance 20 ms). The change from chest to falsetto register is evident in the spectrogram in the diminished intensity of the upper harmonics. Whereas the first two transitions show signs of abruptness, the last downward transition is smooth. The interval between the two dashed lines is studied in more detail in Fig. 6.

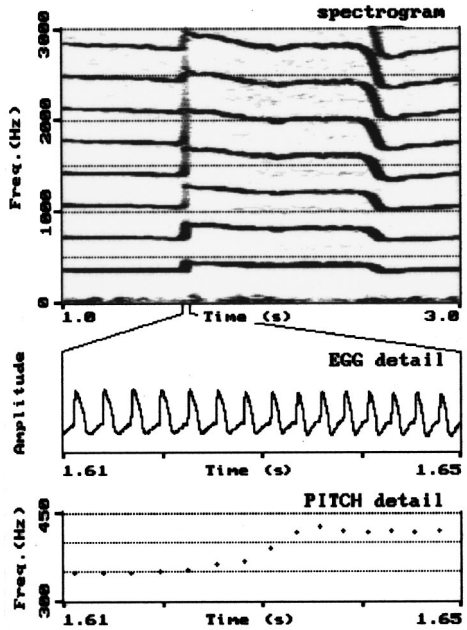


FIG. 6. Top: detail of the change in the fundamental frequency during the upward and downward leap of the female subject (expansion of time between the two dashed lines in Fig. 5). Narrow-band spectrogram of the EGG signal (FFT frame length 102.4 ms, frame advance 7 ms). In the upward transition, the sixth harmonic of chest links with the fifth harmonic of falsetto, revealing that the transient leap ratio is 6:5. The transition downward appears to be smooth with respect to pitch. Middle: EGG data during the upward leap taken from the time interval marked in spectrogram on the top. Bottom: frequency values extracted from the EGG signal above.

III. DISCUSSION

A. Excised larynx: Chest–falsetto jump and the hysteresis

Like the data from excised canine larynges (Berry *et al.*, 1996), our data from the excised human larynx support the bifurcation explanation suggested by the theory of nonlinear dynamics: the chest–falsetto leap may occur abruptly, even in cases when the vocal-fold tension is changed gradually. Figure 2 reveals that instabilities occur at the borders of the hysteresis interval. At value T_B the vibration is highly unstable for falsetto; at T_C, for chest register. Around these values, a tiny change in the longitudinal tension can cause an abrupt, large change in the vocal-fold vibration. In this respect, it is interesting to note that Berry *et al.* (1996) have observed spontaneous chest–falsetto jumps at certain adjustments even when all the phonatory parameters were held constant. Such observations call for further, more detailed research in this area.

The hysteresis observed in the excised human larynx has an important implication for understanding the mechanism of the voice registers: it argues against the prevailing opinion that chest and falsetto vibratory patterns are necessarily produced with different balances of tensions in the intrinsic muscles of the larynx. The different balances are possible, but not necessary: Fig. 2 demonstrates that within the whole hysteresis interval (T_B,T_C), the vocal-fold vibration is bistable and the chest and falsetto vibratory patterns can be produced here with practically identical laryngeal adjustments.

B. Living larynges and the frequency leaps

Although the behavior of a living larynx is more complex than that of the excised larynx, and the role of muscles cannot be excluded completely in our data, it can be supposed that nonlinearities and bifurcations like those found in excised larynges may also be present in living larynges. Indirect evidence of nonlinearities is provided here by the occurrence of period doubling (subject 2, Fig. 4), a bifurcation effect frequently observed in nonlinear systems (Bergé *et al.*, 1984; Baken, 1995). Another indirect evidence was recently given in a study by Miller *et al.* (1998), which has shown that the speed of pitch changes between registers is considerably higher than those within a register. This finding is also visible in our data: the pitch leaps of 12 semitones (subject 1, Fig. 3) or four semitones (subject 3, Fig. 6) were accomplished very quickly, within ca. 30 or 15 ms, respectively, while the maximal speed of pitch changes observed in singing was measured to be much slower: minimum 80 ms for 12 semitones, 40 ms for four semitones (Ohala and Ewan, 1973; Sundberg, 1979).

Our spectrographic data reveal that the changes in the fundamental frequency of the vocal folds during the register jump are made up of two components: (a) a frequency jump, and additionally, in some cases, (b) a gliding frequency change [Figs. 1(d), 3, and 4]. Because of this gliding shift in frequency, the final leap ratio can differ slightly from the transient leap ratio [transient 2:1 versus final 5:3; 2:1 versus 7:3; 3:2 versus 5:3 in Figs. 1(d), 3, and 4, respectively].

Regarding the gliding frequency change, two contributing factors may be considered. Some readjustment in the vocal-fold tension may play a role in living larynges. Such an explanation would, however, be inadequate for the similar shift in frequency that occurred in the excised larynx [Fig. 1(d)]. An alternative explanation is offered again by the theory of nonlinear dynamics: the fundamental frequencies of the chest and falsetto registers at a given vocal-fold adjustment could be determined by two coexisting attractors (limit cycles) of the vocal-fold vibration in a phase space (Bergé *et al.*, 1984). The register transition is then associated with transient vibration of the vocal folds moving from one to another attractor, during which the vibratory period of the vocal folds converges to the new stable value. No change of vocal-fold tension would be required for this explanation.

Each of the three living subjects in the study was observed to have a different range of leap intervals. This can be seen as a reflection of different biomechanical properties of the three sets of vocal folds investigated (see, also, the next section). The small magnitude of the leap interval, as observed here in the female subject, appeared to provide favorable conditions for a smooth register transition: a smooth frequency shift was sometimes observed in this subject instead of a jump. (The distinction smooth/abrupt is based here on the evaluation of spectrograms—identified as continuous/broken frequency lines at the moment of transition, as shown in Fig. 6—as well as on subjective, perceptual evaluations of the voice samples done by the authors.) Some variability of the leap intervals in each of the living subjects (within the ranges reflected in Table I) can be attributed to, e.g., the varied adjustment of the vocal folds during the several in-

stances of the chest-falsetto jump. Further studies are needed to determine the influence of the various adjustments of the vocal folds on the magnitude of the chest–falsetto leap interval.

C. Resonances of the vocal and subglottal tract

The experiments with excised larynges demonstrate that the resonances of the vocal tract are, in principle, not required for the production of chest–falsetto jumps. The vocal tract is completely removed here. Subglottal resonances were shown to have influence on the voice registers (Titze, 1988, 1994; Austin and Titze, 1997). In our case, however, they do not seem to have the primary role: a ''subglottal tract'' is present in the excised larynx (the air-supply tube) but remains constant, while the point of passage from chest to falsetto was observed to vary. Besides the variation caused by the hysteresis as shown here, it was reported that it is possible to extend the range of chest register significantly upward, and thus raise the fundamental frequency of the chest register at which the jump into falsetto occurs, by simply increasing the lateral forces controlling the adduction of the glottis (van den Berg and Tan, 1959; Tan, 1960; van den Berg, 1968a, 1968b). Thus, the factors determining the occurrence of the chest–falsetto jump appear to be primarily influenced by the properties of the vocal folds.

Modeling studies on bifurcations in vocal folds with and without a vocal tract have indicated that the bifurcations occur in both systems. The resonance of the vocal tract was observed to slightly modify the ranges of the distinct vibratory regimes of the vocal folds (Mergell and Herzel, 1996, 1997). In living subjects, our informal experiments on producing chest–falsetto leaps with the same starting pitch and intensity, but different vowels, showed some degree of modification of the leap interval, especially when the first formant approaches the fundamental frequency of the vocal folds. This variability appeared to be small compared to the whole range of the leap intervals in each subject, and was thus neglected in this introductory study.

D. Laryngeal asymmetry, modes of vibrations

Berry *et al.* (1996) have observed the occurrence of spontaneous sudden leaps from chest to falsetto vibratory patterns of the vocal folds in excised larynges, especially under highly asymmetrical conditions, which are considered to represent pathological laryngeal states. The regions of coexistence of chest and falsetto regimes were not found by these authors at symmetrical conditions. In our study, the hysteresis (which is a one-dimensional example of a region of coexistence of the two registers) was observed in the excised larynx experiment while no asymmetrical conditions were voluntarily introduced. Also, laryngoscopic examination revealed no pathological signs of asymmetry in the singers. As no larynx is ideally symmetrical, however, the question remains unresolved as to whether, and to what extent, asymmetry is an important factor in the existence of the hysteresis interval and in the occurrence of spontaneous chest–falsetto jumps. Our early data from videokymography (Švec and Schutte, 1996) indicate that left–right desynchronization of the vocal-fold vibration often occurs in normal subjects. However, instances of phonation have been observed in which the chest–falsetto break occurred with the vocal folds well synchronized (Švec *et al.*, 1997). A generally symmetrical pattern was also observed in our previous detailed study of subharmonic phonation (resembling Fig. 4 in the present study) accompanying the 3:2 chest–falsetto leap (Švec *et al.*, 1996).

The experimental data from register jumps should be helpful in finding a representative model of the vocal folds which would provide a reasonable fit with the variety of the register phenomena. Models have shown that different coupled vibratory modes (eigenfunctions) are responsible for the resulting vocal-fold vibration (Titze and Strong, 1975; Titze, 1976; Berry *et al.*, 1994; Herzel *et al.*, 1995; Berry and Titze, 1996). It is an important further step to relate these vibratory modes to the changes in the vocal-fold vibration during the register transitions. Considering, for instance, two different coupled vibratory modes of the vocal folds playing a role in the register change, one might expect various nonlinear phenomena such as frequency jumps, period doubling, or frequency locking in the vocal-fold vibration.

ACKNOWLEDGMENTS

Parts of the earlier versions of this paper were presented to the Voice Foundation's 24th Annual Symposium: Care of the Professional Voice, June 1995, Philadelphia, PA, and to the 3rd International Workshop: Advances in Quantitative Laryngoscopy, June 1998, Aachen, Germany. The study was carried out during Jan Švec's research stays in the Voice Research Lab in Groningen in 1994, 1997, and 1998, and was supported financially by the Fulbright Commission in the Czech Republic, the Czech Literary Fund, and the Department of Biomedical Technology of the University of Groningen. Valuable comments were provided by R. J. Baken and H. Herzel. These, together with the reviews of the associate editor A. Löfqvist, D. Berry, J. Sundberg, and one anonymous reviewer, resulted in significant improvements of the manuscript. Since 1998, the research carried out in the Center for Communication Disorders, Medical Healthcom, Ltd., in Prague is supported by the Grant Agency of the Czech Republic (GA ČR), Project No. 109/98/K019.

APPENDIX

The exact values of the longitudinal tension are not provided in the film [from the data published by van den Berg and Tan, we learn only that its values are within the range of ca. 10–25 g, corresponding to about 5%–10% strain of the vocal folds (van den Berg and Tan, 1959; Tan, 1960)]. A graph with relative values is presented, in which we consider the fundamental frequencies of the chest and falsetto registers as two functions of relative longitudinal tension of the vocal folds. Taylor's law, which has proven to be a good first approximation for the fundamental frequency F_0 of the vibrating vocal fold (Titze, 1989, 1994), is used for this purpose

$$F_0 = \frac{1}{2L}\sqrt{\frac{\sigma}{\rho}} = \frac{1}{2L}\sqrt{\frac{T}{A\rho}},$$

where L is the length of the vocal fold, ρ is density of the vocal-fold tissue, σ is longitudinal stress ($\sigma = T/A$), T is the external longitudinal tension applied on the vocal fold, and A is the cross-sectional area of the vocal fold. Disregarding the slight changes of L due to elongation of the vocal folds and considering only the longitudinal tension to be variable, we get the proportionality

$$F_0 \sim \sqrt{T},$$

which suggests a linear relationship between the logarithms of tension and frequency

$$\log F_0 \sim \log T.$$

Taking into account the fundamental frequencies of the chest and falsetto registers which are paired in the upward and downward jumps, respectively, and taking into account that each of these pairs of values corresponds to a single value of longitudinal tension, we obtain a graph displaying the frequencies of the chest and falsetto registers as two theoretical functions of the relative longitudinal tension of the vocal folds (Fig. 2).

[1]The electroglottographic (EGG) signal reflects the modulation of the electric impedance of the larynx which is proportional to changes of the vocal-fold contact area during the vibration (Lecluse *et al.*, 1975; Childers *et al.*, 1984; Scherer *et al.*, 1988; Baken, 1992). The contact between the vocal folds is generally smaller in falsetto than in chest register, which is well reflected in the EGG waveform (e.g., Kitzing, 1982; Roubeau *et al.*, 1987; Schutte *et al.*, 1988; Vilkman *et al.*, 1995).

[2]In the literature on singing, the term ''middle'' is the usual name for the female register characterized by a falsetto vibratory pattern, at least in this range of fundamental frequencies (ca. 300–600 Hz, D4–D5).

Austin, S. F., and Titze, I. R. (**1997**). ''The effect of subglottal resonance upon vocal fold vibration,'' J. Voice **11**(4), 391–402.

Baken, R. J. (**1992**). ''Electroglottography,'' J. Voice **6**(2), 98–110.

Baken, R. J. (**1995**). ''Between organization and chaos: a different view of the voice,'' in *Producing Speech: Contemporary Issues for Katherine Safford Harris*, edited by F. Bell-Berti and L. J. Raphael (AIP, Woodbury, NY), pp. 233–245.

van den Berg, Jw., and Tan, T. S. (**1959**). ''Results of experiments with human larynxes,'' Pract. Oto-rhino-laryng. **21**, 425–450.

van den Berg, Jw. (**1960**). Introduction to the instructional film *Voice Production: The Vibrating Larynx*, Booklet accompanying the film, University of Groningen, The Netherlands.

van den Berg, Jw., Vennard, W., Burger, D., and Shervanian, C. C. (**1960**). *Voice Production: The Vibrating Larynx*, Instructional film, University of Groningen, The Netherlands.

van den Berg, Jw. (**1963**). ''Vocal ligaments versus registers,'' NATS Bulletin 20 (December 1963), pp. 16–31. Also in *Contributions of Voice Research to Singing*, edited by J. Large (College-Hill, Houston, Texas, 1980), pp. 146–164.

van den Berg, Jw. (**1968a**). ''Sound production in isolated human larynges,'' Ann. (N.Y.) Acad. Sci. **155** (Art 1), 18–27.

van den Berg, Jw. (**1968b**). ''Register problems,'' Ann. (N.Y.) Acad. Sci. **155** (Art 1), 129–134.

Bergé, P., Pomeau, Y., and Vidal, C. (**1984**). *Order Within Chaos: Towards a Deterministic Approach to Turbulence* (Hermann/Wiley, Paris).

Berry, D. A., Herzel, H., Titze, I. R., and Krischer, K. (**1994**). ''Interpretation of biomechanical simulations of normal and chaotic vocal fold oscillations with empirical eigenfunctions,'' J. Acoust. Soc. Am. **95**, 3595–3604.

Berry, D. A., Herzel, H., Titze, I., and Story, B. (**1996**). ''Bifurcations in excised larynx experiments,'' J. Voice **10**, 129–138.

Berry, D. A., and Titze, I. R. (**1996**). ''Normal modes in a continuum model of vocal fold tissues,'' J. Acoust. Soc. Am. **100**, 3345–3354.

Childers, D. G., Smith, A. M., and Moore, G. P. (**1984**). ''Relationships between electroglottograph, speech, and vocal cord contact,'' Folia Phoniatr. **36**, 105–118.

CSL, Computerized Speech Lab, model 4300B, software version 5.X: instruction manual (**1994**). (Kay Elemetrics Corp., Lincoln Park, NJ).

van Deinse, J. B. (**1981**). ''Registers,'' Folia Phoniatr. **33**, 37–50.

Herzel, H. (**1993**). ''Bifurcations and chaos in voice signals,'' Appl. Mech. Rev. **46**(7), 399–413.

Herzel, H., Berry, D. A., Titze, I. R., and Saleh, M. (**1994**). ''Analysis of vocal disorders with methods from nonlinear dynamics,'' J. Speech Hear. Res. **37**, 1008–1019.

Herzel, H., Krischer, K., Berry, D. A., and Titze, I. R. (**1995**). ''Analysis of spatio-temporal patterns by means of empirical orthogonal functions,'' in *Spatio-temporal Patterns in Nonequilibrium Complex Systems*, edited by P. E. Cladis and P. Palffy-Muhoray (Addison-Wesley, Reading, MA), pp. 505–518.

Hirano, M., Vennard, W., and Ohala, J. (**1970**). ''Regulation of register, pitch and intensity of voice: An electromyographic investigation of intrinsic laryngeal muscles,'' Folia Phoniatr. **22**, 1–20.

Hirano, M. (**1974**). ''Morphological structure of the vocal cord as a vibrator and its variations,'' Folia Phoniatr. **26**, 89–94.

Hirano, M. (**1982**). ''The role of the layer structure of the vocal fold in register control,'' in *Vox Humana: Studies Presented to Aato Sonninen on the Occasion of his Sixtieth Birthday, December 24, 1982*, edited by P. Hurme (University of Jyväskylä, Jyväskylä, Finland), pp. 50–62.

Hollien, H. (**1974**). ''On vocal registers,'' J. Phonetics **2**, 125–143.

Kitzing, P. (**1982**). ''Photo- and electroglottographical recording of the laryngeal vibratory pattern during different registers,'' Folia Phoniatr. **34**, 234–241.

Large, J. (**1972**). ''Towards an integrated physiologic-acoustic theory of vocal registers,'' NATS Bulletin (February/March 1972), pp. 19–36.

Lecluse, F. L. E., Brocaar, M. P., and Verschuure, J. (**1975**). ''The electroglottography and its relation to glottal activity,'' Folia Phoniatr. **27**, 215–224.

Lucero, J. C. (**1996**). ''Chest and falsetto like oscillations in a two-mass model of the vocal folds,'' J. Acoust. Soc. Am. **100**, 3355–3359.

Mergell, P., and Herzel, H. (**1996**). ''Bifurcations in 2-mass models of the vocal folds—the role of the vocal tract,'' in Proceedings: 1st ESCA Tutorial and Research Workshop on Speech Production Modelling and 4th Speech Production Seminar, Autrans, France, 20–24 May 1996 (Institut de la Communication Parlee, Grenoble, France), pp. 189–192.

Mergell, P., and Herzel, H. (**1997**). ''Modelling biphonation—the role of the vocal tract,'' Speech Commun. **22**, 141–154.

Miller, D. G., Švec, J. G., and Schutte, H. K. (**1998**). ''Pitch leaps within and between registers,'' Presentation given at the 27th Annual Symposium: Care of the Professional Voice, Philadelphia, PA, 1–6 June 1998.

Ohala, J., and Ewan, W. (**1973**). ''Speed of pitch change,'' J. Acoust. Soc. Am. **53**, 345(A). (Unpublished data from this study cited by Sundberg, 1979).

Roubeau, B., Chevrie-Muller, C., and Arabia-Guidet, C. (**1987**). ''Electroglottographic study of the changes of voice registers,'' Folia Phoniatr. **39**, 280–289.

Scherer, R. C., Druker, D. G., and Titze, I. R. (**1988**). ''Electroglottography and direct measurement of vocal fold contact area,'' in *Vocal Physiology, Vol. 2: Voice Production, Mechanisms and Functions*, edited by O. Fujimura (Raven, New York), pp. 279–291.

Schutte, H. K., and Seidner, W. W. (**1988**). ''Registerabhängige Differenzierung von Elektroglottogrammen,'' Sprache-Stimme-Gehör **12**, 59–62.

Sundberg, J. (**1979**). ''Maximum speed of pitch changes in singers and untrained subjects,'' J. Phonetics **7**, 71–79.

Švec, J., and Pešák, J. (**1994**). ''Vocal breaks from the modal to falsetto register,'' Folia Phoniatr. Logop. **46**, 97–103.

Švec, J. G., and Schutte, H. K. (**1996**). ''Videokymography: high-speed line scanning of vocal fold vibration,'' J. Voice **10**, 201–205.

Švec, J. G., Schutte, H. K., and Miller, D. G. (**1996**). ''A subharmonic vibratory pattern in normal vocal folds,'' J. Speech Hear. Res. **39**, 135–143.

Švec, J G., Schutte, H. K., and Šram, F. (**1997**). ''Voice registers in the light of videokymography'' (A), ''Videokymography: introduction, vibration of normal vocal folds, voice registers'' (B). Presentations given at the 26th Annual Symposium: Care of the Professional Voice, Philadelphia, PA, 2–7 June 1997 (A), and at the 2nd ''Round Table'' Advances in Quanti-

tative Laryngoscopy using Motion-, Image- and Signal Analysis, Erlangen, Germany, 18–19 July 1997 (B).
Tan, T. S. (**1960**). ''Proeven over geluidsproductie in de menselijke Larynx,'' Doctoral dissertation, University of Groningen, The Netherlands.
Titze, I. R., and Strong, W. (**1975**). ''Normal modes in vocal fold tissues,'' J. Acoust. Soc. Am. **57**, 736–744.
Titze, I. R. (**1976**). ''On the mechanics of vocal-fold vibration,'' J. Acoust. Soc. Am. **60**, 1366–1380.
Titze, I. R. (**1988**). ''A framework for the study of vocal registers,'' J. Voice **2**, 183–194.
Titze, I. R. (**1989**). ''Physiologic and acoustic differences between male and female voices,'' J. Acoust. Soc. Am. **85**, 1699–1707.
Titze, I. R., Baken, R. J., and Herzel, H. (**1993**). ''Evidence of chaos in vocal fold vibration,'' in *Vocal Fold Physiology: Frontiers in Basic Science*, edited by I. R. Titze (Singular, San Diego), pp. 143–188.
Titze, I. R. (**1994**). *Principles of Voice Production* (Prentice-Hall, Englewood Cliffs, NJ).
Vilkman, E., Alku, P., and Laukkanen, A.-M. (**1995**). ''Vocal-fold collision mass as a differentiator between registers in the low-pitch range,'' J. Voice **9**(1), 66–73.

Formants of children, women, and men: The effects of vocal intensity variation

Jessica E. Huber,[a] Elaine T. Stathopoulos, Gina M. Curione, Theresa A. Ash, and Kenneth Johnson

State University of New York at Buffalo, Communicative Disorders and Sciences Department, 122 Cary Hall, 3435 Main Street, Buffalo, New York 14214-3005

(Received 18 December 1998; revised 21 May 1999; accepted 1 June 1999)

The purpose of this study was to investigate the formant frequencies and amplitudes of a wide age range of children and one group of adults at three sound pressure levels (SPLs). Subjects were ten females and ten males in the following age groups (in years): 4, 6, 8, 10, 12, 14, 16, 18, and adults. A sustained /ɑ/ was produced three times by each subject. Formant frequencies were obtained using linear predictive coding analysis. Formant amplitudes were measured from the highest amplitude harmonic in the area of each formant. In addition to following established trends previously published in other studies, it was hypothesized that the first formant frequency would increase with increasing intensity, that females would have higher formant frequencies than same aged males, and that women and children would have reduced formant amplitudes as compared to men. It was found that first formant frequency increased with intensity and changed as a function of age and sex. Second and third formant frequencies changed with age and sex. Formant amplitudes followed the trends set by the SPL differences and did not change as a function of age and sex. Results are discussed in terms of anatomic differences and SPL. © *1999 Acoustical Society of America.* [S0001-4966(99)02109-8]

PACS numbers: 43.70.Ep, 43.70.Aj [AL]

INTRODUCTION

Acoustic measurements provide information about differences that may be present in articulation and glottal configuration between males and females, with changes in age, and with alterations in vocal intensity of speech production. Formant frequencies reflect the size and shape of the supraglottal filter system (Bennett, 1981; Fant, 1970; Gilbert *et al.*, 1997). Amplitudes of the formants provide information pertaining to vocal intensity levels as well as coupling between the supraglottal cavities and subglottal spaces. Vocal tract configurations change throughout development due to alterations in vocal tract size and shape; therefore, it is important to study the acoustic aspects of speech at varying ages (Eguchi and Hirsch, 1969; Kent and Forner, 1979). Additionally, investigation of the differences in formant frequencies and amplitudes with changes in vocal intensity is important since intensity is known to affect articulatory and glottal configuration (Schulman, 1989; Sulter *et al.*, 1996). In all, a developmental design coupled with these acoustic measures provides a noninvasive technique for investigating the growth and development of the human vocal tract and the effects of vocal intensity variation.

A. Formant frequencies

Numerous studies have been completed which have investigated changes in formant frequencies with age. First, second, and third formant frequencies ($F1$, $F2$, and $F3$, respectively) have been found to be highest in children, decrease with increasing age, and be lowest in adult men (Hillenbrand *et al.*, 1995; Kent and Forner, 1979; Lee *et al.*, 1999; Peterson and Barney, 1952). Busby and Plant (1995) studied children and indicated that the change in $F1$, $F2$, and $F3$ was more significant between ages 5 and 7 and became less marked between ages 9 and 11. The decrease in formant frequency appears to be closely linked to size increases of the vocal tract. Therefore, the rate and amount of change is smaller between older age groups than between younger age groups.

Some of these studies have demonstrated trends which varied, depending on the vowel examined. In Eguchi and Hirsch's (1969) classic study, they reported that $F1$ for /ɑ/ did not change with age. However, Busby and Plant (1995) found larger changes for $F1$ of low vowels, like /ɑ/, across age and sex. They concluded that this may be due to the fact that these vowels are produced with relatively open vocal tracts and may be affected by size of the total vocal tract. Eguchi and Hirsch (1969) may not have found age-related changes with the vowel /ɑ/ since they found the greatest amount of variability for $F1$ of the vowel /ɑ/, as compared to other vowels. Therefore, while most studies indicate that $F1$ increases with increased age, the trend for the vowel /ɑ/ is slightly equivocal.

Most previous investigations of formant frequencies treated young children as one group, with little attention to sex distinctions (Eguchi and Hirsch, 1969; Hillenbrand *et al.*, 1995; Petersen and Barney, 1952). More recently, sex differ-

[a]Electronic mail: jehuber@acsu.buffalo.edu

ences in children have been investigated. Lee *et al.* (1999) found that sex differences were not apparent until around age 15 for $F2$ and $F3$. Busby and Plant (1995) reported that within an age group, the values were higher for females than for males for $F1$ in low vowels and for $F2$ in all vowels. Bennett (1981) studied $F1$, $F2$, $F3$, and $F4$ of males and females aged 7–8 years. She found a consistent trend of male children producing lower formant frequencies than the females for all formants. She indicated that the sex differences seen in children are mainly due to differences in the pharyngeal length. Additionally, she hypothesized that sex differences in children, especially for $F1$, may be related to articulatory differences, such as mouth opening size and laryngeal elevation, in addition to the size differences. She suggested that the sex-specific articulatory differences may be used to ''further enhance the perceptual distinction between males and females'' (Bennett, 1981, p. 235). Further, Nordstrom (1977) found that ''anatomical differences, between men and women/children only explain part of the formant differences'' (p. 91).

Formant frequencies have been interpreted to provide information about the size and maturation of the vocal tract. Studies of cephalometric data suggest that males and females grow at differing rates (Boersma *et al.*, 1979). In addition to size factors, it has been suggested that differing proportions of bone growth in males versus females result in face shape differences (Walker, 1994). Finally, some data suggest that even at young ages, male and female children may have slightly different size oral cavities (Boersma *et al.*, 1979), while some data support the view that there are no size distinctions until puberty (Merow and Broadbent, 1990). Therefore, there is evidence to support sex differences in formant frequencies between males and females of all ages.

Except for the Lee *et al.* (1999) investigation, previous studies did not include older children. For example, Eguchi and Hirsch (1969) studied younger children systematically, sampling every year, but stopped studying children at age 13 years. Hillenbrand *et al.* (1995) only studied children at age 10–12 years, and Kent and Forner (1979) studied young children (4 and 6 years) and 12-year-old females. These studies were not developmental as they tended to investigate the endpoints in an age continuum—young children, children almost at puberty, and adults. The current study investigated children throughout the entire development period, including the middle and late teen years, at short, systematic intervals allowing for the discussion of more complete trends concerning the maturation of the vocal tract.

The effect of vocal intensity on formant frequencies has not been systematically studied. Changes in formants were predicted by Schulman's (1989) study which investigated the alterations in articulatory posture with production of loud speech as compared to comfortable speech. He reported that jaw opening increased with increased intensity. Schulman (1989) did not make measures of formant frequency, but if mouth opening increases with loud speech, then tongue height would decrease and $F1$ would increase.

B. Formant frequency amplitudes

The amplitudes of formants have not been studied in detail. However, it is known that formant amplitudes are affected by the (1) intensity of the source, (2) formant frequencies, and (3) damping within the vocal tract (Kent and Read, 1992). When formant frequencies are close together, they boost each other, resulting in higher formant amplitudes (Kent and Read, 1992). If the first formant amplitude is high, then the other formant amplitudes will be higher due to the augmentation of the high-frequency tail from the first formant amplitude (Kent and Read, 1992).

The formant bandwidths are important for predicting the formant amplitudes since, as the formants become more damped, their bandwidths widen and intensity decreases. If there is coupling with the subglottal space during phonation, the amplitudes of the formants will be reduced due to an increased bandwidth causing attenuation of the formant. Stathopoulos and Sapienza (1998) reported that children and women produced larger minimum flows than men. These larger minimum flows may indicate larger glottal openings and greater subglottal coupling. In videostroboscopic studies of laryngeal function, larger posterior glottal openings have been reported in women as compared to men (Sulter *et al.*, 1996). This would result in greater damping on the formants of children and women, resulting in lower formant amplitudes.

However, Peterson and Barney (1952) found no differences in the amplitudes of the formants between the speaker groups. They indicated that there was a large amount of variability in formant frequencies and amplitudes between speakers. Fant's (1973) data demonstrate inconsistent changes in the amplitude of $F1$, $F2$, $F3$, and $F4$ with an increase in overall SPL, with some formant amplitudes increasing or decreasing slightly, but some remaining the same. In sum, the limited data do not show unequivocal trends regarding age, sex, and SPL for formant amplitudes. A systematic investigation of formant amplitudes is still needed.

The present study was designed to collect acoustic data including fundamental frequencies, formant frequencies, and formant amplitudes at various sound pressure levels (SPL) for females and males across a broad age range. The purpose of the present study was to examine information that can be provided by the acoustic measures about the development of the human vocal tract and about articulation across changes vocal intensity. Based on previous findings and the hypothesized physiological effects, the following hypotheses were made.

C. Formant frequencies

1. Age

Previous research has already established that as age increases, formant frequency decreases (see Eguchi and Hirsch, 1969; Hillenbrand *et al.*, 1995; Peterson and Barney, 1952). It was hypothesized the current study would demon-

strate the same trend for $F1$, $F2$, and $F3$, extending this finding to a more complete sampling of subjects.

2. Sex

Prior data regarding size and shape of the vocal tract have demonstrated differences between male and female children of the same age, especially postpuberty (Boersma *et al.*, 1979; Walker, 1994). Additionally, previous studies have demonstrated differences in formant frequencies between same aged males and females, especially for adults (see Bennett, 1981; Busby and Plant, 1995; Petersen and Barney, 1952). Based on this literature, it was hypothesized that all females would have higher $F1$, $F2$, and $F3$ than same aged males.

3. Intensity

Prior research has shown that increased jaw opening is a common articulatory change with increased intensity (Schulman, 1989). Since changes in jaw opening size along with changes in tongue height (increasing oral opening) would most affect the first formant for /ɑ/ (Fant, 1970), it was hypothesized that the $F1$ would increase with increasing SPL. Also, the amount of tongue constriction and the area of the front cavity may affect the $F2$ for /ɑ/ (Fant, 1970). Decreasing tongue height with increased intensity may decrease the amount of tongue constriction and increase the area of the front cavity. Based on this, it was hypothesized that $F2$ would increase as well.

D. Formant amplitudes

1. Sex and age

Based on videostroboscopic and laryngeal aerodynamic data, women and children were expected to have greater coupling between the supraglottal and subglottal cavities than men. Therefore, it was hypothesized that women and children would have lower formant amplitudes.

2. Intensity

The amplitude of the formants is related to changes in the amplitude of the source spectra, slope of the source spectra, filter function, and overall amplitude of the signal. It was hypothesized that amplitude of the formants would increase as SPL increased, with $F2$ and $F3$ increasing proportionally more than $F1$ due to the spread of energy into the higher harmonics, resulting in a less steep slope of the source spectrum.

I. METHODS

A. Subjects

There were ten female subjects and ten male subjects in each of the following age groups (in years): 4, 6, 8, 10, 12, 14, 16, 18, and 20–30 (adults). The mean age for adult females was 23 years, 6 months, and 23 years, 1 month for adult males. Age groups, other than the adults, ranged from six months prior to six months after the year (i.e., 3.5 to 4.5 years). Criteria for subject selection included (1) normal speech, voice, and hearing; (2) general North American dialect, (3) absence of any major health problems, such as head or neck surgery, asthma, pneumonia, or history of smoking, (4) freedom from allergies, infections, or colds on the day of testing, and (5) no professional singing and/or voice training.

B. Speech tasks

Subjects produced sustained phonation (2–3 s) of the vowel /ɑ/ at the following three vocal intensity levels [measured as dB sound pressure level (SPL)]: comfortable, high (10 dB above comfortable), and low (5 dB below comfortable). The experimenter and subject monitored the SPL of the high and low productions on the SPL meter. For younger subjects who could not monitor a number, an intensity bar on the SPL meter was used with stickers showing the child the target length for the bar. Oral instructions and feedback were also provided between trials to assist the subjects in achieving and maintaining the desired SPL. Subjects were given opportunities to practice targeting the correct SPL. Three trials were produced at each SPL.

The vowel /ɑ/ was chosen for two reasons: (1) it is one of the point vowels and, therefore, easily produced repeatably by the young subjects and (2) it has a low fundamental frequency and high first formant, making the linear predictive coding (LPC) analysis less likely to confuse the fundamental frequency with the formant frequency. The average difference between the fundamental frequency and the first formant ranged from 517 to 791 Hz for all age and sex groups, with greatest differences occurring in the data for younger age groups.

C. Equipment and procedures

An acoustic signal was obtained, under C weighting, using a free-field microphone (Quest condenser microphone with Quest model 1700 sound level meter) with a mouth-to-microphone distance of 6 cm at a 45° angle (see Titze and Winholtz, 1993). Mouth-to-microphone distance was held constant by having the subject lightly touch the middle of her/his chin to the top of a bar which was 6 cm from the microphone. The signal was low-pass filtered for anti-aliasing at 8 kHz and digitized to CSpeech 3.0 (Milenkovic, 1989) at a sampling rate of 20 kHz using a computer with an analog-to-digital conversion board (IBM P90, Data Translation 12-bit DT2821), utilizing the maximum voltage resolution (± 10 V).

Calibration of the sound level meter and microphone was completed by coupling a Quest piston phone to the microphone. Two intensities (94 and 110 dB) at two frequencies (1 kHz and 250 Hz) were sequentially applied to the microphone at three gain levels (−20, 0, and +20 dB). The 12 combinations of signal amplitude, frequency, and gain were digitized to CSpeech. The value at 94 dB, 1 kHz, and 0 dB gain was used for calibration, while the other 11 values were used to validate linearity. A calibration constant, converting volts to dB SPL, was computed using an algorithm written within CSpeech.

D. Data measurements

The following measurements were made for each vowel at each intensity level: sound pressure level (SPL), fundamental frequency ($F0$), frequency of formant one ($F1$), frequency of formant two ($F2$), frequency of formant three ($F3$), amplitude of formant one (A1), amplitude of formant two (A2), and amplitude of formant three (A3). The center 200 ms of each vowel was used for measurement. Steady state measurements were used since it has been reported that use of spectral distance measurements (such as subtracting the frequency of one formant from the frequency of another formant) reduces differences between men, women, and children (Hillenbrand and Gayvert, 1993). Similarly, frequencies were not placed on a nonlinear auditory transform scale, since these have also been reported to reduce the differences between groups (Hillenbrand and Gayvert, 1993; Kent and Forner, 1979).

Linear predictive coding analysis (LPC) was used to determine the location of the formant frequencies. A LPC spectrum was computed within CSpeech using a 25-ms Hamming window with preemphasis. Frequency information was taken at the peak of the first three formants identified by the unsmoothed LPC analysis. CSpeech provided a 19.5-Hz resolution in the measurement of the peak formant frequency. In order to increase the resolution of the frequency measurement, a parabolic interpolation was performed using the three frequency bins at the peak of the formant (Hillenbrand *et al.*, 1995; Milenkovic, 1987). The LPC analysis was completed at 50-ms intervals in the center of the vowel for a total of four frames per trial. The mean of the two center frequencies (median) was obtained for each formant and was used as the trial value.

Determining the formants of children's speech is difficult since the LPC analysis has a tendency to miss one of the formants, the second formant for the most part in this study. Methodology for determining the formants from a LPC spectrum was similar to those used by Hillenbrand *et al.* (1995). To avoid missing formants, the LPC spectrum was computed through empirical testing of the following coefficients: 18, 20, 24, 25, 26, 28, and 30. Then formant frequencies identified by each LPC spectrum were compared to the formants identified using the wideband spectrograms as a guide. The spectrographic analysis was obtained from a 25-ms Hamming window with preemphasis, using a bandwidth of 250 Hz for men and older boys, 300 Hz for women and older girls, and 350 Hz for young children. The LPC spectrum with the best match to the spectrographic measurements was used for the data analysis. We chose to use the LPC analysis as the method for determining the formant frequencies since it was an automated routine and determining the peak frequency was more reliable. Spectrograms and previously published data were used as a guide to where the formant was expected to occur. In this way, the LPC and spectrographic analyses were used to cross-check each other.

Once the best LPC coefficient was identified, the LPC spectrum was examined to determine if any formants were missed or misnumbered by the LPC analysis. For example, the most often occurring error was when the LPC analysis missed $F1$ or $F2$ and then mislabeled following formants. If an error was found, the experimenter visually examined the LPC curve to determine the formant frequency. If the formant frequency could not be determined from the LPC analysis, the value from the spectrogram was used.

Formant amplitude information was obtained from a fast Fourier transform (FFT) analysis of the waveform, using a 25-ms Hamming window with preemphasis. The formant amplitudes were taken from the highest harmonic amplitude in dB SPL in the area of the formant. The LPC analysis was used to guide the determination of the location of the formant. The amplitude was not taken from the LPC analysis since it was felt that the LPC analysis may demonstrate slightly reduced amplitudes (Stevens, 1998). Also, the use of a FFT spectra to determine the amplitude of formants is a common methodology (Hanson, 1997; Stevens and Hanson, 1995).

To determine $F0$, the waveform was first inverse filtered. The FFT of the inverse-filtered waveform was obtained, using the setting described above. Fundamental frequency was taken as the frequency of the first harmonic of the FFT spectra.

E. Reliability of measurements

Measurement techniques were examined to ensure confidence in the data collected. For all means and standard deviations, standard deviations will appear in parentheses. First, the measurements of the formant frequencies obtained via the LPC analysis were compared to those from the spectrographic analysis. The average values obtained from the LPC analysis and the spectrographic analysis were very similar. The mean difference for $F1$ was 8 Hz (63 Hz), for $F2$ was 54 Hz (92 Hz), and for $F3$ was 13 Hz (125 Hz). These values were all similar to the difference values reported by Kent and Forner (1979). Further, the formant frequency values obtained via the FFT analysis, which was used to determine the formant amplitudes, were compared to the frequency values obtained from the LPC analysis. Again, these values were very similar. The mean difference for $F1$ was 39 Hz (6 Hz), for $F2$ was 10 Hz (8 Hz), and for $F3$ was 129 Hz (16 Hz).

Intermeasurer reliability was completed for 10% of the LPC analyzed frequency data and for overall sound

TABLE I. Analysis of variance (ANOVA) summary table for the formant measures. Probability level for statistical significance, $p \leq 0.01$. Significant effects are indicated with an asterisk. $F = F$ ratio. Degrees of freedom are in parentheses.

	Age (8)		Sex (1)		Age×Sex (8)		Task (2)		Age×Task (16)		Sex×Task (2)		Age×Sex ×Task (16)	
	F	*p*	F	*p*	F	*p*	F	*p*	F	*p*	F	*p*	F	*p*
*F*0	96.14	<0.00*	374.81	<0.00*	22.54	<0.00*	40.47	<0.00*	0.65	0.84	0.53	0.59	0.44	0.97
*F*1	51.71	<0.00*	125.68	<0.00*	7.54	<0.00*	21.78	<0.00*	1.25	0.23	0.12	0.89	0.71	0.79
*F*2	92.43	<0.00*	151.47	<0.00*	4.55	<0.00*	1.82	0.16	0.96	0.51	0.03	0.97	0.49	0.95
*F*3	204.35	<0.00*	231.46	<0.00*	9.54	<0.00*	0.77	0.46	0.35	0.99	0.29	0.75	0.41	0.98
A1	4.22	<0.00*	7.21	0.01*	5.69	<0.00*	422.19	<0.00*	0.32	0.99	0.02	0.98	0.22	1.00
A2	3.90	0.00*	0.72	0.40	6.34	<0.00*	394.28	<0.00*	0.45	0.97	0.68	0.51	0.25	1.00
A3	7.40	<0.00*	2.00	0.16	5.7	<0.00*	346.66	<0.00*	0.37	0.99	0.55	0.58	0.28	1.00
SPL	2.19	0.03	0.69	0.41	4.64	<0.00*	528.84	<0.00*	0.06	1.00	0.03	0.97	0.01	1.00

pressure level. These comparisons were assessed with *t*-tests (JMP 3.1.5, 1996). The level of significance was 0.05 for these comparisons. Intermeasurer differences were determined to be statistically nonsignificant; therefore, data measurement was deemed reliable. The statistical values were as follows: $F1[t(155) = 0.528,\ p = 0.60]$, $F2[t(154) = -0.306,\ p = 0.76]$, $F3[t(158) = -1.544,\ p = 0.12]$, and SPL $[t(161) = 1.108,\ p = 0.27]$.

To assess the reliability of the amplitude measurement technique, the amplitudes of the formants measured via FFT analysis were compared to those measured with the LPC analysis. These measures were very similar. The mean difference for A1 was 0.60 dB (0.10 dB), for *A*2 was 0.71 dB (0.13 dB), and for *A*3 was 1.12 dB (0.17 dB).

To assess the reliability of the FFT fundamental frequency measurement technique, the fundamental frequencies were measured using a peak picking algorithm. This method allowed for the determination of the fundamental frequency from the periods of ten consecutive cycles of the inverse filtered waveform. The mean difference between the FFT and the peak picking method was 0.2 Hz (0.3 Hz).

F. Statistical analysis

Descriptive and inferential statistics were used to analyze the measurements. The measures listed above were obtained from each trial of the /ɑ/ vowel (3 trials×3 SPL tasks ×180 subjects=1620 observations). The values for each subject were averaged to calculate subject means for each SPL level. The subject means were analyzed using a repeated measures three factorial analysis of variance (ANOVA) with age, sex, and SPL as main effects (JMP 3.1.5, 1996). The repeated measures were the sound pressure levels (low, comfortable, and high). The level of significance was set conservatively at a probability level of 0.01 due to the large number of measures. Tukey-Kramer post-hoc comparisons (JMP 3.1.5) were completed for *post hoc* testing. A sign test was completed on the formant frequencies comparing the female and male values within each age group at each intensity level ($n = 9$ subject groups ×3 SPL tasks ×3 formant frequencies =81). The level of significance for these comparisons was 0.01.

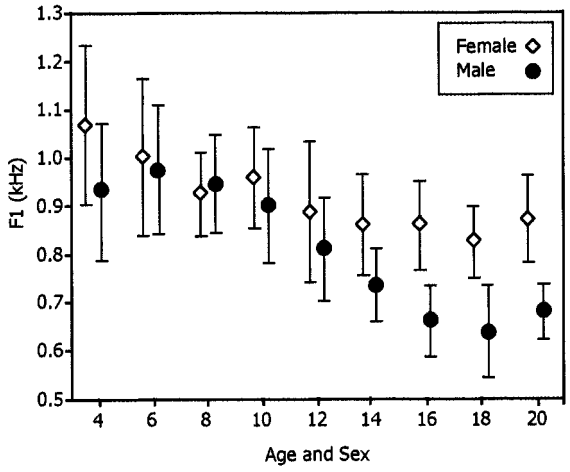


FIG. 1. Average production of first formant frequency for each subject group. Data are averaged across SPL tasks. Symbols represent means; lines designate standard deviations.

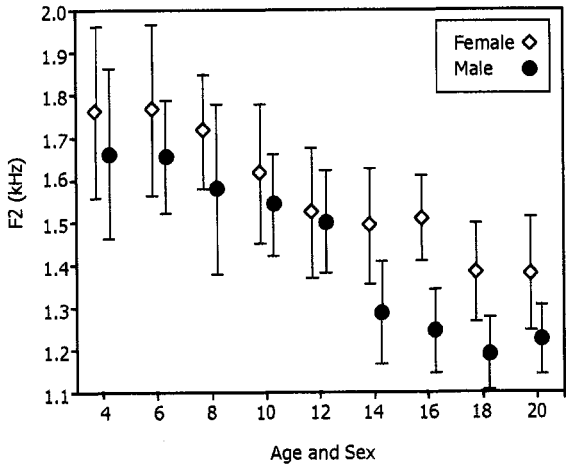


FIG. 2. Average production of second formant frequency for each subject group. Data are averaged across SPL tasks. Symbols represent means; lines designate standard deviations.

II. RESULTS

The means and standard deviations for fundamental frequency and formant frequencies are shown in Appendix A. The means and standard deviations for sound pressure level and formant amplitudes are shown in Appendix B. The results of the ANOVA for the dependent measures are shown in Table I. Significant main effects for SPL were found for $F0$, $F1$, A1, A2, and A3. Significant pairwise age by sex interaction effects were found for $F0$, $F1$, $F2$, and $F3$. Specific main effects and interactions which were of interest to the current study are described below relative to the hypotheses made in the introduction.

The first prediction made was that the current study would extend the previously determined trend of decreasing formant frequency with increasing age for $F1$, $F2$, and $F3$. To determine trends, data were averaged across the three SPL tasks.

$F1$ (see Fig. 1): Girls aged 4 and 6 years produced significantly higher $F1$ than the women. The $F1$ data for the girls over the age of 6 years were similar to the data for the women. The difference between the $F1$ of the 8-year-old girls and the women was 38 Hz. Boys aged 4, 6, 8, 10, and 12 years had significantly higher $F1$ than did the men. After age 12, there was a steady decline in $F1$ to age 16 years. $F1$ dropped 66 Hz between the boys aged 14 years and those aged 16 years. The $F1$ data for the boys aged 16 years and older appeared to be very similar to the $F1$ data for men. A small difference of 27 Hz was noted between the $F1$ of the 16-year-old males and the men.

$F2$ (see Fig. 2): Girls aged 4, 6, 8, and 10 years produced significantly higher $F2$ than the women. From age 12 to 16 years, the $F2$ data for the girls appeared to plateau, with a decrease for the 18 years olds and women. The $F2$ for the 12-year-old girls was 8 Hz higher than for the 16 year olds. There was a 113-Hz difference between the 16-year-old girls and women. The data for $F2$ were similar to the $F1$ data for the males. Boys aged 4, 6, 8, 10, and 12 produced significantly higher $F2$ than the men. After age 12 years, the $F2$ data steadily declined to age 16 years. $F2$ dropped 30 Hz between the boys aged 14 years and those aged 16 years. The $F2$ data appeared to be very similar between the older age groups and the men. $F2$ dropped 30 Hz between boys aged 16 years and the men.

$F3$ (see Fig. 3): Girls aged 4, 6, 8, and 10 years produced significantly higher $F3$ than the women. From age 12 to 16 years, the female $F3$ data appeared to plateau followed by a slight decline. The difference in $F3$ between the 12-and 16-year-old girls was 33 Hz. The difference between the 16-year-old girls and the women was 96 Hz. The $F3$ data for the males were much like the $F1$ and $F2$ data. Boys aged 4, 6, 8, 10, and 12 years had significantly higher $F3$ than the men. The $F3$ data dropped slightly from age 14 years to age 18 years and then rose slightly to the men's $F3$ value. The difference between the 14 year old male and men's $F3$ was 138 Hz, and the difference between the 18 year old male and the men's $F3$ was 188 Hz.

The second prediction made was that the current study would extend the previously determined trend of females postpuberty producing higher $F1$, $F2$, and $F3$ than same aged males. Additionally, it was hypothesized that this sex trend would be present in young children as well. To determine trends, data were averaged across the three SPL tasks.

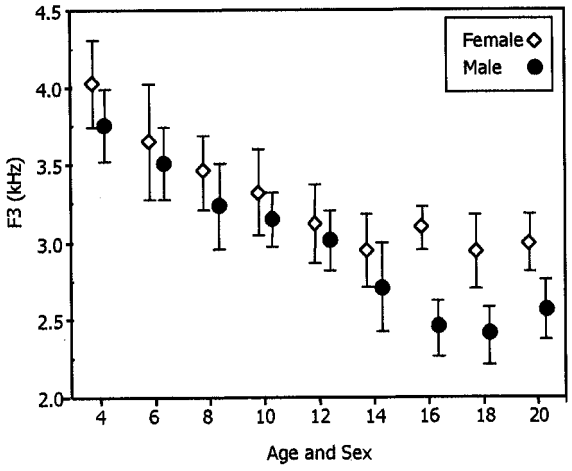


FIG. 3. Average production of third formant frequency for each subject group. Data are averaged across SPL tasks. Symbols represent means; lines designate standard deviations.

$F1$ (see Fig. 1): Females aged 14, 16, and 18 years and adult produced significantly higher $F1$ than same aged males. There was a slight difference between the female and male speakers aged 4 and 12 along the lines of this trend; however, these differences were not significant.

$F2$ (see Fig. 2): Females aged 8, 14, 16, and 18 years and adults produced significantly higher $F2$ than same aged males. There were slight differences between females and males following this trend in the other age groups, except 12 year olds; however, these differences were not significant.

$F3$ (see Fig. 3): Females aged 14, 16, and 18 years and adults produced significantly higher $F3$ than same aged males. There were slight differences following this trend for the other age groups, but these differences were not significant.

In addition to the ANOVA, a nonparametric sign test indicated that within age groups, females produced higher $F1$, $F2$, and $F3$ than males [$Z=8.27$, $p<0.001$]. Seventy-seven of the 81 comparisons showed this effect.

The third prediction made stated that the first and second formant frequencies would increase with increasing SPL. To assess these trends, data were averaged across age and sex (see Fig. 4). For comparison purposes, fundamental frequency trends are reported.

$F1$: The trend for increasing $F1$ with increasing SPL was present; however, the difference was only significant between the high and low conditions. In the high SPL condition, speakers produced a significantly higher $F1$ than in the low condition, a difference of 79 Hz. From the comfortable to the high condition, the $F1$ increased 35 Hz. From the comfortable to the low condition, $F1$ decreased 44 Hz.

$F2$: The trend for increasing $F2$ with increasing SPL was present as well, however, none of the $F2$ data for the SPL conditions were significantly different from one another. The $F2$ increased 18 Hz from the comfortable to high

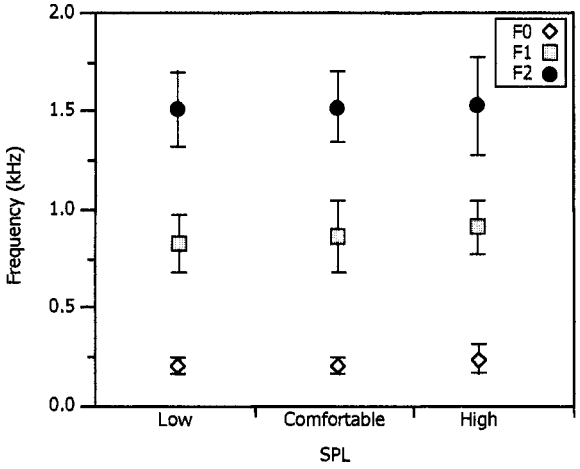


FIG. 4. Production of low, comfortable, and high sound pressure level and its effect on the fundamental frequency, first formant frequency, and second formant frequency of /ɑ/. Data are averaged across subject groups. Sybmols represent means; lines designate standard deviations.

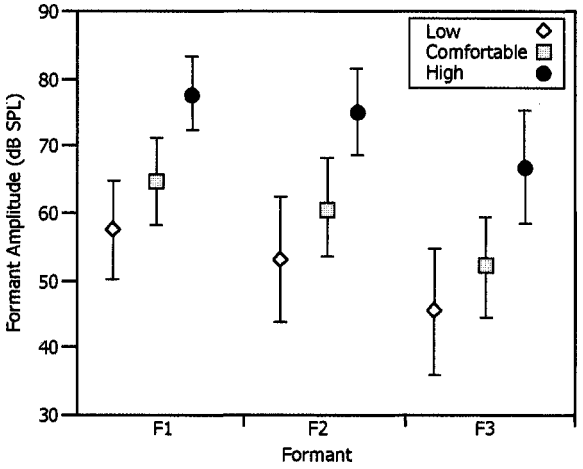


FIG. 5. Production of low, comfortable, and high sound pressure level and its effect on the first, second, and third formant amplitudes of /ɑ/. Data are averaged across subject groups. Symbols represent means; lines designate standard deviations.

condition, and the $F2$ decreased 10 Hz from the comfortable to low condition.

$F0$: The trend for increasing $F0$ with increasing SPL was present. Significant differences were present between the comfortable and high and the low and high conditions, but not between the comfortable and low conditions. $F0$ increased 31 Hz from comfortable to high and increased 7 Hz from comfortable to low. The difference between $F0$ and $F1$ ranged from 617 Hz for low to 672 Hz for high.

The fourth prediction made was that women and children would be expected to have lower formant amplitudes than the older male subjects. To assess these trends, data were averaged across SPL conditions. None of the formant amplitudes demonstrated differences of significant size with changes in age or sex.

The fifth prediction was that the amplitude of the formants would increase as SPL increased, with A2 and A3 increasing proportionally more than A1. To assess these trends, data were averaged across age and sex groups (see Fig. 5).

SPL: The SPL tasks were all statistically different from each other. The average SPL at comfortable was 82 dB, at high 93 dB, and at low 78 dB.

A1: The amplitude of the first formant frequency followed the trends for SPL, significantly increasing with increased SPL. The average A1 at comfortable was 65 dB, while at high it was 78 dB, a difference of 13 dB. At low SPL, average A1 was 56 dB, a difference of 9 dB from comfortable.

A2: The amplitude of the second formant frequency also followed the trends for SPL, significantly increasing with increased SPL. The average A2 at comfortable was 61 dB, while at high it was 75 dB, a 14-dB difference. At low SPL average A2 was 53 dB, a difference of 8 dB from comfortable.

A3: The amplitude of the third formant frequency also followed the trends for SPL, increasing significantly with increasing SPL. The average A3 at comfortable was 52 dB, while at high it was 67 dB, a difference of 15 dB. At low SPL average A3 was 46 dB, a difference of 6 dB from comfortable.

III. DISCUSSION

A. Formant frequencies

It was predicted that the current study would replicate previous formant frequency data and extend the results using a cross-sectional design which allowed a systematic view of the development of the vocal tract across a broad age range. It was stated that $F1$, $F2$, and $F3$ would increase with increasing age. The present developmental data demonstrate that maturation continues through the middle and late teen years. It can be inferred that for males there is a large growth spurt in the vocal tract between the ages of 12 and 14 years followed by a smaller, steady increase in vocal tract size to age 16 or 18. For females, the data suggest that maturation is mostly completed by age 12, though for $F2$ and $F3$ a very slight decline in frequency continued to age 18. Therefore, not only do males and females grow at differing rates, but female growth rate slows down earlier. This finding is supported by the findings of Lee *et al.* (1999) who reported a linear growth trend in males between the ages of 11 and 16, but not in the females.

The inferences about vocal tract growth are substantiated by anatomic growth data. Body height of males has been shown to continue to grow until age 20, with a large growth spurt around age 14, and females have been shown to continue to grow to age 14, with a growth spurt around age 10–12 (Seidel, 1991). Vocal tract growth data follow these trends. According to data in Merow and Broadbent (1990), male facial growth continues through age 18 years, with a growth spurt at age 12–14 years, and female facial growth continues to age 14 years, with a growth spurt at about age 10–12 years.

Cephalometric data address some of these findings regarding differences in male and female maturation. Boersma *et al.* (1979) studied various cephalometric measures of girls and boys ranging from age 4–14 years and found that males and females grow at differing rates. In their study, the distance from the articulare (near the condylar process of the mandible which joins the lower jaw to the temporal bone) to the gnathion (near the menton, the point of the chin) was reported to demonstrate a growth spurt in the boys' data at age $12\frac{1}{2}$ to 14 years, but this trend was not seen for the girls. Girls tended to demonstrate an earlier mature length at about age 10–12 years and a smaller growth spurt. These female and male size/length relationships held for other linear measurements reported, including the distance from the nasion (near the top of the nasal bone) to the menton (near the point of the chin) and the distance from the nasion to the anterior nasal spine (base of the nose near the ala).

In summary, since facial growth continues into the late teens, it is important to continue to study children through these age groups. The unique perspective that these cross-sectional data give is determining the age at which children begin performing like adults. As inferred from the formant frequency data related to age, vocal tract growth in females appears to stabilize at a younger age than males. Girl's formant frequencies became similar to the women's by age 12, indicating a nearly mature vocal tract, whereas for males, formant frequencies significantly lowered at age 14 and continued to decline, indicating growth stabilization at an older age than the females.

Second, it was hypothesized that the current study would replicate previous findings of postpubescent females producing higher formant frequencies than same aged males and extend these findings to younger subjects. The data substantiated a well-known fact: for $F1$, $F2$, and $F3$, females, aged 14, 16, 18, and adults, had significantly higher values than same aged males. This indicates that large differences in formant frequency between females and males occur mainly after puberty. Less known, however, is that for all age groups, female formant frequencies were almost always higher than males in the same age group, even in the very young groups.

While size differences are believed to be the main contributor to formant differences between sexes, it is known that there is not a perfectly linear relationship between vocal tract length and formant frequency change. One reason for this may be that women do not only have shorter vocal tracts than men, but that they have vocal tract shape differences as well. For example, according to Walker (1994), vertical maxillary (upper jaw) facial growth in females is often greater than in males. This is often coupled with a smaller amount of growth in the mandible (lower jaw) of females, thereby increasing facial convexity in females (Walker, 1994). In sum, both vocal tract size and shape differences will affect the formant frequencies, resulting in formant frequency differences.

Cephalometric data address the small differences found between the young boys and girls. According to developmental cephalometric drawings of the face and oral cavity comparing males to females reproduced in Merow and Broadbent (1990), males and females appear to remain comparable in size and shape until age 9–12 years, with boys having only slightly larger oral cavities up until that time. Large size differences appear between ages 12 and 15. However, Boersma *et al.* (1979) reported male/female differences in the articulare to gnathion distance as young as age 4, to varying degrees, with males having a larger length. The differences reported in Boersma *et al.*'s (1979) study may not have been large enough to cause large formant frequency differences. This seems a likely explanation for why statistically significant differences between males and females were not seen in the formant frequency data until age 14.

The third prediction that was made was that $F1$ and $F2$ would increase in frequency with increasing intensity. However, $F1$ was the only formant frequency that clearly increased as SPL increased. Higher vocal intensities are typically produced with an increased jaw opening size with a co-occurring decreased tongue height, thus an increased $F1$. Theoretically, a decrease in tongue height with the vowel /ɑ/ would also cause an increase in $F2$ due to the reduced tongue constriction (Fant, 1970). However, this was not substantiated by the present data. An alternative explanation is that the tongue body is further retracted with increasing intensity. This would lead to a decrease in $F2$ due to an increase in the front cavity volume. Since /ɑ/ has a low and backed articulatory posture (Kent and Read, 1992), these data would indicate that articulatory posture becomes more exaggerated as intensity increases, to a lower, more retracted tongue position.

B. Formant amplitudes

The fourth hypothesis for the current study was women and children would have lower formant amplitudes than the older boys and men. It was originally suggested that if women and children produce voice with more glottal opening than men, then more coupling between tracheal and pharyngeal cavities would produce a more damped waveform. However, the data did not support this hypothesis. There were no significant age or sex differences in the data which is in agreement with the Petersen and Barney (1952) study.

The fifth hypothesis made for the current study was that as SPL increased, A1, A2, and A3 would increase, however, A2 and A3 were predicted to increase to a greater degree due to changes in the slope of the source spectra. In the current study, for a 5-dB increase from low to comfortable SPL, A1, A2, and A3 all increased by approximately 7 dB. For a 10-dB increase from comfortable to high SPL, A1, A2, and A3 all increased by approximately 14 dB. The changes in the formant amplitudes with SPL changes are most likely due to both alterations in the source function and modifications to the filter function.

Relative to changes in the source function, Fant (1973) predicted that with increased vocal intensity there would be a greater amplitude boost in the higher formants due to the spread of energy to the higher harmonics, resulting in a less steep slope of the source. There would be an increase in the harmonic amplitudes of the source, especially in the higher

harmonics. Relative to the changes in the filter function, in the current study there was an increase in the frequency of $F1$ when SPL was increased. Kent and Read (1992) indicated that with an increase in the frequency of $F1$, there would be an increase in the amplitude of all formants, especially the first one.

Fant's (1973) predictions taken along with Kent and Read's (1992) suggestions may provide an explanation for the lack of differential change in amplitude across the formants with SPL change. If the $F1$ increase with high SPL would increase the amplitudes of all three formants, especially A1, and the spread of energy to the higher harmonics would increase A2 and A3, these mechanisms may have combined to increase the three formants by a nearly equivalent amount. Additionally, a decrease in tracheal coupling with high SPL may have narrowed the bandwidths of A1, A2, and A3, increasing all their amplitudes.

Last, since the data were collected using a single vowel /ɑ/, it is important to recognize that the conclusions drawn here may not hold for other vowels (Nordstrom, 1977). However, the vowel /ɑ/ may provide a benefit for examining the development of the vocal tract in that it is produced with a relatively open vocal tract allowing for examination of the entire vocal tract (Busby and Plant, 1995). Vowels such as /i/ and /u/ may not allow this examination due to the striking separation between the front and back cavities of the mouth which is made in their production.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health, DC-2661-01A1. We would like to thank the reviewers for their thoughtful and thorough review of this manuscript which provided for great improvement.

APPENDIX A: MEANS AND STANDARD DEVIATIONS FOR FUNDAMENTAL AND FORMANT FREQUENCIES

TABLE AI. Means and standard deviations (in parentheses) for fundamental frequency and format frequencies in Hz. Data are presented as a function of age, sex, and task.

Age (years)	Task (SPL)	$F0$ (Hz)		$F1$ (Hz)		$F2$ (Hz)		$F3$ (Hz)	
		Females	Males	Females	Males	Females	Males	Females	Males
4	low	272 (32)	251 (33)	926 (210)	897 (157)	1768 (219)	1672 (216)	4074 (272)	3825 (172)
4	comfortable	266 (29)	256 (44)	1139 (137)	981 (141)	1817 (143)	1676 (178)	4064 (289)	3902 (215)
4	high	308 (56)	287 (44)	1162 (97)	1021 (125)	1825 (243)	1753 (185)	4067 (257)	3815 (150)
6	low	275 (37)	254 (26)	970 (192)	942 (139)	1723 (161)	1658 (99)	3675 (517)	3669 (165)
6	comfortable	261 (27)	246 (12)	1030 (216)	992 (142)	1817 (169)	1681 (114)	3735 (258)	3603 (224)
6	high	300 (28)	300 (46)	1081 (110)	1025 (124)	1883 (163)	1727 (104)	3777 (282)	3512 (235)
8	low	267 (49)	240 (26)	875 (105)	921 (95)	1739 (167)	1539 (192)	3504 (248)	3287 (306)
8	comfortable	256 (37)	249 (33)	945 (82)	944 (97)	1738 (152)	1611 (164)	3511 (245)	3324 (353)
8	high	318 (86)	266 (44)	958 (85)	989 (101)	1736 (110)	1619 (176)	3484 (237)	3286 (191)
10	low	262 (45)	244 (29)	1001 (101)	865 (66)	1603 (166)	1582 (98)	3396 (253)	3210 (125)
10	comfortable	256 (41)	237 (26)	935 (94)	938 (131)	1607 (175)	1569 (123)	3380 (263)	3251 (206)
10	high	291 (51)	268 (37)	1016 (89)	992 (115)	1707 (174)	1572 (92)	3417 (282)	3132 (193)
12	low	255 (34)	201 (42)	854 (182)	753 (126)	1533 (165)	1464 (123)	3234 (253)	3050 (239)
12	comfortable	232 (23)	198 (38)	921 (167)	856 (103)	1522 (174)	1525 (121)	3126 (233)	3044 (125)
12	high	263 (32)	232 (52)	933 (120)	875 (104)	1570 (141)	1564 (125)	3116 (271)	2970 (267)
14	low	250 (21)	168 (42)	853 (152)	690 (77)	1502 (152)	1310 (110)	2987 (299)	2745 (328)
14	comfortable	234 (18)	161 (34)	865 (73)	739 (69)	1500 (123)	1292 (125)	2948 (242)	2711 (248)
14	high	253 (14)	185 (35)	927 (85)	780 (89)	1549 (99)	1311 (151)	3066 (177)	2776 (256)
16	low	235 (22)	126 (23)	857 (113)	647 (82)	1559 (127)	1289 (129)	3141 (129)	2511 (167)
16	comfortable	232 (17)	120 (15)	882 (99)	670 (68)	1560 (66)	1273 (86)	3084 (122)	2442 (174)
16	high	247 (29)	143 (28)	905 (71)	694 (72)	1482 (103)	1260 (96)	3152 (131)	2481 (145)
18	low	222 (21)	118 (12)	816 (102)	633 (91)	1471 (81)	1231 (105)	3044 (180)	2409 (210)
18	comfortable	214 (28)	114 (13)	828 (64)	623 (136)	1418 (88)	1209 (71)	2978 (184)	2403 (205)
18	high	250 (37)	139 (22)	858 (37)	679 (77)	1381 (112)	1209 (67)	2894 (333)	2442 (175)
Adults	low	221 (32)	110 (15)	860 (122)	716 (80)	1406 (109)	1261 (68)	3068 (200)	2654 (252)
Adults	comfortable	218 (24)	106 (10)	913 (75)	666 (57)	1441 (128)	1232 (88)	3050 (171)	2585 (164)
Adults	high	230 (30)	128 (19)	892 (83)	709 (47)	1415 (111)	1239 (75)	2971 (196)	2578 (153)

APPENDIX B: MEANS AND STANDARD DEVIATIONS FOR SPL AND FORMANT AMPLITUDES

TABLE BI. Means and standard deviations (in parentheses) for sound pressure level (SPL) and formant amplitudes in dB SPL. Data are presented as a function of age, sex, and task.

Age (years)	Task (SPL)	SPL (dB)		A1 (dB)		A2 (dB)		A3 (dB)	
		Females	Males	Females	Males	Females	Males	Females	Males
4	low	76 (5)	76 (3)	57 (7.2)	56 (4.6)	58 (10.1)	52 (7.4)	50 (13.9)	44 (8.1)
4	comfortable	81 (5)	80 (4)	64 (6.6)	63 (3.9)	63 (7.8)	58 (7.0)	54 (6.7)	50 (5.9)
4	high	91 (5)	91 (3)	79 (6.1)	78 (4.1)	76 (6.0)	74 (5.5)	69 (7.1)	66 (8.0)
6	low	76 (3)	79 (5)	57 (6.7)	63 (6.6)	53 (5.8)	58 (6.6)	47 (5.5)	50 (8.2)
6	comfortable	81 (3)	84 (5)	65 (4.5)	68 (6.0)	61 (5.9)	66 (5.6)	51 (8.3)	56 (6.8)
6	high	92 (3)	94 (4)	78 (5.4)	80 (5.0)	73 (3.9)	79 (5.5)	68 (5.9)	74 (7.2)
8	low	82 (5)	77 (5)	64 (7.1)	59 (8.0)	58 (8.0)	52 (8.3)	52 (8.6)	43 (7.9)
8	comfortable	85 (4)	81 (5)	70 (6.4)	63 (7.9)	64 (6.5)	58 (8.7)	62 (8.9)	50 (6.6)
8	high	97 (5)	92 (5)	83 (5.6)	77 (5.4)	82 (7.0)	74 (7.4)	75 (7.6)	65 (6.6)
10	low	77 (6)	77 (6)	57 (9.4)	61 (7.0)	53 (8.2)	56 (8.8)	49 (8.2)	49 (12.8)
10	comfortable	81 (5)	83 (5)	63 (8.6)	67 (6.5)	60 (6.4)	63 (6.8)	53 (6.6)	55 (7.0)
10	high	92 (5)	94 (5)	78 (5.5)	79 (6.4)	74 (8.0)	78 (7.1)	68 (8.8)	73 (8.5)
12	low	78 (5)	79 (4)	58 (8.8)	59 (6.1)	58 (13.5)	53 (4.7)	47 (7.6)	48 (6.5)
12	comfortable	83 (5)	83 (3)	66 (8.3)	65 (5.5)	61 (8.9)	62 (5.1)	51 (6.7)	53 (6.5)
12	high	93 (6)	94 (3)	78 (5.6)	77 (4.4)	77 (8.4)	76 (5.8)	70 (10.2)	70 (5.3)
14	low	77 (5)	80 (4)	55 (7.1)	59 (5.3)	50 (6.8)	59 (6.8)	43 (9.2)	51 (12.1)
14	comfortable	81 (4)	85 (4)	63 (5.6)	67 (4.6)	60 (4.9)	69 (9.5)	51 (5.2)	54 (5.6)
14	high	91 (5)	95 (4)	76 (6.0)	79 (4.2)	73 (6.7)	78 (4.1)	64 (8.0)	69 (4.3)
16	low	77 (3)	77 (3)	57 (5.6)	52 (10)	52 (6.6)	49 (9.0)	41 (4.1)	38 (6.7)
16	comfortable	82 (4)	82 (4)	65 (5.6)	61 (5.9)	58 (6.2)	59 (7.8)	50 (4.1)	47 (8.1)
16	high	92 (3)	92 (3)	77 (3.7)	74 (5.8)	74 (4.7)	73 (7.8)	64 (6.4)	61 (9.3)
18	low	78 (5)	77 (5)	58 (7.8)	50 (10.9)	52 (8.2)	45 (13.3)	43 (6.5)	39 (12.1)
18	comfortable	82 (4)	82 (5)	67 (6.9)	61 (8.0)	59 (4.9)	58 (9.0)	51 (5.8)	49 (8.6)
18	high	93 (5)	92 (6)	78 (5.7)	74 (7.7)	74 (5.3)	71 (8.0)	65 (8.1)	62 (9.6)
Adults	low	79 (4)	78 (6)	59 (8.6)	52 (10.1)	55 (7.9)	49 (9.0)	46 (9.6)	41 (8.1)
Adults	comfortable	84 (4)	83 (5)	66 (6.5)	61 (7.5)	62 (6.0)	59 (8.7)	52 (7.2)	52 (8.8)
Adults	high	93 (4)	93 (5)	79 (5.2)	75 (6.8)	76 (6.3)	73 (6.4)	64 (9.3)	65 (9.1)

Bennett, S. (**1981**). ''Vowel formant frequency characteristics of preadolescent males and females,'' J. Acoust. Soc. Am. **69**, 231–238.

Boersma, H., van der Linden, F. P. G. M., and Prahl-Andersen, B. (**1979**). ''Craniofacial development,'' in *A Mixed-Longitudinal Interdisciplinary Study of Growth and Development*, edited by B. Prahl-Andersen, C. J. Kowalski, and P. Heydendael (Academic, New York).

Busby, P. A., and Plant, G. L. (**1995**). ''Formant frequency values of vowels produced by preadolescent boys and girls,'' J. Acoust. Soc. Am. **97**, 2603–2606.

Eguchi, S., and Hirsch, I. J. (**1969**). ''Development of speech sounds in children,'' Acta Oto-Laryngol. Suppl. **257**, 5–51.

Fant, G. (**1970**). *Acoustic Theory of Speech Production*, 2nd ed. (Mouton, Paris).

Fant, G. (**1973**). ''Acoustic description and classification of phonetic units,'' in *Speech Sounds and Features* (MIT, Cambridge) (reprinted from Ericsson Technics, 1, 1959).

Gilbert, H. R., Robb, M. P., and Chen, Y. (**1997**). ''Formant frequency development: 15 to 36 months,'' J. Voice **11**, 260–266.

Hanson, H. M. (**1997**). ''Glottal characteristics of female speakers: Acoustic correlates,'' J. Acoust. Soc. Am. **101**, 466–481.

Hillenbrand, J., and Gayvert, R. T. (**1993**). ''Vowel classification based on fundamental frequency and formant frequencies,'' J. Speech Hear. Res. **36**, 694–700.

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). ''Acoustic characteristics of American English vowels,'' J. Acoust. Soc. Am. **97**, 3099–3111.

JMP 3.1.5 (Computer Software). (**1996**). (SAS Institute, Cary, NC).

Kent, R. D., and Forner, L. L. (**1979**). ''Developmental study of vowel formant frequencies in an imitation task,'' J. Acoust. Soc. Am. **65**, 208–217.

Kent, R. D., and Read, C. (**1992**). *The Acoustic Analysis of Speech* (Singular, San Diego, CA).

Lee, S., Potamianos, A., and Narayanan, S. (**1999**). ''Acoustics of children's speech: Developmental changes of temporal and spectral parameters,'' J. Acoust. Soc. Am. **105**, 1455–1468.

Merow, W. W., and Broadbent, B. H., Jr. (**1990**). ''Cephalometrics,'' in *Facial Growth*, 3rd ed, edited by D. H. Enlow (Saunders, Philadelphia).

Milenkovic, P. (**1987**). ''Least squares measures of voice perturbation,'' J. Speech Hear. Res. **30**, 529–538.

Milenkovic, P. (**1989**). CSpeech 3.0 (Speech Analysis Software Program, University of Wisconsin, 1415 Johnson Drive, Madison, WI 53706).

Nordstrom, P. (**1977**). ''Female and infant vocal tracts simulated from male area functions,'' J. Phonetics **5**, 81–92.
Peterson, G. E., and Barney, H. L. (**1952**). ''Control methods used in a study of the vowels,'' J. Acoust. Soc. Am. **24**, 175–184.
Schulman, R. (**1989**). ''Articulatory dynamics of loud and normal speech,'' J. Acoust. Soc. Am. **85**, 295–312.
Seidel, H. M. (**1991**). *Mosby's Guide to Physical Examination*, 2nd ed. (Mosby Year Book, St. Louis).
Stathopoulos, E. T., and Sapienza, C. M. (**1998**). ''Glottal closure characteristics of children and adults as interpreted from minimum flow data,'' Conference on Motor Speech, Tucson, AZ, January.
Stevens, K. (**1998**). Personal communication.
Stevens, K. N., and Hanson, H. M. (**1995**). ''Classification of glottal vibrations from acoustic measurements,'' in *Vocal Fold Physiology Voice Quality Control*, edited by O. Fujimura and M. Hirano (Singular, San Diego).
Sulter, A. M., Schutte, H. K., and Miller, D. G. (**1996**). ''Standardized laryngeal videostroboscopic rating: Differences between untrained and trained male and female subjects, and effects of varying sound intensity, fundamental frequency, and age,'' J. Voice **10**, 175–189.
Titze, I. R., and Winholtz, W. S. (**1993**). ''Effects of microphone type and placement on voice perturbation measures,'' J. Speech Hear. Res. **36**, 1177–1190.
Walker, J. (**1994**). ''Craniofacial changes,'' in *Pediatric Dentistry: Infancy Through Adolescence*, edited by J. R. Pinkham, P. S. Casamassimo, D. J. McTigue, H. W. Fields, Jr., and A. Nowak (Saunders, Philadelphia).

Constant "segmental anchoring" of F_0 movements under changes in speech rate

D. R. Ladd,[a] Dan Faulkner,[b] Hanneke Faulkner, and Astrid Schepman
Department of Linguistics, University of Edinburgh, George Square, Edinburgh EH8 9LL, United Kingdom

(Received 2 November 1998; revised 17 May 1999; accepted 26 May 1999)

This study develops the notion of the ''segmental anchoring'' of F_0 movements, specifically, the finding that both the beginning and the end of a rising pitch accent are anchored to specific points in segmental structure [Arvaniti *et al.*, J. Phonetics **26**, 3–25 (1998)]. If there are segmental anchors for rising accents (and if, as shown by Arvaniti *et al.*, the F_0 level of the beginning and ending points is unaffected by the experimental manipulations), the anchors should be closer together as the speaking rate increases, and the rises should therefore be shorter and steeper. This was tested in two experiments in which speakers were asked to modify rate. The first experiment confirmed the basic prediction that the duration of accentual F_0 rises is affected by changes in rate. The second experiment studied the alignment of the beginning and ending points with their hypothesized segmental anchors. For all speakers there was a strong correlation between the duration of the rise and the time interval between the anchors, but only small and inconsistent effects of rate on alignment. Effects of rate on F_0 excursion size were likewise small and inconsistent. The results support a model in which pitch accents consist of ''tonal targets,'' and in which the alignment and F_0 level of tonal targets are what determine a pitch accent's shape. © *1999 Acoustical Society of America.* [S0001-4966(99)01809-3]

PACS numbers: 43.70.Fq [AL]

INTRODUCTION

A. Intonation and speech rate

To date there has been rather little investigation of the effects of speech rate on intonation, and established findings are few. Perhaps the best documented result is that, as rate increases, the number of pauses and/or prosodic boundaries decreases. This has been demonstrated in various ways and for various languages by, among others, Goldman-Eisler (1968) (for English), Rietveld and Gussenhoven (1987) and Caspers and van Heuven (1993) (for Dutch), and Fougeron and Jun (1998) (for French). However, both the detailed findings and the definition of ''prosodic boundary'' vary considerably from study to study. Another finding is that there are effects of rate on overall level and range of fundamental frequency (F_0) [Cooper and Sorensen (1981)], but again the detailed findings of recent studies [e.g., Caspers and van Heuven (1993) and Fougeron and Jun (1998)] are quite contradictory. The best we can say in our current state of knowledge is that changes in speech rate may have effects on overall F_0 level.

Relative number of prosodic boundaries and relative F_0 level are global properties of contours, and it is therefore difficult to incorporate the findings just summarized into a quantitative model. In particular, knowing about such global effects is of little or no use for predicting or modeling the effect of changes in speech rate on the detailed course of F_0 in individual pitch accents—the localized rises and falls associated with stressed syllables or phrase boundaries. To our knowledge only the two most recent studies just cited [Caspers and van Heuven (1993[1]) and Fougeron and Jun (1998)] are specifically devoted to investigating detailed effects of speech rate modification on the realization of individual pitch accents. This is the topic of the present paper.

Caspers and van Heuven studied the production of pitch accents in short test sentences in Dutch under various kinds of ''time pressure,'' including speaking at a faster rate than normal. Their motivation was to determine which of three phonetic properties characterizing pitch accents—namely their shape, their F_0 level, and their alignment with the segmental string—could be regarded as primary. Their assumption was that speakers would always attempt to preserve the essential features of the accents but might let the less essential features be modified by the articulatory demands of the time pressure. However, they found inconsistent effects of rate on F_0 level, and their central finding for shape and alignment was complex. On the one hand, they found that rising pitch accents are steeper at fast rate than at normal rate, while the slope of falling pitch accents appears unaffected by rate. On the other hand, they report that rising accents show a constant alignment with the segmental structure (the rise begins at the syllable onset irrespective of rate), whereas for falling accents they identified no such constant alignment. They summarize these results by saying that for rises the alignment of the accent is more important than the shape, whereas for falls the shape is more important than the alignment. They do not speculate on why such a difference should exist.

Fougeron and Jun (1998) studied the production of F_0 rises in a narrative passage read aloud in French. They did not find the same asymmetry between rises and falls in

[a]Author to whom correspondence should be addressed; electronic mail: bob@ling.ed.ac.uk
[b]Current address: Aculab plc, Lakeside, Bramley Road, Mount Farm, Milton Keynes MK1 1PT, United Kingdom.

French that Caspers and van Heuven found for Dutch. Moreover, their findings about overall F_0 level suggest a different picture of what the essential features of pitch accents are: whereas Caspers and van Heuven found that Dutch rises at fast rate are steeper and show no consistent effect on excursion size, Fougeron and Jun found that French rises have reduced F_0 excursions at fast rate and slope is unaffected. They speculate [Fougeron and Jun (1998), p. 59f] that this may be because normal French rises are faster than Dutch rises, so that it would be difficult to increase the slope of F_0 change in French. In any case, in the terms proposed by Caspers and van Heuven, Fougeron and Jun find that the shape of an accent is more important than its F_0 level. They do not mention alignment with the segmental string.

It is not the purpose of this paper to address in any detail the difference between rises and falls, or the apparent conflict between Caspers and van Heuven's results and Fougeron and Jun's. Briefly, we believe that the differences may result from the fact that many of Caspers and van Heuven's rises were prenuclear while the falls were all nuclear, whereas most of the accents investigated by Fougeron and Jun were nuclear. Moreover, because of the prosodic patterns of French, all the F_0 movements studied by Fougeron and Jun were on phrase-final syllables, while many of Caspers and van Heuven's F_0 movements, even the nuclear ones, were followed by one or more unstressed syllables within the same phrase. In this paper, as will be seen below, we have concentrated on prenuclear rises, as a way of avoiding the potential confounds that make it difficult to compare Caspers and van Heuven's results with Fougeron and Jun's.

Importantly, however, we assume that similar general principles govern the alignment of F_0 movements with the segmental string in all languages, and that therefore in comparing results across languages it is legitimate to assume that differences result from the interaction of language-specific phonological differences (e.g., the location of word stress in Dutch and in French) with broadly similar or even universal general principles. This view becomes especially plausible in the light of a recent study by Xu (1998), which investigated the alignment of lexical tone contours with the segmental string in Standard (Mandarin) Chinese. (We did not become aware of Xu's study until after this paper was first submitted for publication.) Xu found that the F_0 maximum of the rising tone (tone 2) is invariantly aligned with the end of the syllable with which it is associated, regardless of differences in syllable structure and speech rate. While the function of F_0 movements in Chinese is obviously very different from that of F_0 movements in most European languages, the existence of similar kinds of regularities suggests that the phonetic/phonological principles are shared. Xu himself appears to hold this view, stating that alignment studies are relevant ''for theoretical understanding of tone and intonation in general'' [Xu (1998), p. 179].

Given this point of view, our interest is less in speech rate variation as such, and more in the general principles governing the alignment of F_0 movements with the segmental string. In this we follow Caspers and van Heuven, who used speech rate as a heuristic to get at the linguistic essence of pitch accents—''an experimental tool for focusing on the communicatively important properties of pitch movements'' [Caspers and van Heuven (1993), p. 162]. Xu seems to have had similar motivations. (By contrast, Fougeron and Jun's study was motivated primarily by interest in speech rate and by questions of speech production; they wish ''to understand the global mechanisms involved in speech production at fast rate'' [Fougeron and Jun (1998), p. 48].) More specifically, the goal of this paper is to build on the recent findings of Arvaniti *et al.* (1998) on the ''segmental anchoring'' of F_0 targets, showing that this phenomenon can be generalized beyond the specific language and the specific experimental context on which Arvaniti *et al.*'s report is based.

B. ''Segmental anchoring'' of F_0 targets

In their study of rising prenuclear pitch accents in Modern Greek, Arvaniti *et al.* (1998) found that both the beginning and the end of F_0 rises are anchored to specific points in the segmental string. Like Caspers and van Heuven for Dutch [and like Prieto, van Santen, and Hirschberg (1995) for Mexican Spanish], they found that the beginning of the rise is temporally aligned with the end of the unstressed syllable preceding the accented syllable. In addition, however, they found that the end of the rise is also consistently aligned with a segmental landmark, namely the beginning of the following unstressed vowel. This obviously means that the duration of the F_0 rise is highly dependent on the segmental composition of the accented syllable that it accompanies. In a word like [ro'ditiko], the rise will span the relatively short sequence [dit], whereas in a word like [pa'remvasi], the rise will span the rather longer sequence [remv]. In one of Arvaniti *et al.*'s experiments the rise times ranged from approximately 125 to 325 ms, in close correlation with the duration of the accented syllable and its following consonants.

Moreover, although the rise time is variable, Arvaniti *et al.* found little effect of duration on the size of the F_0 excursion covered by the rise. Rather, for all but one of the five speakers in their study, the minima and maxima at the beginning and end of the rises are essentially constant. The combination of constant F_0 excursion and variable duration means, of course, that the slope of the rise is also highly dependent on the segmental composition of the accented syllable it accompanies. Arvaniti *et al.* argue that their findings on Greek represent evidence in favor of a theory of the structure of intonation contours based on tonal targets, which is widely assumed in current linguistic work [e.g., Bruce (1977), Pierrehumbert (1980), Ladd (1996)]. They argue that the F_0 minima and maxima reflect distinct (though obviously coordinated) phonological events that are independently aligned with specific places in the segmental string. The F_0 rise is, in this view, merely the transition from the minimum or ''low tone'' to the maximum or ''high tone.''

The finding of segmental anchoring—like the phonological consequences that are drawn from it by Arvaniti *et al.*—is somewhat at odds with past experimental and quantitative work on F_0. For example, Fujisaki's model for synthesizing F_0 [e.g., Fujisaki and Sudo (1971), Fujisaki (1983)] assumes that pitch accent rises and falls have char-

acteristic durations: it treats these F_0 changes as a critically damped response to a step function, such that for any given setting of the speaker parameters the F_0 changes has approximately the same duration regardless of the size of the F_0 excursion. Others [notably Ashby (1978) on English nuclear rising accents] have suggested that pitch accents have constant slopes, and that the size of any F_0 excursion therefore depends on its duration. Neither the characteristic duration view nor the constant slope view can be reconciled with the finding of segmental anchoring. More recent work appears somewhat more consistent with Arvaniti *et al.*'s findings; in particular, the notion that alignment is crucial emerges clearly from the work of Caspers and van Heuven, and of Xu just cited above. Alignment is similarly important in Taylor's recent model of F_0 [Taylor (1998)], which is aimed at speech synthesis applications; in this model, slope is subsumed under a single ''tilt'' parameter, which expresses the shape of any pitch accent, including rising–falling ones, with a measure of the relative contributions of rise and fall components to the overall shape.

Nevertheless, both Caspers and van Heuven, and Taylor and his colleagues, draw a clear distinction between the alignment of a pitch accent and its shape. Caspers and van Heuven, as we saw, identify shape, level, and alignment as the three essential properties of a pitch accent. Taylor's model likewise treats ''tilt'' as entirely independent of ''position'' (alignment of the highest point of the accent relative to the onset of the accented vowel). This very distinction between shape and alignment is called into question under a strong version of Arvaniti *et al.*'s conclusions. That is, if it turns out to be possible, for any F_0 movement type in any language, to identify target points that exhibit ''segmental anchoring,'' then the appropriate terms of reference of a general quantitative model should be the target points themselves. Other properties of ''shape,'' such as slope and duration, may be derivable from the segmental alignment and F_0 level of the targets themselves.

C. Overview of the paper

Establishing the generality of the phenomenon of segmental anchoring is clearly relevant to Caspers and van Heuven's, and Xu's, general goals sketched above, i.e., achieving a better understanding of the invariant features of F_0 movements in specific languages, and of the general principles governing the alignment of such movements with the segmental string. The goal of this paper is therefore to adduce a different kind of evidence for this phenomenon. As we saw above, Arvaniti *et al.* demonstrated segmental anchoring by manipulating the temporal interval between putative tonal targets by modifying the segmental composition of accented syllables. In the present study, we achieved the same type of experimental manipulation by keeping the segmental materials constant and asking speakers to modify their speaking rate. (Note that Xu's study combined both techniques.) Broadly speaking, our prediction was that the tonal targets would be closer together at fast rate and farther apart at slow rate, in correlation with the interval between their segmental anchors. In experiment 1, which may be regarded as a kind of pilot study, we tested the rather general prediction that the slope and duration of individual accents would be affected in this way by changes in speech rate. In experiment 2, we tested the more specific predictions that the beginning and ending points of accentual rises would remain anchored to specific segmental landmarks regardless of speech rate, and that the F_0 excursion would not depend on rate.[2]

In both experiments, we compare the segmental anchoring hypothesis with what we might call the ''constant slope'' and ''constant duration'' hypotheses. These are essentially only null hypotheses against which to test our own (though as we saw above, some authors have come close to espousing one or the other as an account of what is invariant about F_0 movements in specific languages). In adopting these as a point of departure, we hope to show that slope and duration, though readily identifiable and quantifiable aspects of the shape of individual accents, are not in any way characteristic of a given accent type. We believe this will help strengthen our contention that any appropriate idealization of accent shape must make reference to the F_0 level and alignment of the targets of which an accent is composed.

I. EXPERIMENT 1

In this experiment, we sought to assess the general plausibility of any model that predicts that the slope and duration of F_0 movements will vary with speech rate. We were not looking for specific segmental anchor points in experiment 1, but only testing the more general prediction that, as speech rate increases, F_0 movements will be shorter and steeper. To some extent this may be regarded as a replication, for English, of Caspers and van Heuven's 1993 results on Dutch rises.

A. Method

Our basic approach in this experiment was to measure the duration and F_0 excursion of rising prenuclear pitch accents in a passage read aloud at three different rates. As noted in the Introduction, we examined only prenuclear rises in order to reduce the amount of uncontrolled variability in our results. We took the hypothesis of constant duration as the null hypothesis in analyzing the duration results: our prediction was that the duration of the pitch accents would be shorter as speaking rate increased. Additionally, we expected that there would be no effect of rate on the size of the F_0 excursion. If both of these predictions are confirmed, then pitch accents cannot be described by either a constant characteristic duration or a characteristic slope.

1. Speech materials

We used a single passage in the style of a news bulletin, specially constructed for the experiment. The passage contained a total of 13 test words, all adjectives or adverbs, set in adjective+noun or adverb+verb phrases such as *minor victory* or *narrowly emerged*. It was expected that speakers would place a rising prenuclear pitch accent on the test words in this context, followed by a nuclear accent on the noun or verb. The stressed syllables of all the test words

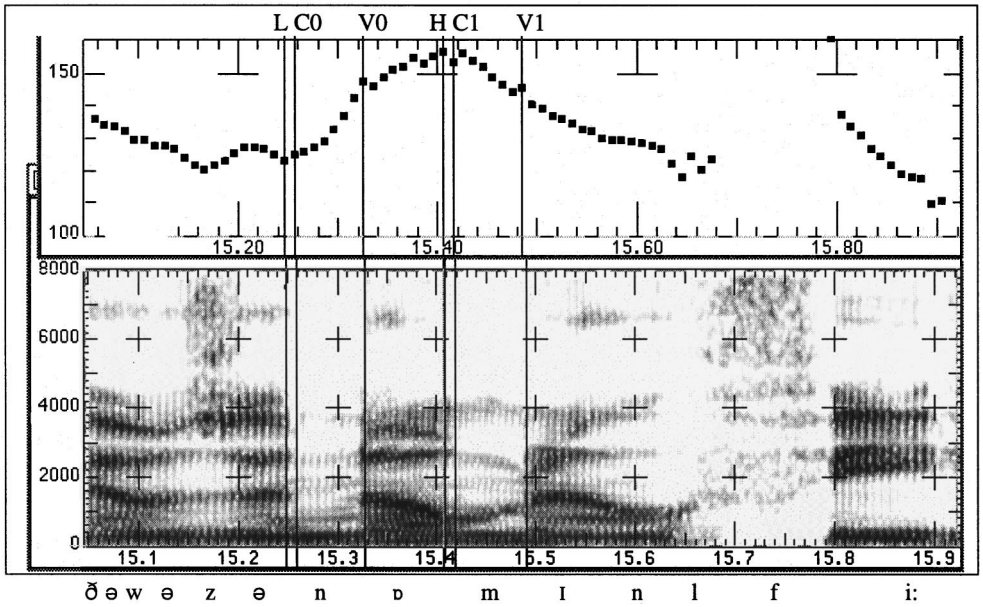


FIG. 1. Measurement points (indicated by labeled vertical lines through the F_0 track and spectrogram) for the fragment ''There was a nominal fee... .''

contained only sonorants to ensure relatively smooth and reliable extracted F_0 contours. The passage, with test words indicated, is given in Appendix A.

2. Speakers

The materials were read by three speakers of Standard Southern British English (''RP''), two males (D.P. and D.F.), and one female (C.L.). D.F. is the second author.

3. Recording and analysis procedures

The materials were recorded on professional equipment in the recording studio of the phonetics laboratory of the Edinburgh University Linguistics Department. Speakers were told that they would be reading aloud at different rates. The speakers determined their own fast, normal, and slow rate. Each speaker was given 15 copies of the passage. Each copy was randomly labeled with one of the required speech rates (fast, normal, slow). There were five copies at each rate.

The recordings were made on DAT tape, and subsequently digitized at 16 kHz and transferred to a Sun Sparc Ultra-1 workstation for analysis using ESPS Waves software. F_0 was extracted using 30-ms Hanning windows, with 20-ms overlap and 10-ms frame steps. F_0 minima and maxima were identified by the second author, in consultation with the first author. All measurements were performed by hand, i.e., not algorithmically. For each test accent, the beginning and the end of the accentual rise were marked on the basis of simultaneous visual displays of the waveform, wideband spectrogram, and F_0 contour. Wherever possible, we identified a clear local F_0 minimum as the beginning of the accent and a clear local maximum as the end, as shown in Fig. 1. These points are referred to collectively in what follows as turning points, and individually as ''*L*'' (for the minimum) and ''*H*'' (for the maximum). Other labels in Fig. 1 are relevant to experiment 2 only. Except in very clear cases, we did not attempt to compensate for spurious values of F_0 or for segmental perturbations; that is, we generally took the absolute minimum and maximum extracted F_0 values in the vicinity of the accented syllable. In most cases it appeared that this procedure gave us something close to the ''true'' beginning and end of the accent. Although microprosodic effects, caused by segment idiosyncracies, may have been present, these were small because of our use of nasals in the test words. In any case, they were similar at all speech rates, as the same passage was used at each of the rates, and thus do not form a confound in the experiment.

B. Results and discussion

In the event, speakers did not realize all the test phrases as intended, and the analyses that follow are based on only nine test words. Even for those nine test words, not all speakers produced usable renditions every time. Unusable renditions were primarily productions of the target word without pitch accents, with hesitations or with flat F_0 contours. As a consequence of discarding unusable renditions there is not a full set of 45 (5 repetitions$\times$9 words) realizations of each word by each speaker. Instead, C.L had 29, 33, and 30 usable renditions at fast, normal, and slow rates respectively, while D.F. had 23, 28, and 31, and D.P. had 33, 38, and 37, respectively. In our ANOVAs we have therefore not used test word as a grouping factor. This, of course, increases the variance and works against our hypotheses.

We first confirmed that the rate manipulation had been successful. This was done, in the first instance, by editing out all pauses and disfluencies from each passage and taking the total duration of the remaining segmental material. Note that, if the amount of pausing contributes to perception of rate, this method actually underestimates the degree of difference between rates. Results are summarized in Fig. 2. A two-way ANOVA with speaker and rate as independent variables confirmed the existence of a large effect of rate $[F(2,36) = 565.682, p < 0.001]$. There was also a smaller main effect of speaker $[F(2,36) = 28.198, p < 0.0001]$ and a small but still significant interaction $[F(4,36) = 7.153, p < 0.0001]$. It

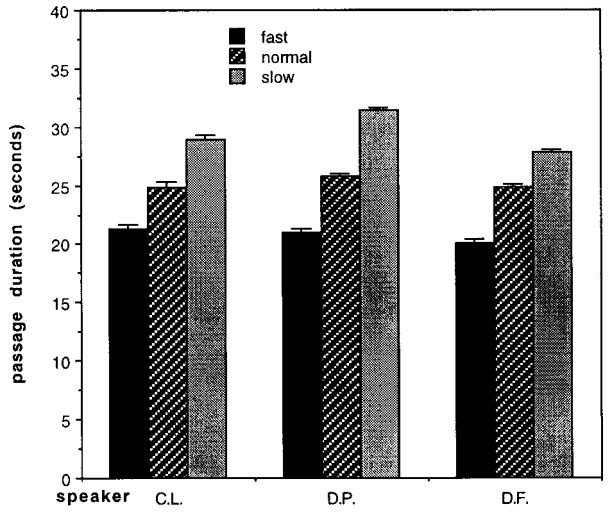


FIG. 2. Passage durations for experiment 1, with standard error bars.

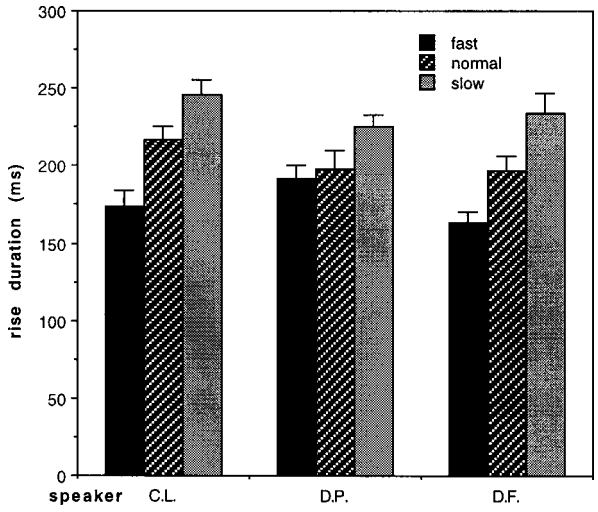


FIG. 3. Mean rise times for experiment 1, with standard error bars.

can be seen from Fig. 2 that these speaker effects are due to difference in overall speaking rate and to differences in the extent of slowing down at slow rate. (Different degrees of slowing were also found in experiment 2, as will be seen below.) An additional analysis was also run, using a smaller portion of the words, as indicated in Appendix A. This served to check whether the portions of the words on which F_0 movements were measured were individually affected by rate changes, as rate does not necessarily affect all portions of a stretch of speech equally. This analysis confirmed that these portions, too, showed a robust effect of rate. The mean duration of the relevant portions at fast rate were 393, 402, and 408 ms for C.L., D.P., and D.F., respectively. The normal rate had means of 480, 457, and 480, while the slow rate had means of 571, 550, and 543, respectively. An ANOVA with rate and speaker as factors showed a main effect of rate, $F(2,273)=13.279, p<0.0001$, but no effect of speaker, and no interactions ($F<1$ and $F<1.3$, respectively).

Having confirmed that the speakers produced the passage and individual portions of the passage at different rates, we then investigated whether rate had an effect on the rise time of the pitch accent. The segmental anchoring hypothesis would predict a clear effect of rate, since the segmental durations (as just shown) were shorter, whereas under a constant-duration model rate should have no effect. We looked for an effect of rate by means of a two-way ANOVA (speaker×rate), with rise time in ms as the dependent variable. There was a clear effect of rate on rise time $[F(2,264)=24.31, p<0.001]$, with no effect of speaker $[F(2,264)=1.6, p>0.2]$ nor any interaction of speaker and rate $[F(4,264)=1.17, p>0.3]$. Mean rise times are shown in Fig. 3.

Finally, we investigated whether there was an effect of rate on F_0 excursion—the difference in Hz between *L* and *H*. Strictly speaking, the segmental anchoring hypothesis makes no predictions about the size of the F_0 excursion, yet the idea of independent *L* and *H* targets leads to the expectation that rate will have little effect on excursion size. By contrast, a constant-slope model would expect a large effect of rate, because (as we just showed) the duration of the accent is not constant: the longer the rise time, the farther the F_0 should rise.

Again, we used a two-way ANOVA (speaker×rate), with F_0 excursion in Hz as the dependent variable. There was a strong effect of speaker [means: 34, 27, and 27 Hz for C.L., D.P., and D.F., respectively; $F(2,264)=9.61, p<0.001$], which is to be expected since speakers have different F_0 ranges, but no effect of rate $[F(2,264)=2.42, p=0.09]$ and no interaction $[F(4,264)=1.492, p>0.2]$. Similar results were obtained when the F_0 excursions were measured in semitones rather than Hz.

In summary, there is a strong effect of rate on rise time, suggesting that a constant-duration model is seriously inadequate. The effect is in the direction predicted by the segmental anchoring hypothesis. There is no effect of rate on F_0 excursion, which, given the effect of rate on rise time, means that a constant-slope model also fails to account for the data.

II. EXPERIMENT 2

Given that the results of experiment 1 were broadly compatible with the segmental anchoring hypothesis, it seemed appropriate to do a larger-scale experiment, using speech materials that would permit us to look for specific segmental anchors for the beginning and ending points of pitch accent rises. This was the goal of Experiment 2.

A. Method

1. Speech materials

There were 14 test sentences. A typical test sentence is *There was a nominal fee for his services*, in which the rising pitch accent on *nominal* occurs after two or three unaccented sentence-initial syllables and is immediately followed by another (rising–falling) pitch accent on the next word (*fee*). As in experiment 1, the rising pitch accents were the subject of our investigation. In what follows we will refer to them as ''test accents,'' and to the word and syllable on which they occur (*nominal* and *no-*, respectively) as the ''test word'' and ''test syllable.''

The test words were chosen according to the following criteria: (1) they have a ''short'' or ''lax'' vowel (/ɪ/ /ɛ/ /ɒ/

or /æ/) in the stressed syllable; (2) there are two (or in a few cases three) unstressed syllables following the stressed syllable in the same word; (3) the consonants flanking the stressed vowel are nasals or /l/. Criteria (1) and (2) were adopted because of evidence that the alignment of F_0 targets is affected by both vowel length [Caspers and van Heuven (1993); Ladd *et al., submitted*[3]] and position of stressed syllable in word [Arvaniti *et al.* (1998)]; criterion (2) also ensured that the F_0 would drop at least slightly after the test accent before rising again for the next pitch accent, thus creating a clear F_0 peak for measurement. Criterion (3) was intended to ensure smooth F_0 contours in the vicinity of the test accent.

The test words were all incorporated into noun phrases such as a *nominal fee* or a *phenomenal interest*. We expected speakers to place two pitch accents on these phrases (i.e., on the two content words, in this case the adjective and the noun). The full list of test sentences is included in Appendix A. In addition to the test sentences, there were 86 fillers of similar length but different structure and phonetic content, for a total of 100 sentences. (Two examples of fillers are: *He read the instructions and they were hopeless* and *Weather forecasts in November are always depressing.*) The reason for using so many fillers was to eliminate any possibility that the speakers would notice the structural and phonetic similarity of the test sentences. The test sentences were randomly interspersed among the fillers, except that the first 21 sentences and the last 11 sentences were exclusively fillers. The list was intended to be read through once, at each of three rates.

2. Speakers

The materials were recorded by eight speakers of Standard Southern British English, four female and four male, all of whom were undergraduates at Edinburgh University in their early 20s. The recordings of two of the female speakers were discarded because they were too monotone to allow reliable marking of the beginnings and ends of pitch accents. The results are therefore based on the recordings of six speakers. The speakers are identified in this report as FJ, FL, MG, MM, MN, and MR, where F or M stands for female or male, and the second letter is an initial. None of the authors served as speakers in this experiment.

3. Recording and analysis procedures

The materials were recorded and transferred to a Sun workstation in the same way as in experiment 1, and acoustic analysis was again done using ESPS Waves software.

Each speaker read the entire list of sentences once in each of three speaking rates. On the first reading no mention was made of speaking rate, and the rate obtained on this reading was taken as ''normal.'' Following the first reading, speakers were asked to read the list again, with rate specified, and then a third time, again with rate specified. Half the speakers were asked to read fast on the second reading and slowly on the third, and the other half were asked to read slowly on the second reading and fast on the third. This was purely a counterbalancing measure and did not feature as a factor in the analyses.

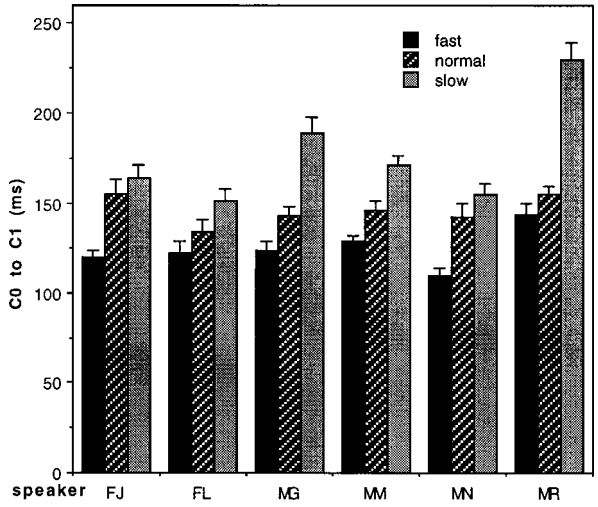


FIG. 4. Mean durations for $C0$ to $C1$, with standard error bars.

F_0 turning points were located for each accent as in experiment 1 (see Fig. 1). In addition, we marked the onset of the initial consonant of the stressed syllable and the offset of the stressed vowel (e.g., the beginning of the /n/ and the end of the /ɒ/ in *nominal*), since informal observations had suggested that those were the approximate segmental anchor points for the beginning and the end of the accent. Subsequently we also marked the onset of the following unstressed vowel (the beginning of the /ɪ/ in *nominal*). In what follows these three anchor points are identified as ''$C0$'' (for the onset of the initial consonant), ''$C1$'' (for the offset of the stressed vowel), and ''$V1$'' (for the onset of the following unstressed vowel). Examples of these segmental measurement points are also displayed in Fig. 1.

B. Results and discussion

1. Confirmation of rate manipulation

As in experiment 1, we began by checking that the subjects had indeed produced significantly different speech rates for the three conditions. This time, we focused on the rate at the test syllable, calculating the intervals between $C0$ and $C1$ for each sentence in each condition. Mean durations for each speaker are shown in Fig. 4. We analyzed these data in a two-way (6×3: speaker×rate) repeated-measures ANOVA, with items as the random factor. Results showed a large main effect of rate [$F(2,26)=240.653$, $p<0.0001$], a main effect of speaker [$F(5,65)=18.245$, $p<0.0001$], and an interaction between speaker and rate [$F(10,130)=9.152$, $p<0.0001$]. Once again, this means that the speech rate manipulation was successful. There were some individual differences, but these were differences in the magnitude of the duration increase across conditions by subject.

2. Alignment and rate

To test the hypothesis that alignment is not affected by speech rate, we did two analyses of the alignment of the turning points. First, we correlated the temporal interval be-

TABLE I. Correlation between duration of test accent and duration of stressed syllable, for each speaker separately. $N=42$.

Speaker	R	P
FJ	0.46	0.0023*
FL	0.41	0.0071*
MG	0.58	0.00001*
MM	0.560	0.00001*
MN	0.86	0.00001*
MR	0.87	0.00001*

tween the turning points (the duration of the rise of the test accent) with the temporal interval between the anchor points (the duration of the test syllable). If the duration of F_0 movements is relatively fixed, then there should be no correlation between the segmental duration of the stressed syllable and the duration of the test accent. If, on the other hand, the duration of the F_0 movement is controlled by fixed segmental anchors, then there should be a strong positive correlation between the two measures.

We did this correlation for each speaker separately in order not to confound the results by speaker-specific differences of speaking rate. Results are shown in Table I. It can be seen that for all speakers there is a strongly significant positive correlation between syllable duration and accent duration, showing that changes in speech rate had a similar effect in this experiment as in experiment 1, i.e., as rate decreases, the durations of rises increase.

As in experiment 1, we also examined the effect of rate on rise times. Initially, we conducted ANOVAs for each speaker separately, with items as the random factor, and rate as a repeated measures fixed factor, and with the duration in ms of the stretch between L and H as the dependent variable. The means, F-ratios, and p-values for each speaker are shown in Table II. In addition, we conducted an ANOVA of rise times as above, but combining all speakers, and including speakers as a second repeated measures fixed factor with six levels. The latter analysis showed that there was a significant main effect of rate $[F(2,26)=84.75, p<0.0001]$. There was also a significant effect of speaker $[F(5,65)=17.76, p<0.0001]$ and a significant interaction $[F(10,130)=10.26, p<0.0001]$. These analyses and the patterns in the means show that speakers all showed longer rise times as rate slowed. The interaction was due to individual differences in magnitude of this lengthening of rise times across different rates.

Given the presence of the correlations and the effect of rate on rise time, we then tested whether speech rate affected

TABLE II. Means rise times in ms for experiment 2, with standard errors in parentheses.

Speaker	Fast	Normal	Slow	$F(2,26)$	P
FJ	163 (15)	209 (16)	211 (11)	6.286	0.006
FL	160 (11)	173 (16)	224 (25)	5.975	0.007
MG	158 (9)	186 (9)	229 (11)	11.450	0.0001
MM	168 (14)	176 (7)	228 (8)	10.954	0.0001
MN	135 (5)	172 (9)	204 (8)	33.612	0.0001
MR	191 (10)	231 (11)	411 (28)	51.381	0.0001
Overall	162 (5)	191 (5)	251 (11)		

TABLE III. Alignment of L (F_0 minimum) relative to $C0$ (beginning of onset consonant of stressed syllable), in ms (standard errors in parentheses), for each of three rates. Negative values indicate that L precedes $C0$.

Speaker	Fast	Normal	Slow	$F(2,26)$	P
FJ	2 (10)	1 (10)	−9 (10)	0.543	ns
FL	23 (8)	7 (11)	12 (9)	1.707	0.172
MG	16 (8)	25 (10)	12 (5)	0.708	ns
MM	−1 (4)	−1 (6)	14 (4)	3.773	0.36*
MN	4 (1)	1 (4)	5 (2)	0.565	ns
MR	−6 (7)	0 (5)	7 (10)	0.804	ns
Overall	6 (3)	5 (3)	7 (3)		

the alignment of the turning point relative to its anchor point. We did this by means of an ANOVA on the alignment of each turning point relative to its anchor point, with speaker and speech rate as the main independent variables. Our prediction here was that there would be no effect of speech rate, i.e., that the alignment of the turning point and corresponding anchor point would remain constant under changes in rate. We did these analyses separately for the F_0 minimum (L) and the F_0 maximum (H).

To examine the effect of rate on the alignment of the L, we ran ANOVAs for each subject, with items as a random factor, and rate (fast, normal, slow) as a fixed factor (repeated measures), with the duration in ms of the stretch between L and $C0$ as the dependent variable. The means, F-ratios and p-values for each speaker are presented in Table III.

An additional ANOVA, incorporating the factors above, but including all speakers, and adding speakers as a within-items fixed factor, showed a significant main effect of speaker $[F(5,65)=3.015, p=0.017]$, but no significant main effect of rate $[F(2,26)<1]$, nor an interaction between speaker and rate $[F(10,130)=1.430, p=0.174]$. This shows that, although speakers have slightly idiosyncratic ways of aligning, the interval between L and $C0$ remains stable across speech rates, except in the case of speaker MM's slow rate.

We used the same approach to investigate the effect of rate on the alignment of the F_0 maximum (H), this time based on the interval between $C1$ and H. The segmental anchoring hypothesis predicts that, as for the alignment of L with $C0$, there should be no significant effect of rate on the alignment of the H with $C1$. By contrast, given the results just reported for L alignment, the hypothesis that accent movements are of fixed duration would predict earlier alignment of H as rate decreases. Results of the analysis are given in Table IV.

It can be seen that three speakers show a significant effect of rate—H is aligned later relative to $C1$ as rate decreases. One speaker shows a nonsignificant tendency in the same direction, with the other two showing earliest alignment at normal rate, slightly later at fast rate, and considerably later at slow rate. An overall analysis with rate and speakers as fixed factors (as for L alignment above) showed a significant main effect of rate $[F(2,26)=31.032, p<0.0001]$, a significant main effect speaker $[F(5,65)=11.8837, p<0.0001]$, and a significant interaction between speaker and rate $[F(10,130)=6.657, p<0.0001]$. While the

TABLE IV. Alignment of H (F_0 maximum) relative to $C1$ (offset of vowel of stressed syllable), in ms (standard errors in parentheses), for each of three rates. In all cases H follows $C1$ by the interval shown, i.e., the F_0 maximum is reached during or after the consonant that follows the stressed syllable.

Speaker	Fast	Normal	Slow	$F(2,26)$	P
FJ	41 (10)	53 (12)	57 (4)	0.942	ns
FL	16 (13)	32 (13)	60 (17)	4.632	0.019*
MG	18 (10)	17 (12)	28 (11)	0.477	ns
MM	39 (13)	31 (4)	43 (7)	0.804	ns
MN	22 (3)	29 (3)	44 (6)	6.414	0.005*
MR	52 (7)	76 (10)	174 (23)	25.076	0.0001*
Overall	31 (4)	40 (4)	69 (8)		

segmental anchoring hypothesis is therefore not fully confirmed, it nevertheless fares better than the competing assumption that F_0 movements have a fixed duration, because the effect of rate is in the opposite direction to what would be predicted under that assumption. Moreover, there are two reasons to suspect that the picture presented in Table IV sheds an unnecessarily unfavorable light on the segmental anchoring hypothesis.

First, as we noted above, the mean alignment values for H are always some distance after the end of the stressed syllable. It may be that $C1$ (the end of the stressed vowel) is not the ''true'' segmental anchor for the H and that therefore it is not the most appropriate reference point for our analyses. We therefore redid the analysis of H alignment, this time measuring the alignment relative to ''$V1$,'' the onset of the unstressed vowel following the stressed syllable. Results are shown in Table V. It can be seen that the effect for speaker MN completely disappears and that for FL is reduced so that it ''misses significance,'' though a new, significant effect is created for MM. MR's effect is as strong as ever. Except for MR, however, it could be suggested that there is a ''true'' anchor point sometime late in the consonant immediately following the stressed syllable, probably varying somewhat from speaker to speaker, and that alignment of H relative to this point is substantially unaffected by speaking rate. This would certainly predict the results more successfully than any theory based on constant rate or (as we shall see) constant slope.

To explore this possibility further, we calculated the alignment of H as a proportion of the stretch from $C1$ to $V1$(H-$C1$/$V1$-$C1$). The rationale is based on the idea that the alignment of H may be relative to a particular point in a segment, which is not necessarily the segment boundary. As the compression of segments across speech rates may not be the same, this means that the use of a segment boundary may mask a regularity that would be noticed if a proportional measure is used. Table VI shows the means for this measure, along with results of analyses of the effect of rate for each speaker individually. An overall analysis of these same data, including speaker as a factor, showed that there was a significant main effect of rate ($F2,26) = 4.149, p = 0.027$, speaker [$F(5,65) = 5.425, p < 0.0001$], and a significant interaction [$F(10,130) = 2.232, p < 0.02$]. However, simply removing the exceptional speaker MR from this analysis removes both the effect of rate [$F(2,26) < 1$] and the interaction between speaker and rate [$F(8,104) = 1.28$, $p = 0.26$], leaving only the effect of speaker significant ($F4.52) = 3.308$, $p = 0.017$. The latter analysis, along with the fact that four out of six speakers show no significant effect of rate on the alignment of H using this measure, strengthens our claim that the alignment of H to a target somewhere in the segmental stream is reasonably constant, regardless of speech rate. For the two speakers that do show effects of rate, the alignment difference is in the wrong direction for the constant rise time hypothesis.

TABLE V. Alignment of H (F_0 maximum) relative to $V1$ (onset of unstressed vowel following stressed syllable), in ms (standard errors in parentheses), for each of three rates. Negative values indicate that H precedes $V1$.

Speaker	Fast	Normal	Slow	$F(2,26)$	P
FJ	10 (12)	15 (14)	15 (3)	0.083	ns
FL	−20 (12)	−3 (14)	19 (8)	3.047	0.065
MG	−16 (10)	−30 (13)	−27 (13)	0.610	ns
MM	11 (14)	−14 (5)	−4 (7)	3.499	0.045*
MN	−20 (6)	−18 (9)	−10 (5)	0.919	ns
MR	10 (6)	27 (11)	117 (23)	20.310	0.0001*
Overall	−4 (4)	−4 (5)	18 (7)		

TABLE VI. Alignment of H as a proportion of $C1$ to $V1$. A value of *1* indicates alignment at $V1$. Lower values indicate earlier alignment, and higher values indicate later alignment.

Speaker	Fast	Normal	Slow	$F(2,26)$	P
FJ	2.44 (1.0)	1.84 (0.58)	1.43 (0.09)	0.704	ns
FL	0.39 (0.35)	0.97 (0.36)	1.58 (0.42)	5.137	0.013*
MG	0.46 (0.39)	0.42 (0.22)	0.63 (0.23)	0.204	ns
MM	1.13 (0.54)	0.76 (0.09)	0.94 (0.12)	0.381	ns
MN	0.61 (0.10)	0.74 (0.10)	0.84 (0.08)	2.141	ns
MR	1.31 (0.17)	1.81 (0.37)	3.19 (0.44)	10.968	0.0001*
Overall	1.05 (0.22)	1.09 (0.15)	1.44 (0.14)		

The second reason not to take the results in Table IV as outright refutation of the segmental anchoring hypothesis is that the data for the slow rate are probably affected by the presence of prosodic boundaries in the middle of the test phrases. That is, in the slow rate some of the speakers often appeared to break up phrases like *a nominal fee* with a prosodic boundary between the adjective and the noun. (This, of course, is entirely consistent with the findings cited at the beginning of the paper.) These boundaries were manifested in the F_0 trace as a leveling-off or very gradual rise following the accentual rise, rather than the expected fall. The difference between the two types of contours can be seen in Fig. 5. Because our acoustic analysis procedures looked strictly for the F_0 maximum, they would have been affected by intonational realizations of this sort: the F_0 maximum in contours like those of Fig. 5(b) would have been measured quite far to the right of the point that, impressionistically, is the end of the accentual rise and the beginning of the level stretch. We could, of course, have attempted to estimate the end of the accentual rise by eye, but this would have introduced an unacceptably large element of subjectivity into the measurements.

Instead, as an indirect way of testing the idea that added phrase boundaries in the slow condition were responsible for the effect of rate observed in Tables IV and V, we redid the ANOVAs using only fast and normal rates. The data using

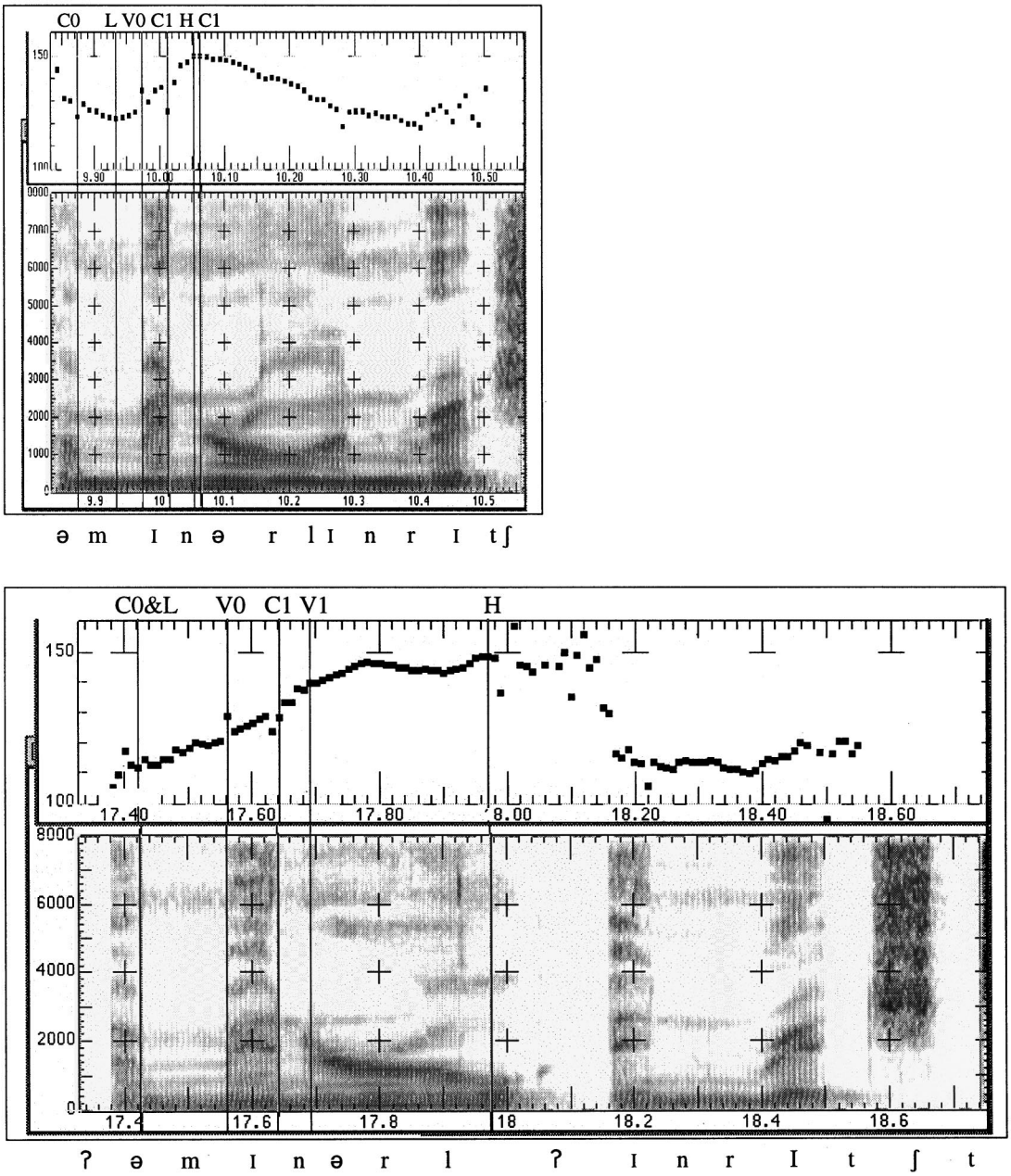


FIG. 5. Time-aligned F_0 contours and spectrograms for the fragment ''...a mineral enriched.... .'' The fast rate is displayed in the upper panel and the slow rate in the lower panel. Measurement points are indicated by labeled vertical lines through the F_0 track and spectrogram. Note the clear peak for the fast rate and the plateau for the slow rate.

$C1$ as an anchor showed a much reduced effect of rate [$F(1,13)=4.824, p=0.047$], which did not interact with the factor speaker [$F(5,65)=1.146, p>0.3$], although the factor speaker itself was still strongly significant [$F(5,65)=4.531, p=0.001$]. However, in the same analyses using $V1$ as an anchor, only the factor speaker remained significant [$(F5,65)=3.914, p=0.004$], whilst the effect of rate disappeared [$(F1,13)<1$, ns]. There was no interaction between the two factors [$(F5,65)=1.878, p>0.1$]. This seems to confirm that there is something special about slow rate.

3. Target F_0 level and rate

So far we have been testing the segmental anchoring hypothesis against what we have called the constant duration hypothesis, i.e., the view that F_0 movements should take an approximately fixed amount of time. We have not yet addressed the constant slope hypothesis. Even if there is something like fixed segmental anchoring, it could be that F_0 nevertheless changes at a roughly constant rate, and that therefore F_0 excursions would be greater as rate decreases, because there would be more time between anchor points. As we saw in the Introduction, this was explicitly claimed by Ashby (1978) for English nuclear rises and strongly suggested for French by Fougeron and Jun (1998). This question is addressed in the following analyses.

To examine the effect of speech rate on F_0 excursions, we calculated the F_0 differences between L and H in semitones [cf., e.g., 't Hart *et al.* (1990), p. 24] and did a one-way

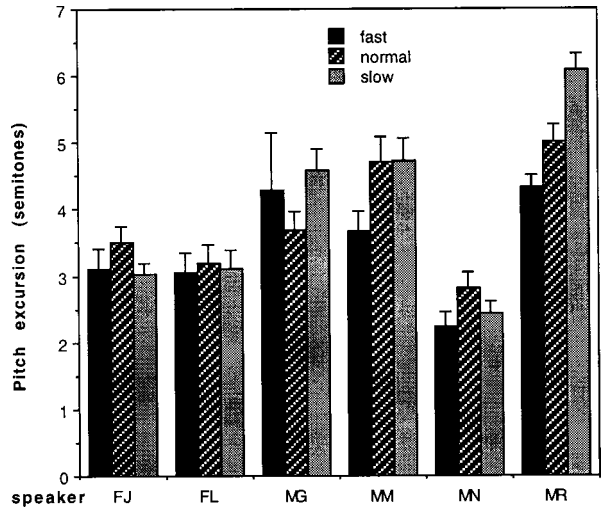


FIG. 6. F_0 excursions for experiment 2, with standard error bars.

ANOVA with rate as the fixed factor for each speaker separately. We also ran the same analysis measuring F_0 differences in equivalent rectangular bandwidth (ERB) units[4] [Hermes and van Gestel (1991)] rather than in semitones, but the pattern of results was unaffected. Results for the semitone analysis are shown in Fig. 6.

Upon analysis, MR showed a highly significant effect of rate on F_0 excursion size [$F(2,26)>13, p<0.0001$], in the direction that would be predicted by the constant-slope hypothesis. For all other speakers there is no significant effect of rate on F_0 excursion, though for MM and MN the effects approach significance [$F(2,26)=3.4, p=0.051$ for MM; $F(2,26)=3.0, p=0.067$]. MM's effect is the same as MR's (increasing F_0 excursions with decreasing rate), but MN's effect shows normal rate to have the largest F_0 excursions, with both fast and slow showing smaller F_0 excursions. There is thus no strong support for the constant-slope hypothesis, though it cannot be ruled out entirely. Note that the larger size of MR's F_0 excursions at slow rate could be largely explained by the presence of a ''high boundary tone'' or additional boundary-related rise following the accentual rise itself.

Another approach to determining whether F_0 movements are aiming at a constant slope is to correlate F_0 excursion size with accent duration. Obviously, the constant-slope hypothesis would predict a high correlation between duration and excursion size. Results of this analysis, for each speaker separately, are shown in Table VII.

Once again, MR shows a considerable dependence of excursion size on accent duration; so, surprisingly, does MG. The other speakers show no significant relationship. As before we redid the same analyses in ERB and semitone units. The correlations are slightly larger using these two measures, but there is no difference in the pattern of results.

TABLE VII. Correlations between rise time (duration L to H in ms and excursion size (F_{10} increase L to H in Hz) by speaker (all rates combined).

Speaker	R	P
FJ	0.26	0.10 ns
FL	0.21	0.19 ns
MG	0.39	0.009*
MM	0.05	0.77 ns
MN	0.07	0.64 ns
MR	0.54	0.0002*

Overall, there clearly are effects of the sort predicted by the constant-slope hypothesis, but as a general account of the slope and duration of pitch accents it appears that the constant-slope hypothesis by itself is inadequate. If, on the other hand, we start from the segmental anchoring hypothesis, we will assume that F_0 excursion size (which depends on the F_0 level of the tonal targets) is roughly fixed for a given speaker and linguistic context. Within that approximate phonological framework, we can then acknowledge the existence of small phonetic effects of speech rate on the precise details of target level. Since the durational evidence for the segmental anchoring hypothesis is so strong, we feel that such an account is the most appropriate way to deal with the limited evidence in favor of the constant-slope view. The efforts of studies like Fougeron and Jun's should therefore be directed at ascertaining the effects of rate on the scaling of individual targets, not on slope or size of F_0 excursion as such.

III. GENERAL DISCUSSION

As we noted in the Introduction, it is almost universally assumed that the ''shape'' of a pitch accent is independent of its alignment with the segmental string. ''Shape'' has been defined in a number of ways by different investigators, but at least some have taken either slope or duration as essential attributes of accent types and have predicted that these attributes will remain roughly constant in a variety of contexts. Rather than attempt to evaluate many different models separately, we have therefore taken what we call the ''constant-slope'' and ''constant duration'' hypotheses as a baseline against which to evaluate our own approach. The essential claim of our approach—and the logical conclusion of the autosegmental ''phonological'' approach to intonation developed by Bruce, Pierrehumbert, and others—is that accent shape has no definition independent of the linguistically specified alignment and F_0 level of the targets of which the accent is composed. F_0 movements are merely transitions from one independently specified F_0 target to another. (It is, of course, possible that the ''mere transitions'' could have distinctively different shapes, but this does not change the fact that the F_0 level and segmental anchoring of the targets severely constrain the ''shape'' of any accent, not the other way around.)

Empirical support for this view comes from Arvaniti *et al.*'s finding that both the beginning and the end of rising accents in Greek are anchored to segmental landmarks. In the present paper, we have used a different methodology, a different experimental task, and a different language to test the generality of the phenomenon of ''segmental anchoring.'' Our results make it clear that there is no evidence for anything like the assumption that F_0 movements have a constant characteristic duration. There is a small amount of evidence that F_0 excursions are larger as rate slows down, but the effect is too small and too restricted to certain speakers to

offer any serious support for the assumption that F_0 movements have a constant characteristic slope. By contrast, the effects we have found are almost entirely in line with the segmental anchoring hypothesis suggested by Arvaniti *et al.*'s work on Greek. For all but one of our speakers, there was clear evidence that both the beginning and the end of the accentual rises investigated are anchored to specific locations in the segmental string: the beginning (*L*) of the rise is timed to co-occur with the beginning of the onset consonant of the stressed syllable, and the end (*H*) is timed to occur somewhere late in the consonant following the stressed vowel. For the one exceptional speaker, there is reason to believe that his results were distorted by the extreme slowness of his slow rate delivery, with the attendant addition of many prosodic boundaries.

We believe that the results presented here provide support for the notion that the primary determinants of the shape of a pitch accent are the alignment and F_0 level of the specific targets that make it up. We also believe that our results would justify attempts to design a model of intonation based on the segmental anchoring of targets for use in speech technology.

ACKNOWLEDGMENTS

We would like to thank Ineke and Johanneke Caspers Mennen for her help and advice. We also thank Paul Taylor for useful discussions of theoretical issues. We thank Anders Löfqvist and two anonymous reviewers for their helpful comments on an earlier version of this paper. We are grateful for a small grant from the Arts, Divinity and Music Faculty Group Research Fund at Edinburgh University, which made possible the acoustic analysis of much of the data from experiment 2. Astrid Schepman's collaboration was funded by the UK Economic and Social Research Council through Grant No. R000237447 to Edinburgh University, as part of the project ''Alignment of Fundamental Frequency Targets in English and Dutch.''

APPENDIX A: SPEECH MATERIALS

Experiment 1—A number in the text indicates that the word immediately following it was a target word in the experiment. Numbers in bold italics indicate that the test word was discarded (see Sec. II A for more details). Letters printed in italics represent segments that were entered as portions in the segmental rate manipulation analysis (not indicated for discarded items).

Bertie Aherne (1) *narrowly emerged* as the (2) *newly elected* Irish Prime Minister yesterday, although the (***3***) worrying size of his (***4***) minor victory and *the* (5) *remaining* dependence on the *support of* (6) *meaningless* parties and (7) minor independents (***8***) really seemed likely to herald instability in the Republic. Mr. Aherne's (***9***) rolling majority left his (10) *many* adversaries astounded. The (11) *roaming* nationalist threat continues to be their shared political bandwagon and this (12) *mainly* accounted for their poor performance in the polls. *A* (13) *mori*bund Mr. Bruton conceded defeat early in the proceedings.

Experiment 2—test syllables are printed in bold.

(1) There was a **no**minal fee for his services.
(2) There is a phe**no**menal interest in the products.
(3) She got a u**na**nimous vote for the proposal.
(4) They got an a**no**nymous call from a witness.
(5) She's a **mi**nister's wife in the Home Counties.
(6) There was an a**no**malous reading in the data.
(7) There were **mo**nogrammed sheets in the hotel rooms.
(8) There is a **mi**nuscule chance of surviving a plane crash.
(9) I need a **mo**nosyllabic word in my crossword puzzle.
(10) They sentenced the **mi**litant splinter group to five years.
(11) You need a **mi**neral and vitamin supplement to get well.
(12) They charge a **mi**nimum rate for the use of their phone lines.
(13) He took a **mi**neral enriched supplement every day.
(14) They showed a **mi**nimal interest in what he had to say.

APPENDIX B. SUPPLEMENTARY DATA: EFFECTS OF SPEAKING RATE ON OVERALL LEVEL

Although it is not strictly relevant to the segmental anchoring hypothesis, we also analyzed the absolute values of *L* and *H* in both experiments, to see if there was any overall raising or lowering effect of rate changes. As noted in the Introduction, Cooper and Sorensen (1981) found that when speakers were asked to read sentences at different rates, overall fundamental frequency (F_0) was slightly but significantly lower at slow rate than at fast or normal rates. Rietveld and Gussenhoven (1987), in what may be a related finding, showed that higher-pitched monotone contours were perceived as faster than lower-pitched ones. However, more detailed recent findings on overall F_0 level are somewhat contradictory. Caspers and van Heuven found an overall raising of level at fast rate (which is consistent with Cooper and Sorensen and Rietveld and Gussenhoven), but they note that their two speakers raised F_0 level in different ways: one

TABLE BI. High and low tones (Hz) for experiments 1 and 2. Standard errors in parentheses.

Speaker	Fast	Normal	Slow
High tones:			
CL (1)	207 (3)	208 (3)	197 (3)
DP (1)	151 (3)	150 (2)	147 (1)
DF (1)	110 (1)	113 (1)	109 (1)
FJ (2)	340 (5)	364 (7)	328 (4)
FL (2)	297 (4)	315 (5)	299 (5)
MG (2)	149 (4)	151 (2)	153 (2)
MM (2)	139 (3)	149 (2)	137 (2)
MN (2)	123 (1)	135 (2)	120 (1)
MR (2)	145 (2)	146 (2)	148 (2)
Low tones:			
CL (1)	174 (2)	169 (2)	165 (2)
DP (1)	122 (1)	123 (1)	122 (1)
DF (1)	86 (1)	83 (1)	80 (2)
FJ (2)	284 (3)	297 (7)	275 (3)
FL (2)	249 (4)	262 (3)	249 (4)
MG (2)	118 (5)	122 (1)	118 (2)
MM (2)	113 (1)	114 (2)	104 (2)
MN (2)	108 (1)	115 (1)	104 (1)
MR (2)	113 (1)	109 (1)	104 (2)

raised F_0 minima more than maxima and so had smaller F_0 excursions (measured in semitones) at fast rate, while the other raised F_0 maxima more than minima and therefore had larger F_0 excursions. Fougeron and Jun (1998) found for all three of their speakers that F_0 excursions were smaller at fast rate than at normal rate (like the first of Caspers and van Heuven's speakers), but they also found that speakers varied in the details: two speakers lowered both maxima and minima but lowered maxima more, while one speaker kept maxima constant and raised minima.

Our experiments provide data that are directly relevant to these issues. In both experiments, clear effects of rate on excursion size and overall level were found. Both *L* and *H* are highest at the normal rate and somewhat lower at both fast and slow rates, though the details vary substantially from speaker to speaker, and there are some speakers whose fast rate productions are higher rather than lower, in line with the findings of Caspers and van Heuven. We report our detailed findings in Table BI for the benefit of interested researchers.

[1]Caspers and van Heuven (1993) is actually only one of several papers based on Caspers' Ph.D. research [Caspers (1994)]. We deal only with this one, on the grounds that it is the most widely accessible and most directly relevant to our hypotheses.

[2]Experiment 1 was carried out as part of the second author's dissertation project for the M.Sc. in Speech and Language Processing at Edinburgh University in the summer of 1997. Part of experiment 2 was the basis of the third author's Honours dissertation project for the degree in English Language and Linguistics at Edinburgh University in the academic year 1997–1998.

[3]Ladd, Mennen, and Schepman's work (submitted) shows that, in Dutch, the peak of a prenuclear rising accent is typically aligned at the end of the accented vowel if the vowel is phonologically ''tense'' or ''long,'' but aligned during the following consonant if the vowel is phonologically ''lax'' or ''short.'' It appears that the difference cannot be attributed to differences in the actual physical duration of the vowel, but depends on the syllable structure which in turn is determined by the phonological length or tenseness of the vowel.

[4]Hermes and his colleagues [see also Hermes and Rump (1994)] have argued that the ERB (equivalent rectangular bandwidth) scale, based on psychophysical research by Glasberg and Moore (1990), is the most appropriate scale for expressing the perceptual equivalence of speech F_0 excursions in different voice ranges.

Arvaniti, A., Ladd, D. R., and Mennen, I. (**1998**). ''Stability of tonal alignment: The case of Greek prenuclear accents,'' J. Phonetics **26**, 3–25.

Ashby, M. G. (**1978**). ''A study of two English nuclear tones,'' Language and Speech **21**, 326–336.

Bruce, G. (**1977**). *Swedish Word Accents in Sentence Perspective* (Gleerup, Lund).

Caspers, J. (**1994**). ''Pitch movements under time pressure: Effects of speech rate on the melodic marking of accents and boundaries in Dutch,'' Doctoral dissertation, University of Leiden, The Netherlands.

Caspers, J., and van Heuven, V. J. (**1993**). ''Effects of time pressure on the phonetic realization of the Dutch accent-lending pitch rise and fall,'' Phonetica **50**, 161–171.

Cooper, W., and Sorensen, J. (**1981**). *Fundamental Frequency in Sentence Production* (Springer-Verlag, New York).

Fougeron, C., and Jun, S. A. (**1998**). ''Rate effects on French intonation: Prosodic organization and phonetic realization,'' J. Phonetics **26**, 45–69.

Fujisaki, H. (**1983**). ''Dynamic characteristics of voice fundamental frequency in speech and singing,'' in *The Production of Speech*, edited by P. F. MacNeilage (Springer-Verlag, New York), pp. 39–55.

Fujisaki, H., and Sudo, H. (**1971**), ''A generative model for the prosody of connected speech in Japanese,'' Annual Report of the Engineering Research Institute, University of Tokyo, Vol. 30, pp. 75–80.

Glasberg, B. R., and Moore, B. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hearing Res. **47**, 103–138.

Goldman-Eilser, F. (**1968**). *Psycholinguistics: Experiments in Spontaneous Speech* (Academic, New York).

Hermes, D., and van Gestel, J. (**1991**). ''The frequency scale of speech intonation,'' J. Acoust. Soc. Am. **90**, 97–102.

Hermes, D., and Rump, H. H. (**1994**). ''Perception of prominence in speech intonation induced by rising and falling pitch movements,'' J. Acoust. Soc. Am. **96**, 83–92.

Ladd, D. R. (**1996**). *International Phonology* (Cambridge University Press, Cambridge).

Ladd, D. R., Mennen, I., and Schepman, A. (submitted). ''Phonological conditioning of peak alignment in rising pitch accents in Dutch.''

Pierrehumbert, J. (**1980**). ''The phonology and phonetics of English intonation,'' Doctoral dissertation. Distribution: Bloomington, IN: Indiana University Linguistics Club.

Prieto, P., van Santen, J., and Hirschberg, J. (**1995**). ''Tonal alignment patterns in Spanish,'' J. Phonetics **23**, 429–451.

Rietveld, A. C. M., and Gussenhoven, C. C. (**1987**). ''Perceived speech rate and intonation,'' J. Phonetics **15**, 273–285.

Taylor, P. A. (**1998**). ''The Tilt Intonation Model,'' *Proceedings of ICSLP V*, Sydney, December 1998.

't Hart, J., Collier, R., and Cohen, A. (**1990**). *A Perceptual Study of Intonation: An Experimental-Phonetic Approach to Speech Melody* (Cambridge University Press, Cambridge).

Xu, Yi (**1998**). ''Consistency of tone-syllable alignment across different syllable structures and speaking rates,'' Phonetica **55**, 179–203.

The relative perceptual distinctiveness of initial and final consonants in CVC syllables

Melissa A. Redford[a] and Randy L. Diehl
Department of Psychology, University of Texas, Austin, Texas 78712

(Received 24 November 1997; revised 3 February 1999; accepted 3 May 1999)

Among the world's languages, syllable inventories allowing only initial consonants predominate over those allowing both initial and final consonants. Final consonants may be disfavored because they are less easy to identify and/or more difficult to produce than initial consonants. In this study, two perceptual confusion experiments were conducted in which subjects identified naturally produced consonant–vowel–consonant syllables in different frame sentences. Results indicated that initial consonants were significantly more identifiable than final consonants across all conditions. Acoustic analyses of the test syllables indicated that the relative identifiability of initial and final consonants might be explained in terms of production differences as indicated by the greater acoustic distinctiveness of initial consonants. © *1999 Acoustical Society of America.* [S0001-4966(99)04009-6]

PACS numbers: 43.71.Es [WS]

INTRODUCTION

It is natural to assume that sound patterns that are widely attested among languages may be explained in terms of human biology. According to one such class of explanations, preferred sound patterns arise in response to performance factors involving perception, production, or both (e.g., Lindblom *et al.*, 1984; Diehl and Lindblom, 1996). This class of explanations has been used to account for sound patterns in infant babbling (Davis and MacNeilage, 1995), prevalent feature cooccurrences within sounds in languages (Diehl and Kluender, 1989), and the structure of vowel systems (Lindblom, 1986) and consonantal systems (Willerman, 1994).

Bell and Hooper (1978) made a variety of observations regarding cross-language syllable structure with reference to word units. For instance, they noted that (1) initial consonants predominate over final consonants and that final vowels predominate over initial vowels, (2) the consonants that occur in final position are generally a subset of the consonants that can occur in initial position, and (3) consonant clusters are disfavored across languages, but are more likely to occur in initial position rather than final position. These typological facts suggest that syllable inventories across languages favor syllables with simple (single) consonantal onsets over those with consonantal offsets.

It is also likely that syllables with initial consonants predominate even in languages that allow syllable-final consonants. For instance, the most frequent syllable type in English is the consonant–vowel (CV) syllable (Greenberg, 1998). Historical processes also indicate a preference for initial consonants over final consonants. For example, a common sound change is final consonant weakening and loss (Hock, 1986). In this paper, we examine whether the syllable structure preference for initial consonants over final consonants may be attributed to perceptual and/or production factors. We first report results of two perceptual confusion experiments designed to assess the relative identifiability of initial and final consonants, and then provide evidence, in the form of acoustic measurements on the test stimuli, that the perceptual results may be explained in part by production differences between initial and final consonants.

Ample evidence suggests that syllable-initial and syllable-final consonants are perceived differently by listeners. The selective adaptation paradigm, for example, has shown that repetitive presentation of CV syllables fails to elicit adaptation effects in the perception of vowel–consonant (VC) syllables composed of the same vowel and consonant (Ades, 1974; Wolf, 1978; Samuel, 1989). However, evidence regarding a perceptual advantage for syllables with consonantal onsets over those with consonantal offsets is relatively scant and inconclusive because the question has not been the focus of research.

Van Wieringen (1995) reviewed studies of plosive identification as a function of formant transitions into or out of the syllabic nucleus and reported that eight studies found that the initial (CV) transition was perceptually more distinctive, while the remaining six studies found that the final (VC) transition was more distinctive. Evidence based on other acoustic cues also leads to conflicting conclusions regarding the relative identifiability of initial and final consonants. For instance, Raphael and Dorman (1980) found that the shorter the interval of silence between a final consonant and a following initial consonant, the more subjects report hearing only the initial consonant. On the other hand, Winitz *et al.* (1972) found that voiceless stop release bursts of similar duration with and without accompanying vowel formant transitions provide at least as much information about consonant place of articulation for phrase-final as for phrase-initial consonants.

It should be noted that Raphael and Dorman (1980),

[a]Electronic mail: mredford@utxsvs.cc.utexas.edu

Winitz *et al.* (1972), and each of the studies reviewed by Van Wieringen (1995) used either synthetic or digitally altered stimuli which were typically presented in isolation. As such, the studies were not designed to address the question of whether naturally produced consonants are generally more identifiable in one syllable position versus another.

Results of perceptual confusion studies that use naturally produced, unaltered stimuli have also been somewhat mixed, although they tend to show a perceptual advantage for initial consonants over final ones. Miller and Nicely (1955) found that, overall, initial stop consonants in isolated CV syllables (V=/a/) were less confusable than final stops in isolated consonant–vowel–consonant (CVC) syllables (with CV =/ta/). Wang and Bilger (1973) reported that consonants were less confusable in CV than in VC syllables when the vowel was /a/ or /u/, but the reverse was true when the vowel was /i/. Unlike Miller and Nicely (1955) or Wang and Bilger (1973), Helfer and Huntley (1991) embedded CV and VC test syllables in a carrier phrase, thus approximating more natural listening conditions. In many cases they found no position differences in consonant identifiability, but when such differences did occur, initial consonants were favored over final consonants.

While previous studies lend some support to the claim that initial consonants are more identifiable than final ones, a more complete test is needed. First, it is appropriate that the target syllables occur within a carrier phrase (as in Helfer and Huntley, 1991) so as to approximate more natural conditions of phonetic production and perception than those found in previous perceptual confusion experiments. Constraints that apply only to the production and perception of isolated syllables would appear to be unlikely sources of cross-language syllable structure preferences. Second, once a decision is made to embed the target syllables in a carrier phrase, a question arises about how best to structure the phonetic context in the immediate vicinity of each syllable. With the exception of /s/, English consonants in initial position occur invariably before a vowel or sonorant consonant. Final stop consonants often occur before a syllable-initial consonant and therefore might lack place cues associated with a vocalic or sonorant environment. Accordingly, a general test of the relative identifiability of consonants across initial and final syllable position should include opportunities for the final consonant to appear before both a vowel and a consonant. Finally, if consonant identifiability turns out to be position dependent, acoustic analyses should be performed to evaluate whether this perceptual effect may be related to position-dependent differences in consonant production.

We therefore designed a perceptual confusion experiment to directly test the relative identifiability of initial and final consonants under conditions that could be generalized to the conditions under which phonological systems emerge. Naturally spoken CVC syllables were presented in frame sentences, which provided both a consonantal and a vocalic context for the final consonant.

The segments /p, t, k, f, θ, s, ʃ/ and /i, u, a/ were used to create the CVC syllables. With the exception of /θ/, these segments have a wide distribution in the world's languages. According to Maddieson's (1984) analyses of 317 languages from diverse language families, the plain voiceless stops /p, t, k/ are the most commonly occurring consonants in languages (92% of the languages examined have all three), /s/ is the most common voiceless fricative in languages which have fricatives (84%), followed by /ʃ/ and /f/ (46% and 43%). The vowels /i/, /u/, and /a/ are by far the most common among world's languages (92%, 84%, and 88%, respectively).

Predictions were based on the hypothesis that syllable structure preferences across languages reflect differences between initial and final consonants in perceptual distinctiveness. Specifically, it was predicted that initial consonants would yield fewer overall perceptual confusions errors than final consonants. We also predicted that this position effect would occur across consonants, vowel nuclei, and talkers. This assumes that differences in perceptual distinctiveness are associated with syllabic position generally and not with phoneme- or talker-specific production contexts and conditions. A final prediction was based on the typological fact that clusters are disfavored, but more likely to occur in initial position rather than in final position. Because preconsonantal final consonants might behave as though they formed part of a consonant cluster, it was predicted that they would be more confusable than prevocalic final consonants.

I. PERCEPTUAL EXPERIMENT 1

A. Method

1. Stimuli

The consonants /p, t, k, f, θ, s, ʃ/ and the vowels /i, a, u/ were combined to form CVC target syllables. Each consonant occurred with every other consonant in both syllable positions and with three different vowel nuclei ($7\times7\times3=147$).[1] Two male and two female speakers of American English produced each target syllable in two frame sentences ($147\times4\times2=1176$). The frame sentences which yielded a consonantal context for the syllable-final consonant was *Say* CVC *some more*. The vocalic context was produced with *Say* CVC *again*. The sentences were recorded using the Kay Elemetrics Computerized Speech Laboratory with a Shure SM48 microphone. The root-mean-squared (rms) amplitude of each stimulus sentence was normalized across talkers using a waveform editor developed in our laboratory. The rms amplitude was calculated for the entire utterance, which was then scaled so that the average amplitude of the entire utterance was consistent across all utterances.

2. Subjects and procedure

Subjects were seven female and seven male native American-English-speaking college students. They were instructed to listen for the CVC target syllables embedded in the two frame sentences and to write each syllable in normal orthography on the response sheet provided. Subjects were informed what consonants and vowels would appear in the syllable set, but were encouraged to write down whatever they heard even if it did not correspond to these sounds. To simulate natural listening conditions, stimulus sentences were presented in a low level of pink noise[2] (15-dB signal-to-noise ratio). Subjects were seated in a second-attenuated

TABLE I. Confusion matrices for initial consonants (top) and for preconsonantal (middle) and prevocalic (bottom) final consonants. Subjects' response types are listed horizontally at the top of each matrix. The ''other'' and ''none'' columns refer to responses outside of the token set and null responses, respectively.

	p	t	k	f	θ	s	ʃ	Other	None	
p	**2295**	15	23	9		1			9	
t	6	**2254**	5	24	3	1	1	5	53	T
k	39	14	**2251**	4	6	2		2	34	a
f	384	24	20	**1678**	113	65		14	54	r
θ	129	44	12	670	**1274**	197	3	7	16	g
s	10	25	1	84	87	**2106**	27	3	9	e
ʃ	1		2	2	3	38	**2302**	2	2	t

	p	t	k	f	θ	s	ʃ	Other	None	
p	**1007**	37	38	32	42	2		15	3	
t	76	**950**	38	11	72	2		20	7	T
k	32	13	**1063**	15	14	1		30	8	a
f	431	35	90	**366**	184	19	6	37	8	r
θ	249	52	104	122	**589**	23		23	14	g
s	17	29	9	41	180	**748**	81	30	41	e
ʃ	1			3	7	31	**1119**	11	4	t

	p	t	k	f	θ	s	ʃ	Other	None	
p	**1005**	4	3	73	60	7		18	6	
t	25	**1090**	7	3	24	1		16	10	T
k	27	10	**1084**	9	11	1		26	8	a
f	192	15	23	**660**	218	28	3	31	6	r
θ	53	29	18	184	**810**	54	1	22	5	g
s	11	10	5	34	90	**911**	84	23	8	e
ʃ					1	23	**1131**	19	2	t

room and listened to the stimuli over earphones. The sentences occurred in a randomized sequence at interstimulus intervals of 3.5 s.

B. Results

There were a total of 2352 observations for each consonant in each syllable position. Perceptual confusions in manner or place (but not voicing[3]) were recorded as errors, as were nonresponses. Confusion matrices are presented in Table I for all initial consonants and for final consonants in a consonantal context and final consonants in a vocalic context. Visual inspection of the matrices reveals that the overall pattern of confusions was broadly similar between the initial and the two final positions: in all three cases, the error rates were lowest for /ʃ/, somewhat higher for the stop consonants /p/, /t/, and /k/, higher still for /s/, and highest for the nonsibilant fricatives /f/ and /θ/.

The relative percentage of errors were subjected to an arcsine transform to ensure a normal distribution, and a repeated measures analysis of variance (ANOVA) was performed with four within-subjects factors (syllable position/context, target consonant, vowel nucleus, and talker). The syllable position factor consisted of three levels: initial, final before a consonant, and final before a vowel. There were significant main effects of syllable position [$F(2,26) = 38.18$, $p < 0.001$], target consonant [$F(6,78) = 52.68$, $p < 0.001$], and talker [$F(3,39) = 142.18$, $p < 0.001$], but not of vowel nucleus [$F(2,26) = 1.73$, $p > 0.1$].

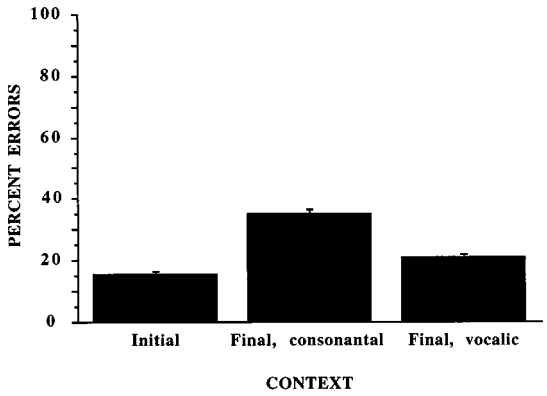


FIG. 1. Identification error rate on initial and preconsonantal and prevocalic final consonants in experiment 1.

Figure 1 displays consonant identification error rates in initial position, final position before a consonant, and final position before a vowel. As predicted, subjects made more errors identifying final consonants than initial consonants irrespective of context. Although final consonants were more identifiable when preceding a vowel, they were still less identifiable than the initial consonants. Planned comparisons revealed significant differences between initial position and final position before a consonant [$F(1,26) = 21.96$, $p < 0.001$], between initial position and final position before a vowel [$F(1,26) = 5.61$, $p < 0.05$], and between the two final position contexts [$F(1,26) = 37.21$, $p < 0.001$].

There were significant interactions between syllable position and consonant [$F(12,156) = 17.48$, $p < 0.001$] and syllable position and speaker [$F(6,78) = 5.07$, $p < 0.001$]. Mean comparisons statistically confirm what visual inspection of Fig. 2 (top panel) indicates, namely, that the perceptual advantage associated with initial position over final consonants in either context was observed for all four talkers. The position effect occurred for all consonants except /θ/ (Fig. 2, middle panel). Mean comparisons indicated, however, that apart from /θ/, two of the seven consonants (/k/ and /ʃ/) were not statistically more identifiable in initial position than in preconsonantal final position. In addition, three of the seven consonants (/t/, /k/, and /ʃ/) were as identifiable in prevocalic final position as in initial position. It is possible that failure to achieve statistical significance in these cases resulted from insufficient power (i.e., not enough errors). The interdental fricative /θ/ was the only consonant that reversed the general finding that initial consonants were more perceptible than final consonants in any context. Prevocalic final /θ/ was correctly identified significantly more often than initial /θ/ or preconsonantal final /θ/. No significant differences in identifiability occurred, however, between initial /θ/ and preconsonantal /θ/.

There was also a significant interaction between syllable position and vowel nucleus [$F(4,52) = 5.46$, $p < 0.001$] (Fig. 2, bottom panel). Although mean comparisons indicated that preconsonantal final consonants were statistically less identifiable than initial consonants across all vowel nuclei, final consonants that preceded a vowel were only statistically less

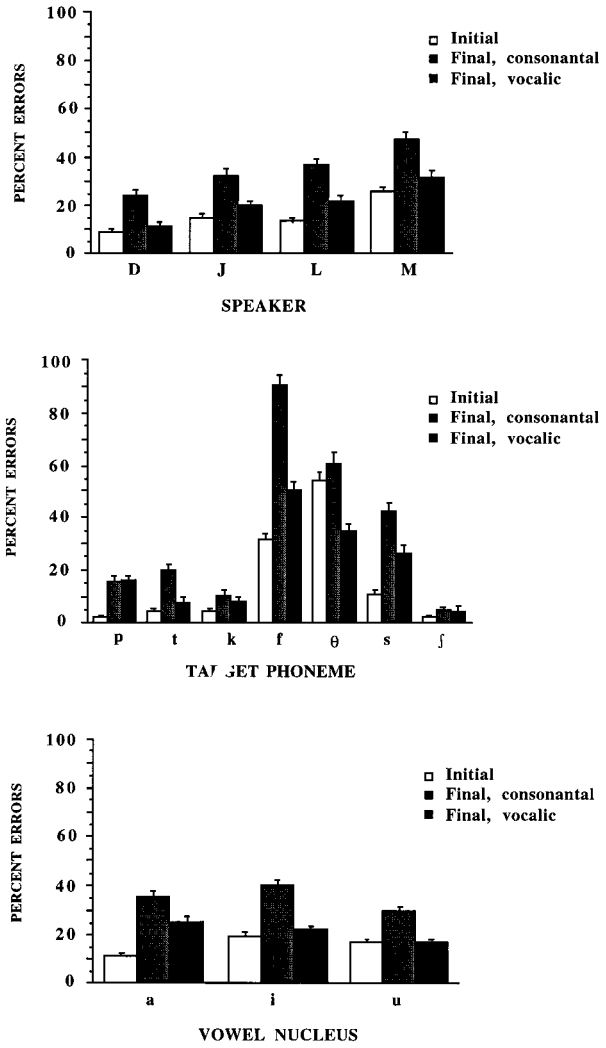


FIG. 2. Identification error rates on initial and final consonants across the different talkers (top), different phonemes (middle), and different vowel nuclei (bottom) of experiment 1.

identifiable than initial consonants when the vowel nucleus was /a/.

C. Discussion

Results of this experiment provide support for the prediction that, in naturally produced CVC syllables, initial consonants are less confusable than final consonants. The prediction that this position effect would occur across consonants, vowel nuclei, and talkers was also generally upheld. Even though the effect did not reach significance in every case, it was reversed only in the case of /θ/. The final prediction—that final consonants that occur before a consonant would be less identifiable than those that occur before a vowel—was also supported, albeit not as strongly. In the overall analysis, final consonants in a consonantal context were less identifiable than those in a vocalic context, but a significant interaction with syllable position and vowel nucleus suggested that this effect might depend on the preceding vowel context.

The fact that prevocalic final consonants that follow /i/ or /u/ nuclei were found to be as identifiable as initial consonants also points to a potential problem in this experiment: the perceptual advantage observed here for consonants in initial position could instead be a perceptual advantage for intervocalic consonants. This alternative interpretation of the main result is supported by evidence that suggests that vowel transitions both into and out of a vowel carry information about the adjacent consonant (Winitz *et al.*, 1972; Ostreicher and Sharf, 1976). This possibility is not necessarily contradicted by the fact that prevocalic final consonants are less identifiable than initial consonants when the vowel nucleus is /a/ because the pre- and postvocalic environments of the initial and final consonants are not identical. In initial position, as in final position, the prevocalic environment is always a stressed vowel—in initial position this vowel is the /e/ of ''say'' and in final position it is one of the three vowel nuclei. The postvocalic environment, however, differs in the two conditions. For initial consonants the postvocalic environment is also always a stressed vowel, but for final consonants the following vowel is always the unstressed, reduced vowel /ə/ of ''again.'' The discrepancy between the pre- and postvocalic environments of the final consonants may be especially important when the vowel nucleus is /a/ given that this vowel is generally produced with greater amplitude than other vowels. The greater discrepancy between the environments might explain why prevocalic final consonants are less perceptible than initial consonants when the vowel nucleus is /a/.

In light of the possibility that the main result from experiment 1 may reflect greater identifiability of intervocalic consonants rather than of initial consonants *per se*, we conducted another perceptual confusion experiment with the same CVC tokens embedded in different frame sentences. To test whether the main result of experiment 1 indeed reflected a perceptual advantage for initial consonants over final consonants, experiment 2 placed both initial and final consonants in a consonantal context as well as in a stressed, vocalic context.

II. PERCEPTUAL EXPERIMENT 2

A. Method

1. Stimuli

As in the first experiment, the consonants /p, t, k, f, θ, s, ʃ/ and the vowels /i, a, u/ were combined to form CVC target syllables. Each consonant occurred with every other consonant in both syllable positions and with three different vowel nuclei ($7\times7\times3=147$). Two male and two female speakers of American English produced each target syllable in two frame sentences ($147\times4\times2=1176$). Only one of the speakers was the same as in experiment 1. The frame sentences differed from experiment 1. The sentence which yielded a consonantal context for the target syllable was *I said* CVC *definitively*, which provided a similar consonantal context for both the initial and final consonants of the target syllable. The vocalic context was produced with *Say* CVC *eight times*, which ensured that the vocalic context was as similar as possible for the initial and final consonants of the target syllable. As in experiment 1, the sentences were recorded using the Kay Elemetrics Computerized Speech Laboratory with a

TABLE II. (a) Confusion matrices for postconsonantal (top) and postvocalic (bottom) initial constants. Subjects' response types are listed horizontally at the top of each matrix. The 'other' and 'none' columns refer to responses outside of the token set and null responses respectively. (b) Confusion matrices for preconsonantal (top) and prevocalic (bottom) final consonants. Subjects' response types are listed horizontally at the top of each matrix. The ''other'' and ''none'' columns refer to responses outside of the token set and null responses respectively.

(a)

	p	t	k	f	θ	s	ʃ	Other	None	
p	**1109**	11	17	19	1	1			18	
t	45	**1082**	12	5	7	4	1	1	19	T
k	25	17	**1108**	3				4	19	a
f	87	16	7	**981**	23	22		1	39	r
θ	25	44	9	435	**340**	257	10	2	54	g
s	4	59	6	45	21	**971**	47	2	21	e
ʃ	1		1	3	2	72	**1071**	7	19	t

	p	t	k	f	θ	s	ʃ	Other	None	
p	**1105**	17	11	23	1	1			18	
t	36	**1075**	35	6	1		1		22	T
k	31	15	**1090**	11	1	1	1	3	23	a
f	113	13	20	**941**	28	20		1	40	r
θ	24	41	18	456	**340**	241	7	3	46	g
s	2	17	4	68	29	**981**	49	3	23	e
ʃ	1	2	1	1	1	63	**1087**	6	14	t

(b)

	p	t	k	f	θ	s	ʃ	Other	None	
p	**709**	153	172	29	53	5		11	44	
t	117	**875**	88	8	34	4	1	4	45	T
k	152	96	**750**	26	69	6	1	11	65	a
f	174	56	199	**322**	228	10	10	12	165	r
θ	118	111	122	141	**424**	73	10	16	161	g
s	19	24	38	25	61	**794**	54	95	66	e
ʃ		5		3	11	40	**1059**	40	18	t

	p	t	k	f	θ	s	ʃ	Other	None	
p	**720**	102	139	49	71	17	1	5	72	
t	90	**830**	65	17	61	12	1	1	99	T
k	35	48	**1019**	9	9	4	1	2	49	a
f	162	43	128	**292**	281	66	13	14	177	r
θ	107	89	118	154	**428**	90	8	13	169	g
s	11	16	44	12	45	**853**	84	37	74	e
ʃ			1	1		27	**1088**	39	20	t

Shure SM48 microphone. The root-mean-squared (rms) amplitude of each stimulus sentence was normalized across talkers. The rms amplitude was calculated for the last half of the frame sentence (e.g., for *definitively* and *eight times*). The entire utterance was then scaled so that the average amplitude of the latter part of the frame sentence would be consistent across all utterances. In this way the stimuli were normalized across talkers, but the inherent amplitude differences of the different CVC target syllables were preserved even better than in experiment 1.

2. Subjects and procedure

Subjects were seven female and seven male native American-English-speaking college students, none of whom had served in the previous experiment. The instructions and procedure were identical to those of experiment 1.

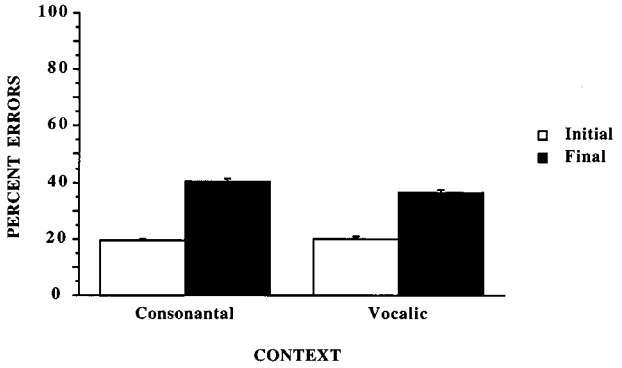


FIG. 3. Identification error rates on initial and preconsonantal and prevocalic final consonants in experiment 2.

B. Results

There were a total of 2352 observations for each consonant in each syllable position. Perceptual confusions of manner and place (but not voicing) were recorded as errors, as were nonresponses. Confusion matrices for initial and final consonants in the two conditions are presented in Table II. As in experiment 1, the pattern of confusion errors was qualitatively similar across conditions: the fricative /ʃ/ was confused least often of any phoneme, and the stop consonants and /s/ were confused less often than the fricative /f/ and /θ/.

A repeated measures analysis of variance (ANOVA) was performed on error rate with five within-subjects factors (syllable position, context, target consonant, vowel nucleus, and talker). There were significant main effects for syllable position [$F(1,13)=105.01$, $p<0.001$] and context [$F(1,13)=12.41$, $p<0.005$], target consonant [$F(6,78)=74.8$, $p<0.001$], vowel nucleus [$F(2,26)=13.75$, $p<0.001$], and talker [$F(3,39)=174.37$, $p<0.001$].

There was a significant interaction between syllable position and context [$F(1,13)=35.5$, $p<0.001$] (see Fig. 3). Planned comparisons reveal that subjects made more errors identifying final consonants than initial consonants irrespective of context ([$F(1,13)=1534.67$, $p<0.001$] in a consonantal context and [$F(1,13)=945.5$, $p<0.001$] in a vocalic context), and on preconsonantal final consonants than on prevocalic final consonants [$F(1,13)=55.25$, $p<0.001$], but that postconsonantal initial consonants were no less identifiable than postvocalic initial consonants [$F(1,13)=0.99$, $p>0.1$]. In spite of the statistically significant difference between the identifiability of prevocalic and preconsonantal final consonants, the graph shows clearly that this difference is small. Thus, further analyses deal only with the difference between initial and final positions.

Although significant interactions occurred between syllable position and target consonant [$F(6,78)=27.66$, $p<0.001$], syllable position and vowel nucleus [$F(2,26)=11.7$, $p<0.001$], and syllable position and talker [$F(3,39)=16.86$, $p<0.001$], the perceptual advantage associated with initial position was observed for all four talkers (Fig. 4, top panel), every consonant except /θ/ (Fig. 4, middle panel), and all three vowel nuclei (Fig. 4, bottom panel). Mean comparisons indicated that these differences

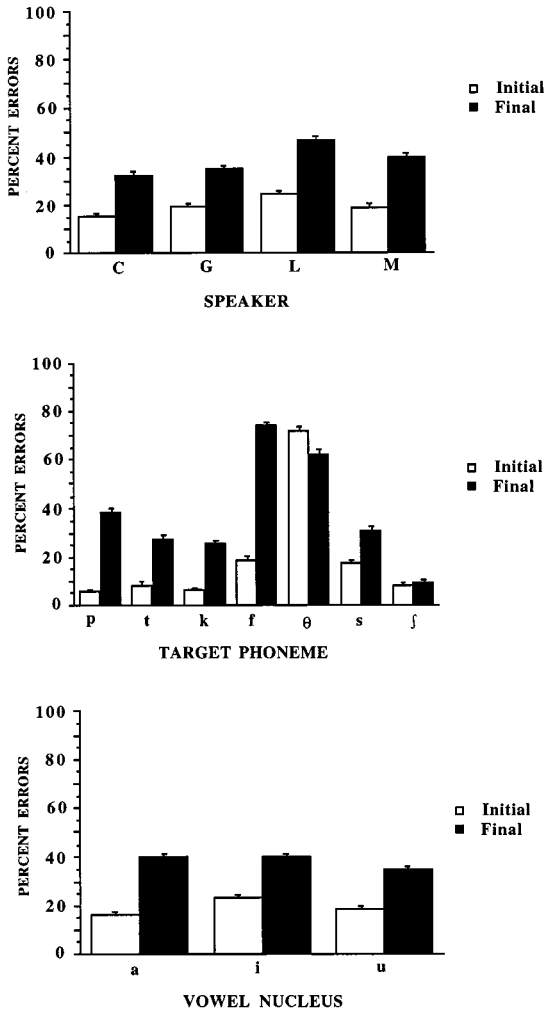


FIG. 4. Identification error rates on initial and final consonants across the different talkers (top), different phonemes (middle), and different vowel nuclei (bottom) of experiment 2.

were statistically significant ($p<0.05$) in all conditions for all cases except /ʃ/, which had the fewest errors of any of the consonants in either position.

C. Discussion

1. Predictions

Results from the second perceptual confusion experiment confirm that the perceptual advantage associated with initial consonants over final consonants is attributable to position rather than to context. The findings from the two experiments are consistent with demonstrations of a perceptual advantage for syllable onsets over syllable offsets (e.g., Raphael and Dorman, 1980) and with results of some earlier perceptual confusion studies (Miller and Nicely, 1955; Helfer and Huntley, 1991). The results do, however, deviate in one respect from those of Wang and Bilger (1973), who reported a perceptual *dis*advantage for initial consonants when the vowel nucleus was /i/. The reasons for this discrepancy between the two studies are unclear. In any case, the observed positional effect on consonant identifiability is consistent with the hypothesis that cross-language preferences for initial over final consonants emerge through selection for communicative value.

Another prediction was that the position effect would occur across consonants, vowel nuclei, and talkers. This was also confirmed. Although in both experiments there were significant interactions between position and each of the other variables, the greater identifiability of initial consonants occurred for six of the seven consonants, all three vowel nuclei, and all four talkers, and thus appears to be quite general. Moreover, the overall pattern of confusions was similar across positions, with sibilant fricatives and stop consonants being more accurately identified than nonsibilant fricatives in both cases.

It is not clear why the position effect was reversed in the case of /θ/. In any case, on the basis of this result, it might be expected that /θ/, already a rare phoneme in the world's languages, will occur preferentially in inventories, like English, that allow for final consonants.

The prediction that syllable-final consonants would be less confusable before a vowel than before another consonant was strongly upheld in experiment 1, but only weakly in experiment 2. This discrepancy between experiments may stem from differences between the vowels that followed the final consonants. In Experiment 1, the final consonant preceded an unstressed vowel. Phonologists have noted that consonants in this context are often treated as ambisyllabic (Kahn, 1976) or as geminates (Burzio, 1994), making the consonant syllable-initial as well as syllable-final. In experiment 2, the final consonant preceded a stressed vowel. Hoard (1966) noted that glottal stops are often inserted before stressed, word-initial vowels. Visual inspection of our stimulus waveforms confirmed that glottal stops were typically inserted in this context. The presence of a glottal stop would preclude resyllabification of the final consonant across the word boundary. Thus, whereas the prevocalic final consonants in experiment 1 may have been treated (also) as syllable-initial consonants, those of experiment 2 were likely to be treated only as syllable-final ones.

2. Explanations

There are at least two types of explanation for why initial consonants are more easily identified than final consonants. One explanation attributes the position effect to differences in auditory processing and assumes that similar acoustic information is available for both initial and final consonants. A candidate auditory processing account of this type comes from a study by Delgutte and Kiang (1984), who presented isolated synthetic CV and VC stimuli to anesthetized cats and measured discharge patterns of auditory-nerve fibers. They found that the discharge rate was lower for consonants in syllable-final position because of short-term adaptation to the preceding vowel. On the basis of this evidence, it might be argued that syllable onsets are more readily perceived than syllable offsets because onsets yield a greater neuronal response. However, such an account does not appear to be applicable to consonant perception in phrase contexts, where syllable-initial consonants are also susceptible to the effects of short-term adaptation from preceding vowels

or consonants. Thus, the account does not readily explain either the present findings or the cross-language position preferences in syllable structure that motivated this study.

Another possible explanation for the position effect on identification accuracy is that initial consonants may be produced differently than final consonants, resulting in different acoustic correlates. For example, there is some evidence that consonants tend to be longer in syllable-initial than in syllable-final position, provided the syllable is not in phrase-final position (Hoard, 1966; Anderson and Port, 1992). In the case of fricatives, a shorter duration might yield less overall information for segment identity. For stops, a short duration could negatively affect identification in the context of other consonants. Raphael and Dorman (1980) found that shorter silent intervals (corresponding to stop closure) reduced perceptibility of the first stop in a sequence of two consonants straddling a syllable boundary. Also, segments that are produced more quickly tend to be more reduced and more variable (Moon and Lindblom, 1994; Byrd and Tan, 1996), especially in final position (Byrd, 1996).

Another potentially relevant factor in the perception of consonants is amplitude. Sounds that are harder to hear are harder to identify. It is likely that the repeatedly observed difference in identifiability between the sibilant fricatives, /s/ and /ʃ/, and the nonsibilant fricatives, /f/ and /θ/, is largely attributable to their differences in relative amplitude. With respect to stop consonants, Henderson and Repp (1982) noted that although preconsonantal stops usually have release bursts, these are of much lower amplitude than prevocalic stops and therefore difficult to detect. Given that bursts are important (and sometime sufficient) cues for the identification of voiceless stops (Winitz *et al.*, 1972), differences in perceptibility may result if syllable position affects their amplitude (Malecot, 1968).

Winitz *et al.* (1972) also found that the average frequency of the stop release burst provided good place-of-articulation information. Place information is also provided by the average frequency value of the noise in fricative consonants (Harris, 1958). Given that average frequency plays a role in the identification of consonants, positional variation in frequency values may affect the relative perceptibility of the consonants to be identified.

Another perceptual cue for consonants is the vowel formant transitions into and out of the consonant. These transitions reflect the coarticulation of the consonant with the following or preceding vowel and provide information for consonant place of articulation (Delattre *et al.*, 1955; Harris, 1958) and manner (Liberman *et al.*, 1956). In a review of the acoustic and perceptual literature on stop consonants, Pickett *et al.* (1995) conclude that CV transitions are generally more informative about stop consonant identity than VC transitions. Results of studies on ''locus equations'' are consistent with this conclusion. Locus equations are regression functions describing the relationship between the onset (or offset) frequency of the second formant ($F2$) in a CV (or VC) transition and the $F2$ value of the adjacent vowel nucleus. These variables have been shown to be linearly related for a given place category across different vowels, and different place categories are associated with different locus equation slopes (Lindblom, 1963). Krull (1988) and Sussman *et al.* (1997) found that these slopes differ less across place of articulation categories for VC transitions than for CV transition, indicating that $F2$ trajectories are more distinctive for place value in CV than in VC syllables. In addition, Sussman *et al.* (1997) found that the variance explained by locus equations is significantly greater for stops in initial than in final position. Finally, it is important to note that locus equation parameters have been shown to be relevant for listeners' identification of place categories (Fruchter and Sussman, 1997, but see commentaries associated with Sussman *et al.*, 1998).

In view of the preceding discussion, five different types of acoustic measurements were made on the CVC stimuli used in experiment 2 to determine whether differences in production could account for the results obtained. A discriminant analysis was then performed to determine how accurately consonants could be categorized on the basis of these measurements.

III. ACOUSTIC MEASUREMENTS

A. Method

Five acoustic measurements—one of duration, one of amplitude, an average frequency measure, and two frequency measures related to $F2$ transitions—were made on each of the initial and final consonants of the CVC stimuli produced by the four talkers in experiment 2. Only the stimuli from experiment 2 were analyzed because this experiment provided the stronger test of the relative identifiability of initial and final consonants. Fricative duration was measured from the onset to the offset of friction noise. In the case of stops, the durational measurement included the release burst and aspiration interval as well as the closure interval. (In cases where unreleased final stops preceded another stop consonant, duration of the target consonant could not be determined.) Root-mean-squared (rms) amplitude was calculated for energy present during the entire constriction interval for fricatives and for the energy present in a 51.2-ms window centered at the onset of the release burst for stops. A Hamming window was applied to these same intervals, and the frequency centroid [$=\Sigma(x_i y_i)/\Sigma y_i$, where x_i = frequency in Hz of ith sample and y_i = corresponding amplitude], or average frequency, was calculated. The $F2$ transition measures were taken at the onset of $F2$ for initial consonants (or the offset of $F2$ for final consonants) and at the midpoint of the following (or preceding) vowel. These last two measures correspond to the data points used in studies of locus equations. Following Sussman and Shore (1996), the $F2$ onset (or offset) values were measured at the point where $F2$ was first (or last) clearly distinguishable in a visual inspection of the spectrogram.

B. Results

Table III presents the average values of the five acoustic measurements for the initial and final consonants of the CVC stimuli from experiment 2. The values presented are averaged across 84 observations for each of the fricatives in each condition and for each of the stops in the two initial conditions. The values presented for final stops are averaged

TABLE III. Acoustic measurements of the initial and final consonants of the CVC stimuli from experiment 2. Average values are displayed for each of the phonemes in the two contexts: consonantal (top) and vocalic (bottom). Duration measures are in milliseconds, frequency measures are in Hz, and amplitude measures are rms values calculated from 16-bit integer signals.

	Duration		Amplitude		Average frequency		$F2$ onset/offset		$F2$ midpoint	
	Initial	Final	Initial	Final	Initial	Final	Initial	Final	Initial	Final
p	153	103	599	270	1774	1776	1664	1661	1807	1859
t	161	111	808	542	2664	2381	2181	1962	1953	1902
k	152	103	634	315	2304	2012	2161	2012	1847	1876
f	136	105	286	381	2480	2358	1673	1731	1781	1824
θ	135	106	268	270	2265	2058	1933	1896	1913	1904
s	155	115	666	821	3356	3365	1958	1919	2155	1897
ʃ	172	121	1996	1813	3387	3253	2112	2069	1898	1868
p	151	122	575	314	1764	1545	1721	1644	1801	1873
t	158	109	706	467	2653	2422	2254	2032	1952	1906
k	159	122	614	550	2327	2432	2196	2116	1853	1874
f	153	133	280	326	2407	2324	1675	1719	1795	1840
θ	146	136	212	245	2187	1989	2024	1902	1904	1884
s	175	140	630	633	3342	3330	2074	1956	1918	1885
ʃ	178	140	2042	1771	3362	3294	2182	2103	1910	1873

across the released final stops only and are based on 63, 33, and 52 observations for preconsonantal /p/, /t/, and /k/, respectively, and 83, 73, and 84 observations for prevocalic /p/, /t/, and /k/, respectively.

Repeated measures analyses of variance (ANOVA) were performed on each of the measurements on initial and final consonants (except for $F2$ of the vowel nucleus) with five factors (syllable position, context, target consonant, vowel nucleus, and talker). The analyses indicated that initial consonants were greater both in duration [$F(1,6)=937.82$, $p<0.001$] and amplitude [$F(1,6)=101.3$, $p<0.001$] than final consonants. Average frequency and frequency of $F2$ onset/offset also differed between initial and final consonants [$F(1,6)=349.2$, $p<0.001$; $F(1,6)=9.4$, $p<0.05$].

A significant interaction between syllable position and target consonant occurred for all measurements. Significant interactions between syllable position and context occurred for duration, amplitude, and average frequency. Significant interactions also occurred for duration and amplitude between syllable position and talker and for average frequency between syllable position and vowel nucleus. Mean comparisons indicated that (1) consonant duration was significantly greater in initial position than in final position across all conditions; (2) consonant amplitude was significantly greater in initial position than in final position for both contexts, all vowel nuclei, and for /p/, /t/, /k/, and /ʃ/, but not for /f/, /θ/, or /s/; (3) average frequency and the onset/offset frequency of $F2$ transition were significantly higher in initial position than final position, but varied considerably across the different conditions. Even though there was no effect for intensity for three of the seven consonants, there were no significant reversals of the tendency for initial consonants to be longer and more intense than final consonants.

Pearson correlations on 168 token-measure pairs were made for each of the consonants in the initial and final conditions. Table IV lists coefficients for correlations between

TABLE IV. Correlation coefficients between number of errors and each of the measures for each of the initial and final consonants.

	Duration		Amplitude		Average frequency		$F2$ onset/offset	
	Initial	Final	Initial	Final	Initial	Final	Initial	Final
p	0.133	0.251[b]	−0.185[a]	−0.218[b]	0.091	−0.060	0.039	0.220[b]
t	−0.050	−0.192[a]	−0.255[b]	−0.244[b]	−0.180[a]	−0.105	0.250[b]	0.219[b]
k	0.138	−0.045	−0.362[b]	−0.484[b]	0.036	−0.346[b]	0.049	−0.044
f	0.175[a]	−0.068	−0.084	−0.317[b]	−0.055	−0.203[b]	−0.126	−0.178[a]
θ	0.084	−0.043	−0.063	−0.237[b]	−0.149	−0.098	0.252[b]	−0.060
s	−0.133	−0.039	−0.439[b]	−0.508[b]	−0.543	−0.618[b]	0.342[b]	0.102
ʃ	−0.103	0.095	−0.172[a]	−0.079	−0.132	−0.003	−0.111	0.071

[a] $p<0.05$.
[b] $p<0.01$.

each of the acoustic measures (excluding $F2$ of the vowel nucleus) and listeners' identification error rate for each consonant. A quite consistent inverse relationship was observed between consonant amplitude and error. However, shorter consonant durations were not associated with more errors.

A discriminant analysis was performed to assess the degree to which the tokens could be correctly classified on the basis of the five acoustic measures. Classification accuracy is presented in Fig. 5. Overall correct classification for initial consonants was 66.0% in the consonantal context and 67.2% in the vocalic context. Correct classification for final consonants was 54.3% in the consonantal context and 58.8% in the vocalic context. The performance patterns are similar to those of the human subjects, but the human subjects outperform discriminant analysis in most cases (see Fig. 4 for comparison with human subjects). Only for /f/ and /θ/ did discriminant analysis outperform human subjects.

Table V shows the standardized canonical discrimination coefficients of the first two discriminant functions. These coefficients reflect the degree to which a given acoustic variable contributes successfully to the classification of stimulus tokens. Across consonant class and position, the most highly weighted coefficients corresponded to the frequency measures, especially $F2$ transition onset or offset and $F2$ of the preceding or following vowel. The average frequency measure and amplitude were somewhat less important than the $F2$ transition measures in discriminating phoneme categories, but more important than duration.

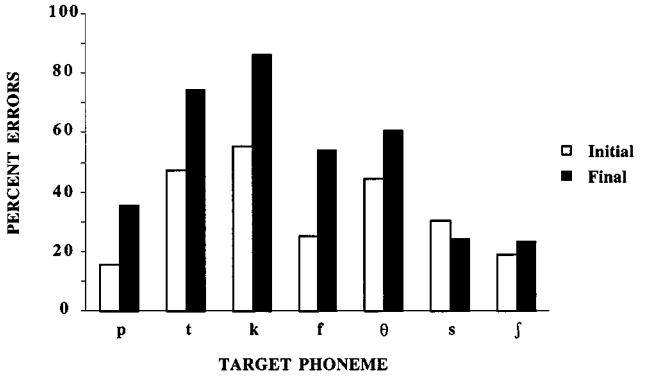


FIG. 5. Correct classification by disciminant analysis of each phoneme in initial and final position.

TABLE V. Standardized canonical discriminant coefficients for the five discriminant functions.

	Duration		Amplitude		Frequency		$F2$ onset/offset		$F2$ midpoint	
	Initial	Final	Initial	Final	Initial	Final	Initial	Final	Initial	Final
1	0.085	0.071	0.643	0.483	0.592	0.765	0.439	0.091	−0.438	−0.149
2	−0.017	0.029	0.658	0.898	−0.866	−0.701	0.685	0.116	−0.329	0.028
3	0.077	−0.611	0.419	−0.088	−0.034	0.037	−1.235	0.812	0.804	0.118
4	1.016	0.785	−0.187	−0.121	−0.096	−0.062	−0.041	0.583	−0.156	−0.221
5	0.046	0.153	0.01	0.025	−0.032	−0.015	−0.099	−0.174	1.067	0.995

C. Discussion

Analyses of the acoustic measurements made on the CVC stimuli suggest that the perceptual advantage associated with initial consonants may have a basis in production. Initial consonants were found to be longer and louder than their final counterparts. Similar position differences in duration and amplitude have been previously reported in the literature (Hoard, 1996; Anderson and Port, 1992; Malecot, 1968; Henderson and Repp, 1982). The link between amplitude and perceptibility is obvious and was supported by the tendency for number of errors to be significantly inversely correlated with amplitude. The link between frequency differences and identifiability might reflect the relative role that frequency plays in differentiating consonant types. For example, the fact that higher frequencies correlated positively with identification accuracy for /s/ indicates that subjects probably used frequency as a cue to distinguish /s/ from the other fricatives. The expected inverse correlation between duration and number of errors was not found.

A discriminant analysis based on acoustic measurements of duration, amplitude, average frequency, and $F2$ transition yielded more accurate classification of initial consonants than final consonants. In this respect, discriminant classification performance was similar to human performance in the identification experiment. The fact that initial consonants were better classified suggests they are produced with greater acoustic distinctiveness than final consonants.

Despite the parallels, humans generally outperformed discriminant analysis except in the classification of nonsibilant fricatives. The advantage for human listeners is not surprising since they have access to a considerably richer array of acoustic cues than are available to a discriminant classifier. For instance, one important cue not available to the classifier was the presence or absence of a silent interval, normally used by human listeners to distinguish stops from fricatives. Also, because the discriminant analysis only included released stops, the relative contribution of a release burst *per se* is unknown.

IV. GENERAL DISCUSSION

The purpose of this study was to evaluate whether cross-language preferences for syllable-initial consonants may be influenced by perceptual and/or production factors. Consonant identification experiments were conducted to test three predictions: (1) that syllable-initial consonants are identified more accurately than syllable-final consonants; (2) that this perceptual advantage occurs across consonants, vowel nuclei, and talkers; and (3) that syllable-final consonants preceding a vowel are identified more accurately than those preceding a consonant. The first prediction was clearly supported. The second prediction was supported in that the position effect on identification accuracy was observed for all four talkers, all three vowel nuclei, and for six of the seven consonants, albeit not always at statistically significant levels. In addition, the overall pattern of identification errors was similar across initial and final positions in both experiments: sibilant fricatives and stops were more accurately identified than nonsibilant fricatives.

The third prediction—that syllable-final consonants are more identifiable before a vowel than before a consonant—received strong support in experiment 1, but not in experiment 2. As discussed earlier, it is possible that the discrepancy between the experiments arises from a difference in vocalic environment in the two cases. Specifically, in experiment 1 the target final consonant occurred before a schwa—an environment conductive to ambisyllabicity (or gemination)—making the consonant syllable-initial as well as syllable-final. However, in experiment 2, the target final consonant was followed by a stressed word-initial vowel, an environment that typically prompts the insertion of a glottal stop and thus precludes resyllabification of the final consonant.

Acoustic properties of the stimuli were analyzed to determine whether or not production differences could account for the observed differences in identification accuracy. A significant positive correlation between amplitude and accuracy, but not between duration and accuracy, indicated that amplitude contributed more than duration to differences in perceptibility of initial and final consonants. The discriminant analysis indicated that frequency measures, especially those associated with the $F2$ transition, also contributed strongly to the separation of the consonant categories. Importantly, discriminant classification performance was greater for initial than for final consonants, suggesting that the perceptual advantage for initial consonants may be attributable to their greater acoustic distinctiveness.

The present study was motivated by the view that factors of perception and production shape cross-language preferences in sound structure. The result that initial consonants are more identifiable than final consonants is consistent with the typological observation that CV syllables are preferred to syllables with final consonants. The fact that this result held in consonantal as well as vocalic contexts is consistent with the cross-language observation that consonant sequences,

though disfavored overall, are more likely to occur in initial position than in final position.

The perceptual advantage associated with initial consonants held across consonant and adjacent vowel categories, indicating that the syllable position effect is not attributable to specific target phonemes and contexts. However, the general nature of the effect does not preclude the possibility that different consonants are more or less suited for final position. This possibility is consistent with the typological observation that consonants permitted in final position represent a small subset of a language's total consonant inventory.

It is, of course, possible that the results of the present study are limited to the language environment in which they were observed. On the other hand, it might be expected that because English syllable structure incorporates many different syllable types with final consonants, English speakers and listeners would be especially adept at the production and perception of final consonants. Nevertheless, the broad cross-language arguments made here would be considerably strengthened by replicating these findings with speakers of other languages.

Although our results indicate a perceptual advantage for syllable-initial consonants, it is possible that the advantage is restricted to word-initial consonants. Further research involving multi-syllabic words is necessary to clarify this issue.

In summary, cross-language preferences for syllable-initial consonants may be attributable to their relatively high identifiability. This perceptual advantage may derive in part from production factors giving rise to greater acoustic distinctiveness.

ACKNOWLEDGMENTS

This paper benefited from the insightful criticisms and comments of three anonymous reviewers. The work was supported by a NSF Graduate Research Fellowship to the first author and by Research Grant No. 5 R01 DC00427-09, -10 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health, to the second author.

[1]Unavoidably, some of the combinations corresponded to actual English words. Since lexical status of the syllables is independent of the syllable position and context variables, there is no reason to expect any biasing effect of lexical status with respect to the hypotheses under test.

[2]Pink noise was used because it has a spectrum that drops off in amplitude at the same rate as speech. The flat spectrum of white noise would mask the higher more than the lower frequencies of speech.

[3]Voicing errors were not recorded because the stimuli did not vary along this dimension and subjects were told in advance which consonants and vowels they would hear in the experiment.

Ades, A. E. (**1974**). ''How phonetic is selective adaptation? Experiments on syllable position and vowel environment,'' Percept. Psychophys. **16**, 61–66.

Anderson, S., and Port, R. (**1992**). ''Evidence for syllable structure, stress and juncture from segmental durations,'' J. Phonetics **22**, 283–315.

Bell, A., and Hooper, J. B. (**1978**). ''Issues and evidence in syllabic phonology,'' in *Segments and Syllables*, edited by A. Bell and J. B. Hooper (North–Holland, New York), pp. 9–22.

Burzio, L. (**1994**). *Principles of English Stress* (Cambridge U. P., Cambridge).

Byrd, D. (**1996**). ''Influences on articulatory timing in consonant sequences,'' J. Phonetics **24**, 209–244.

Byrd, D., and Tan, C. C. (**1996**). ''Saying consonant clusters quickly,'' J. Phonetics **24**, 263–282.

Davis, B. L., and MacNeilage, P. F. (**1995**). ''The articulatory basis of babbling,'' J. Speech Hear. Res. **38**, 1199–1211.

Delattre, P. C., Liberman, A. M., and Cooper, F. S. (**1995**). ''Acoustic loci and transitional cues for consonants,'' J. Acoust. Soc. Am. **27**, 769–773.

Delgutte, B., and Kiang, N. Y. S. (**1984**). ''Speech coding in the auditory nerve: IV. Sounds with consonantlike dynamic characteristics,'' J. Acoust. Soc. Am. **75**, 897–907.

Diehl, R. L., and Kluender, K. (**1989**). ''On the objects of speech perception,'' Ecological Psychol. **1**, 121–144.

Diehl, R. L., and Lindblom, B. (**1996**). ''Explaining the structure of feature and phoneme inventories,'' in *Speech Processing in the Auditory System*, edited by S. Greenberg, W. A. Aisnworth, A. Popper, and R. Fay (Springer-Verlag, New York).

Fruchter, D., and Sussman, H. (**1997**). ''Perceptual significance of locus equations,'' J. Acoust. Soc. Am. **105**, 2997–3008.

Greenberg, S. (**1998**). ''A syllable-centric framework for the evolution of spoken language,'' Brain Behav. Sci. **21**, 518.

Harris, K. S. (**1958**). ''Cues for the discrimination of American English fricatives in spoken syllables,'' Lang. Speech **1**, 1–7.

Helfer, K. S., and Huntley, R. A. (**1991**). ''Aging and consonant errors in reverberation and noise,'' J. Acoust. Soc. Am. **90**, 1786–1795.

Henderson, J. B., and Repp, B. (**1982**). ''Is a stop consonant released when followed by another stop consonant?'' Phonetica **39**, 71–82.

Hoard, J. E. (**1966**). ''Juncture and syllable structure in English,'' Phonetica **15**, 96–109.

Hock, H. H. (**1986**). *Principles of Historical Linguistics* (Mouton de Greater, The Hague).

Kahn, D. (**1976**). ''Syllable-based generalizations in English phonology,'' Ph.D. thesis, MIT.

Krull, D. (**1988**). ''Acoustic properties as predictors of perceptual responses: A study of Swedish voiced stops,'' Phonetic Experimental Research at the Institute of Linguistics, University of Stockholm, PERILUS VII, pp. 66–70.

Liberman, A. M., Delattre, P. C., Gerstman, L. J., and Cooper, F. S. (**1956**). ''Tempo of frequency change as a cue for distinguishing classes of speech sounds,'' J. Exp. Psychol. **52**, 127–137.

Lindblom, B. (**1963**). ''On vowel reduction,'' Report No. 29, The Royal Institute of Technology, Speech Transmission Laboratory, Stockholm, Sweden.

Lindblom, B. (**1986**). ''Phonetic universals in vowel systems,'' in *Experimental Phonology*, edited by J. Ohala and J. Jaeger (Academic, New York), pp. 13–44.

Lindblom, B., MacNeilage, P. F., and Studdert-Kennedy, M. (**1984**). ''Self-organizing processes and the explanation of phonological universals,'' in *Explanations for Language Universals*, edited by B. Butterworth, B. Comrie, and O. Dahl (Mouton, Berlin), pp. 181–203.

Maddieson, I. M. (**1984**). *Patterns of Sounds* (Cambridge U. P., London).

Malecot, A. (**1968**). ''The force of articulation of American stops and fricatives as a function of position,'' Phonetica **18**, 95–102.

Miller, G. A., and Nicely, P. E. (**1955**). ''An analysis of perceptual confusions among some English consonants,'' J. Acoust. Soc. Am. **27**, 338–352.

Moon, S., and Lindblom, B. (**1994**). ''Interaction between duration, context, and speaking style in English stressed vowels,'' J. Acoust. Soc. Am. **96**, 40–55.

Ostreicher, H. J., and Sharf, D. J. (**1976**). ''Effects of coarticulation on the identification of deleted consonant and vowel sounds,'' J. Phonetics **4**, 285–301.

Pickett, J. M., Bunnell, H. T., and Revoile, S. G. (**1995**). ''Phonetics of intervocalic consonant perception: retrospect and prospect,'' Phonetica **52**, 1–40.

Raphael, L. J., and Dorman, M. F. (**1980**). ''Silence as a cue to the perception of syllable-initial and syllable-final stop consonants,'' J. Phonetics **8**, 269–275.

Samuel, A. G. (**1989**). ''Insights from a failure of selective adaptation: Syllable-initial and syllable-final consonants are different,'' Percept. Psychophys. **45**, 485–493.

Sussman, H. M., Bessell, N., Dalston, E., and Majors, T. (**1997**). ''An investigation of stop place articulation as a function of syllable position: a locus equation perspective,'' J. Acoust. Soc. Am. **101**, 2826–2838.

Sussman, H. M., Fruchter, D., Hilbert, J., and Sirosh, J. (**1998**). ''Linear correlates in the speech signal: The orderly output constraint,'' Behav. Brain Sci. **21**, 241–299.
Sussman, H. M., and Shore, J. (**1996**). ''Locus equations as phonetic descriptors of consonantal place of articulation,'' Percept. Psychophys. **58**, 936–946.
Van Wieringen, A. (**1995**). ''Perceiving dynamic speechlike sounds: psychoacoustics and speech perception,'' Ph.D. thesis, Universiteit van Amersterdam, Holland.
Wang, M. D., and Bilger, R. C. (**1973**). ''Consonant confusions in noise: a study of perceptual features,'' J. Acoust. Soc. Am. **54**, 1248–1266.
Willerman, R. (**1994**). ''The phonetics of pronouns: articulatory bases of markedness,'' Ph.D. thesis, University of Texas, Austin.
Winitz, H., Scheib, M. E., and Reeds, J. A. (**1972**). ''Identification of stops and vowels for the burst portion of /p,t,k/ isolated from conversational speech,'' J. Acoust. Soc. Am. **51**, 1309–1317.
Wolf, C. (**1978**). ''Perceptual invariance for stop consonants in different positions,'' Percept. Psychophys. **24**, 315–326.

Preference judgments of artificial processed and hearing-aid transduced speech

Niek J. Versfeld, Joost M. Festen, and Tammo Houtgast
Department of Otolaryngology, University Hospital VU, P.O. Box 7057, 1007 MB Amsterdam, The Netherlands

(Received 27 April 1998; revised 10 April 1999; accepted 3 May 1999)

In order to assess the relative importance of various signal processing algorithms and distortions on hearing-aid preference, male and female speech was manipulated in a number of ways and subsequently presented to normal-hearing and hearing-impaired subjects (the latter having a mild sensorineural high-frequency hearing loss). Signal manipulations were artificial (e.g., filtering, compression, peak clipping, or adding noise) or were actual dummy-head recordings of five different hearing aids. Listeners judged the sounds in a pairwise-comparison format. Their task was to indicate the ''hearing aid'' they would prefer assuming they had to wear it all day. The data were analyzed with multidimensional scaling techniques; Principal Components Analysis revealed that the first two dimensions on which preference judgments were based, can be interpreted as (1) intelligibility or clarity, and (2) distinction between signal distortion and added background distortion. Furthermore, the results showed that normal-hearing subjects generally preferred the original signal, whereas hearing-impaired subjects were inclined to choose the signals with a high-frequency emphasis. Severe band-pass filtering or low-frequency emphasis were disliked, as was to be expected. Surprisingly, however, a soft background noise (S/N ratio of 25 dB) was often among the least preferred of all signals. The differences in preference between the five hearing aids were small, but consistent. For hearing-impaired subjects, hearing-aid ordering could be accounted for by the amount of low-frequency cutoff; for normal-hearing subjects both high- and low-frequency cutoff played a role. Results of a retest experiment with normal-hearing subjects, about one year later, showed that subjects' criteria remain remarkably stable. © *1999 Acoustical Society of America.* [S0001-4966(99)04409-4]

PACS numbers: 43.71.Gv, 43.71.Ky, 43.38.Xm, 43.66.Ts [WS]

INTRODUCTION

The sound quality of reproducing systems (e.g., loudspeakers) or transducers (e.g., hearing aids) has been the subject of research in several papers (e.g., Eisler, 1966; McDermott, 1969; Staffeldt, 1974; Gabrielsson and Sjögren, 1979a; Punch *et al.*, 1980). With respect to hearing aids, it is recognized that, next to speech intelligibility, sound quality plays an important role in user satisfaction (Kochkin, 1992). Nonetheless, the topic of sound quality has received only limited attention, and, up to the present, hearing-aid design, selection, and adjustment still aims at optimization of speech intelligibility. However, current speech intelligibility tests cannot differentiate well between different hearing-aid settings. The audiologist is thus faced with the problem of how to select the best setting. Usually, remaining degrees of freedom are utilized to optimize the hearing aid with respect to sound quality. However, these adjustments are largely based on the experience of the audiologist, rather than on objective or standardized procedures. With the great increase in the number of adjustment parameters available with modern hearing aids, the demand for such procedures increases. Unfortunately, no such procedures exist, not only due to the complex and multidimensional nature of sound quality, but probably also due to large individual differences in listener preference. For the development and application of such procedures, the nature of sound quality and the degree to which individuals differ from one another needs to be studied in more detail.

Gabrielsson and co-workers (1979a, b, 1988) have tried to unravel the underlying dimensions that constitute the sound quality of hearing aids by the use of verbal descriptions. They found that a limited number of dimensions could account for their results. The most important dimensions were labeled as ''sharpness,'' ''clarity,'' ''brightness,'' ''fullness,'' ''spaciousness,'' ''nearness,'' ''loudness,'' and ''disturbing sounds.'' Although no real attempt was made to relate these dimensions to actual physical characteristics of the hearing aids, they noted that, for example, ''sharpness'' and ''fullness'' were related to high- and low-frequency boost, respectively, whereas the ''disturbing sounds'' probably originated from nonlinear distortions (as peak clipping) and the internal noise of the hearing aid.

In another study to investigate the nature of the underlying dimensions of sound quality of speech transmitted through a variety of circuits, McDermott (1969) conducted pairwise-comparison experiments, where listeners had to give either similarity ratings or preference judgments. Multidimensional analyses revealed three factors that were interpreted as ''overall clarity,'' ''distinction between signal distortion and background distortion,'' and ''loudness.'' Other experiments, similar to that of McDermott (1969), were conducted by Yonovitz *et al.* (1978) and Punch *et al.* (1980) to

identify the electroacoustic characteristics that affect the sound quality of hearing aids. Yonovitz *et al.* (1978) were able to identify and label five dimensions (viz., low- and high-frequency cutoff, bandwidth, third-harmonic distortion, and internal noise), whereas Punch *et al.* (1980) were able only to label the first dimension as the low-frequency cutoff, despite the fact that more than one significant dimension could be extracted. Detailed analysis of the data showed that all listeners took into account low-frequency cutoff, but beyond that individuals used different electroacoustic variables.

Although several papers on sound quality have mentioned individual differences (Eisler, 1966; Punch *et al.*, 1980; Byrne and Walker, 1982; Studebaker and Sherbecoe, 1988; Sullivan *et al.*, 1988; Kuk *et al.*, 1990; Neuman *et al.*, 1995; Preminger and van Tassel, 1995a, 1995b), actually few view this as a consequence of the multidimensional structure of sound quality. Preminger and van Tassel (1995a) have shown convincingly that intelligibility plays a predominant role in speech quality judgments. They asked listeners to judge, among others, the intelligibility, pleasantness, loudness, effort, and overall impression of a set of stimuli. If the stimuli differed with respect to intelligibility, the ratings on all subjective dimensions correlated highly, and no large inter-listener differences were found. If, on the other hand, the stimuli were equated with respect to intelligibility, similarities between the different subjective dimensions were much less evident, and substantial differences between listeners were observed. These findings may explain the lack of hearing-aid–listener interaction in experiments with hearing aids that differ with respect to perceived intelligibility (e.g., Jeffers, 1962; Studebaker and Sherbecoe, 1988; Sullivan *et al.*, 1988), and indicate that in studying the factors underlying the sound quality of hearing aids one should keep the perceived intelligibility as constant as possible.

The experiments reported in this paper deal with preference judgments of normal-hearing and hearing-impaired listeners for speech that either has been transduced through actual hearing aids, or has been processed artificially to mimic aspects that may occur in real hearing aids (as, for example, filtering, peak clipping, or compression). Both male and female speech served as input signals. The aim of the experiment was (1) to measure the effect of the different processing schemes, (2) to assess the differences between a male speaker and a female speaker as input signals, (3) to compare the responses of a group of normal-hearing subjects and a group of hearing-impaired subjects, (4) to determine the differences between individuals within these two groups, and (5) to get an impression about the test–retest reliability for the present method of measuring sound quality judgments. The results may provide better insight into the factors that govern the preference of hearing aids, and to the degree to which hearing aids should be fitted on an individual basis.

I. EXPERIMENT I: NORMAL-HEARING SUBJECTS

A. Method

1. Stimuli

Two sentences served as stimuli for the present experiment. In one session, subjects had to base their judgments on the Dutch sentence ''De bal vloog over de schutting'' (''The ball flew over the fence''), uttered by a female speaker. In the other session, judgements were based on the sentence ''De voordeur bij de buren klemt'' (''the neighbors' front door jams''), uttered by a male speaker. The two sentences were the very first sentences from two sets originally developed for a reliable measurement of the Speech Reception Threshold or SRT [Plomp and Mimpen (1979) for the female speaker, and Smoorenburg (1986) for the male speaker]. The material was digitized at a sample frequency of 15 620 Hz, and the cutoff frequency of the anti-aliasing filter was set to 6.5 kHz. The two sentences were processed in a number of ways typical of the processing that may occur in real hearing aids.

The normal-hearing subjects received 24 conditions. Each condition was administered once with the male speaker and once with the female speaker. The 24 conditions were divided in seven subgroups (symbols relate to the figures presented and discussed below):

(1) *Unprocessed condition* (∗).
(2) *Reduction of the bandwidth* (●). The signals were band-pass filtered; filter slopes had a roll-off of approximately 72 dB/oct. The signals were band-pass filtered from 500 Hz to (a) 6500 Hz (the cutoff frequency of the anti-aliasing filter), (b) 4000 Hz, (c) 2000 Hz, or (d) 1500 Hz.
(3) *Spectral slope* (□). A spectral tilt of −12, −6, +6, or +12 dB/oct was imposed on the speech signals.
(4) *Peak clipping* (△). Symmetrical hard peak clipping occurred, for the male speech for 5%, 10%, or 20% of the time, and for the female speech 1%, 3%, or 10% of the time.[1]
(5) *Amplitude compression* (▽). The signal's amplitude was compressed if the temporal envelope exceeded a given threshold. Compression thresholds were −5, −18, −38 dB (male speaker), and 1, −10, −36 dB (female speaker) *re*: the rms level of the sentences.[1] The compression ratio was kept fixed to 5:1. The attack and release time were 0 and 50 ms, respectively.
(6) *Noise* (■). Noise with a spectral shape equal to the long-term average spectrum of either male or female speech was added to obtain a signal-to-noise (S/N) ratio of 10, 15, 20, or 25 dB.
(7) *Hearing-aid recordings* (○). Five hearing aids (labeled A to E) were placed on the artificial ear of a KEMAR manikin, and were coupled via a B&K DB2012 External Ear Simulator to a B&K 4157 Ear Simulator, mounted on a B&K 4134 1/2-in. microphone. The output of the microphone was amplified by a B&K 2610 Measuring Amplifier, and the signals were eventually stored on the computer hard disk at a sampling frequency of 15 620 Hz. The manikin was placed in an anechoic room, at 1 m in front of a Vifa custommade loudspeaker. For each of the five hearing aids, the two sentences (male and female voice) were played through the loudspeaker at a typical speaking level of 65 dB SPL (about 60 dBA SPL). The types of hearing aid and their parameter settings were chosen so that they could be used as prostheses for persons with a mild sensorineural hearing loss (typical au-

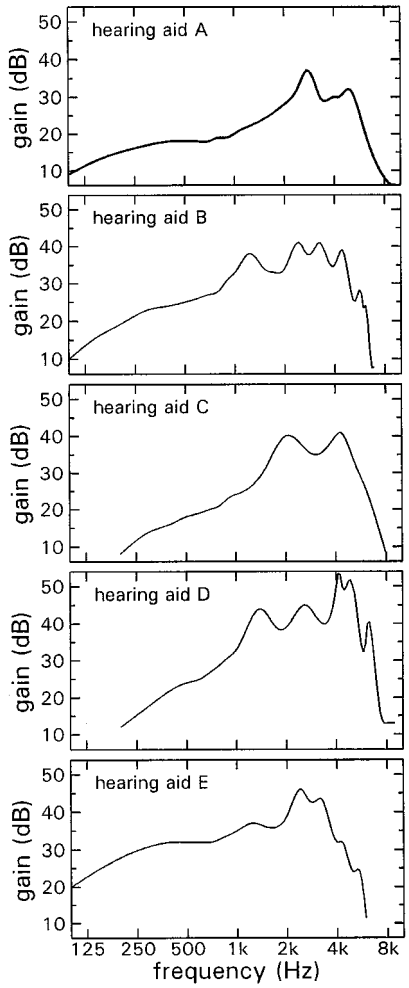


FIG. 1. Frequency response curves for the five hearing aids, measured with an O.E.S. coupler according to IEC 711. Hearing aids A and C were in-the-ear aids, Types B, D, and E were behind-the-ear aids.

diogram: 40 dB HL for frequencies below 1 kHz, −7 dB/oct slope beyond 1 kHz). The gain was adjusted such that the hearing aid acted as a linear amplifier—that is, compression, peak clipping, and saturation were avoided as much as possible.

Frequency response gain curves for the five hearing aids were measured with an Occluded Ear Simulator according to IEC 711, and are plotted in Fig. 1. Types A and C were in-the-ear aids; the other three were behind-the-ear aids. In order to correct for the frequency response of the entire audio-chain (amplifiers, loudspeaker, etc.) as well as the open-ear gain, the recorded material was filtered digitally with a 512-point Finite Impulse Response (FIR) filter, which was approximately the inverse function of the frequency response obtained with the KEMAR head without hearing aid. Since in the present experiment the stimuli were presented over headphones, this filtering yielded a frequency response at the eardrum of the subject comparable to that if the actual hearing aid was worn (cf. Killion and Revit, 1993; Dillon, 1997).

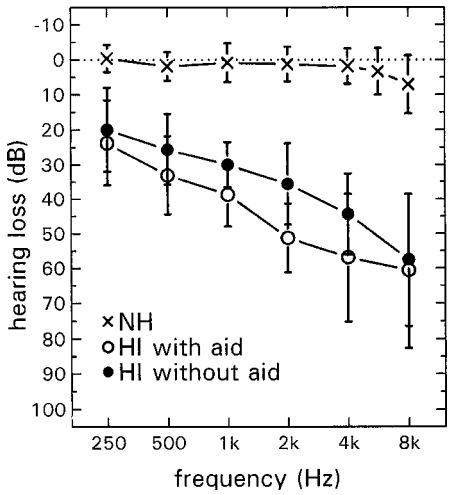


FIG. 2. Audiogram for three groups of listeners. (×) average of 24 normal-hearing subjects; (●) average of 8 hearing-impaired subjects who do not use a hearing aid; (○) average of 8 hearing-impaired subjects who do use a hearing aid. Error bars denote the standard deviation between subjects.

All conditions were calculated in advance, and signals were played out via an OROS-AU21 signal-processing card controlled by a 286 PC.

2. Subjects

Twenty-four subjects (12 male and 12 female, age ranging from 18 to 25 years) participated. Their pure-tone thresholds (as measured with a Madsen OB822 Clinical Audiometer) were 15 dB HL or better in their best ear at octave frequencies between 250 and 8000 Hz. Only one subject had a 30-dB loss in his best ear at 8000 Hz, but had no loss at 6000 Hz. The average pure-tone thresholds are shown in Fig. 2 by crosses. Error bars indicate the between-subject standard deviations. Subjects were mostly university students, and were paid for their services.

3. Procedure

Subjects were tested individually. The subject was seated in a soundproof booth, and received the stimuli monaurally via Sony MDR-CD999 headphones. Before each session, the subject was asked to carefully adjust the unprocessed stimulus to a comfortable level. Subjects adjusted the level of this stimulus on average to 75 dBA SPL (with a between-subject standard deviation of 6 dB). All subsequent stimuli were presented at the same A-weighted level. For normal-hearing listeners signals with the same A-weighted level are perceived as approximately equally loud.

The subject's task was to choose from a pair of sentence presentations the one they preferred. After the presentation of a pair of sentences (which was presented only once), the subject was requested to respond to the question: *You've just listened to two different hearing aids. Suppose that you have to listen all day with one of these two types. Which one would you prefer?* Listeners responded by pressing a ''1'' or a ''2'' on a PC keyboard. They received 12 practice trials in which all stimuli occurred only once, followed by a round robin tournament, in which all 24*23/2=276 possible pairs of the 24 conditions were presented. The sequential order of the pairs was balanced according to a scheme given by Phil-

TABLE I. Results of three principal-components analyses on the data sets obtained with normal-hearing subjects. (a) Male voice, (b) male voice (retest), (c) female voice. For each principal component the eigenvalue, proportion of variance accounted for, and cumulative proportion of variance is given.

	Male voice (test)				Male voice (retest)				Female voice		
Principal component	1	2	3	4	1	2	3	4	1	2	3
Eigenvalue	15.3	2.9	1.7	1.2	14.7	3.0	2.0	1.5	16.8	2.1	1.4
Variance	64.1	12.2	7.2	5.1	61.3	12.4	8.5	6.1	69.8	8.7	5.9
Cumulative variance	64.1	76.3	83.4	88.5	61.3	73.7	82.2	88.3	69.8	78.5	84.3

lips (1964). The conditions were arranged according to a 24×24 Latin Square (Wagenaar, 1969). Subjects were allowed to pause whenever they wished, but were required to take a break after every 50 pairs. The experiment was divided into two sessions: one with a male speaker and one with the female speaker. Each session lasted about 1 h. This experiment was conducted in conjunction with another experiment dealing with the intelligibility of time-compressed speech. The order of these three sessions was also counterbalanced: The six possible orders were evenly distributed over the subjects. In total, the entire experiment lasted about 5 h, breaks included.

4. Retest

All 24 subjects repeated the pairwise-comparison experiment for only the male speaker. The time between test and retest was between 1.0 and 1.5 years. Subjects did not know in advance that they would be asked to participate in the retest.

B. Data analysis

As will become evident in Section I C, between-subject differences were substantial, so that simple averaging across subjects was an inappropriate method of data reduction. Instead, multidimensional scaling was applied. The data obtained in the present experiment were analyzed with Principal Components Analysis (PCA). This technique extracts the most salient dimensions on which the subjects based their preferences, together with the relative weights that each subject attached to each of these dimensions.

In the present experiment, the subject is asked to make preference judgments. According to Thurstone's Law of Comparative Judgment (Thurstone, 1927; Bock and Jones, 1968) this implies that in order to come to a judgment, the subject has to map the different stimuli on a unidimensional preference scale. Suppose that a subject i has received N pairs that comprised stimulus j, and that this stimulus was preferred k_{ij} times to the alternative. The associated preference score z_{ij} then is simply obtained by a z-transformation (e.g., Abramowitz and Stegun, 1970, Table 26.1) given by[2]

$$z_{ij}=\sqrt{2}\Phi^{-1}\left(\frac{k_{ij}}{N}\right). \tag{1}$$

Here, $\Phi^{-1}(u)$ is the inverse function of the Gaussian or Normal probability function $\Phi(z)$, given by

$$\Phi(z)=\int_{-\infty}^{z}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2}\,dt. \tag{2}$$

For each individual subject, the preference scores z_{ij} were determined with Eq. (1). If k_{ij} equaled N (i.e., stimulus j was always preferred), or if k_{ij} equaled zero (stimulus j was never preferred), k_{ij} was set to $(N-\frac{1}{2})$ or $\frac{1}{2}$, respectively. Due to the nature of the pairwise-comparison experiment, absolute values for z_{ij} have little meaning; only the *difference* between two preference scores is a measure of the ability of a subject to discriminate between two conditions with respect to sound quality. In order to make the data suitable for the Principal Components Analysis (PCA), the preference scores were shifted such that, for each subject, $\Sigma_{j=1}^{P}z_{ij}=0$, where P is the number of stimuli in the experiment.

Preference scores z_{ij} can be obtained for each of the M subjects and P stimuli, forming an $M\times P$ matrix, that can be viewed as P stimuli in M-dimensional space, each subject accounting for one dimension (viz., his or her preference axis). This matrix can be subjected to a PCA (see, e.g., Harman, 1970), which is a data-reduction technique that decomposes this matrix into three matrices of rank $R(R<M)$. One matrix is of size $R\times P$, and contains the coordinates of the P stimuli in R-dimensional space (''stimulus space''). The second matrix is of size $M\times R$, and represents the weight that each subject assigns to each dimension (''subject space''). The third matrix, finally, is a diagonal matrix and contains the R eigenvalues λ. The eigenvalue λ of a dimension is equal to the variance accounted for by that dimension divided by the average variance. Dimensions with $\lambda>1$ are considered to be significant.

In summary, PCA reduces the number of dimensions by extraction of the dimensions that subjects have in common, and separates the data into two (R-dimensional) spaces, viz., a ''stimulus space,'' containing the relative position of the different stimuli, and a ''subject space,'' containing information about the relative importance of each dimension for the individual subject.

C. Results and discussion

Inter-individual differences were substantial, which was reflected by the between-subject correlations that ranged from -0.15 to 0.9. Therefore, the data were analyzed with PCA. Three separate PCAs were run—one for the male speaker, one for the female speaker, and one for the retest (with the male speaker). The results of these PCAs are given in Table I, where for each PCA the significant principal components are given together with their eigenvalue, the percentage of the variance accounted for, and the cumulative percentage of the variance.

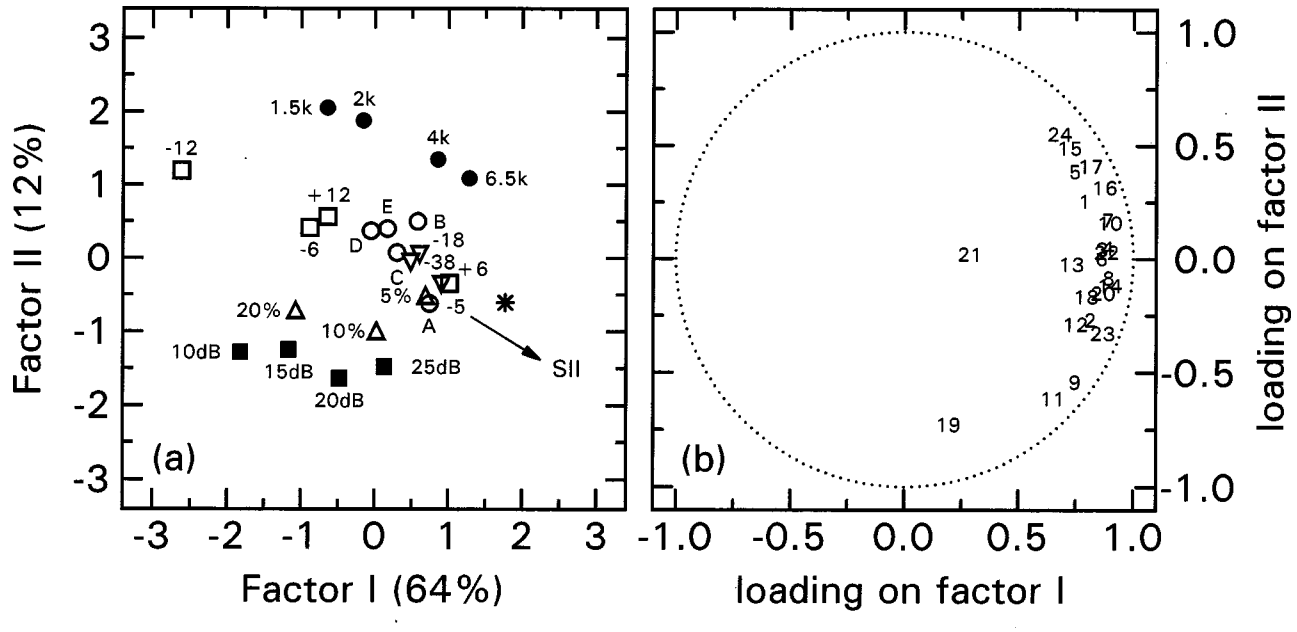


FIG. 3. Results of a principal components analysis for the group of normal-hearing subjects, listening to a male voice. (a) Stimulus space for Factor II versus Factor I. Different symbols indicate a different type of processing, and each symbol is annotated shortly. (*) unprocessed signal; (●) band-limited signals; (□) signals with an imposed spectral slope; (△) peak-clipped signals; (▽) compressed signals; (■) signals with added background noise; (○) signals recorded through a hearing aid. The arrow indicates the direction of maximum increase in the speech intelligibility index (SII). (b) Subject space. Each subject is indicated with a number. The coordinate values correspond to the weight attached to each factor in panel (a).

1. Male speaker

The outcome of the PCA on the data for the male speaker obtained with normal-hearing subjects indicated four significant dimensions. Figure 3(a) displays the so-called ''stimulus space.'' In this panel, Factor II has been plotted versus Factor I. The percentage of variance accounted for by each dimension is given between parentheses. Each group of conditions is indicated with a different symbol (as described in Sec. I A 1). Also, the stimuli of each group are briefly annotated to indicate the specific signal processing condition. The stimulus configuration is such that distances between stimuli reflect subjective distances (with respect to preference). For example, the speech signal that has been band-pass filtered from 500 to 1500 Hz (indicated with a filled circle and labeled ''1.5k'') is fairly similar to the speech signal filtered from 500 to 2000 Hz (indicated with ''2k''). On the other hand, it appears to be very dissimilar to the signals with added background noise (filled squares). The arrow with the label ''SII'' will be discussed in Sec. III A, and deals with the Speech Intelligibility Index.

The arrangement of the stimuli in Fig. 3(a) shows some clear trends. First of all, increased processing causes a shift away from the unprocessed signal (asterisk). As will be shown below, for the majority of normal-hearing subjects this means that increased processing decreases preference. Additional removal of the low-frequency portion of the signal up to 500 Hz has a considerable effect on preference judgments. Removal of the upper part of the frequency spectrum (4000–6500 Hz) does not have a large effect on preference, probably because speech contains little information in that region. However, if the low-pass cutoff frequency is shifted to 2000 or 1500 Hz, a substantial effect is observed. A change in the spectral slope (□) produced large decreases in preference judgments. A spectral slope of −12 dB/oct was for most subjects the least preferred of all stimuli. Positive spectral slopes were preferred to negative slopes of the same magnitude. The position of the spectral-slope stimuli can be described by a parabolalike curve, with an optimum for stimuli with a slightly positive spectral slope of 2.3 dB/oct. Preference was affected by peak clipping (△) more than by compression (▽), even though, in the present experiment, thresholds for compression were much lower than for peak clipping. Compression threshold had only a slight effect on preference, whereas the peak clipping threshold had a much stronger effect. Addition of a soft background noise (■), resulting in a signal-to-noise ratio of 25 dB, substantially decreased preference, and a stepwise decrease in signal-to-noise ratio decreased preference accordingly. The ordering for the hearing aids (○) along Factor I is DECBA (from least to most preferred). The frequency response characteristics of the various hearing aids are rather different (cf. Fig. 1), yet preference scores did not differ very much.

The right panel of Fig. 3 displays the so-called ''subject space.'' In this panel, a number represents an individual subject. The coordinate values of that number indicate the weight assigned to each factor. Thus for example, subject #13 loads mainly on Factor I, meaning that mainly the ordering of the stimuli along Factor I was taken into account, ignoring the stimulus ordering along Factor II. Subject #19, on the other hand, hardly takes Factor I into account, and bases his decision mainly on Factor II. Subject #24 uses both factors with equal weight.

The proportion of the variance accounted for by the two factors is given by the sum of squared weights (and is indicated in the middle row of Table I). If these two factors could account for all the variance in the data of a given

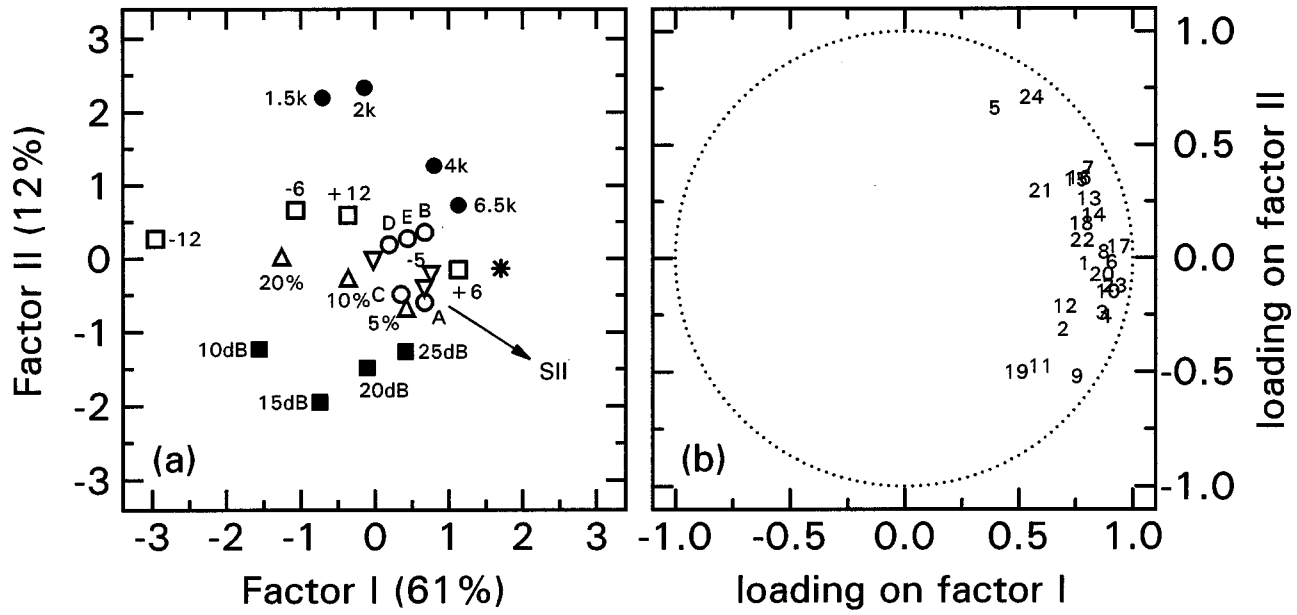


FIG. 4. Results of a principal components analysis for the group of normal-hearing subjects, listening to a male voice. Retest one year later. See caption of Fig. 3 for details.

subject (i.e., could perfectly describe the subject's behavior), the sum of squared weights would be equal to one, and the point representing that subject would lie on the dotted circle. From Fig. 3, it can be seen that most points are close to the edge of this circle, indicating that most of the variance can be accounted for by two factors. Only the preferences of subject #21 and, to a lesser degree, that of subject #19, cannot be described completely with these two factors. Inspection of the data revealed that just these two subjects caused the significance of Factors III and IV. Subject #21 was the only subject that preferred signals with a slightly negative spectral slope as well as the stimuli that were clipped or compressed. He was the sole subject that loaded significantly on Factor III. Subject #19 was the only one that loaded significantly on Factor IV. This subject was inclined to prefer signals with added background noise.

For the individual subject, the ordering of stimuli with respect to preference can be obtained by drawing a vector in stimulus space [Fig. 3(a)] followed by a projection of the stimuli onto this vector. The direction of the vector is given by the relative weights in subject space. For example, the vector for subject #13 is directed to the right, and projection of all stimuli to this vector (which is about a projection of all stimuli to the abscissa) yields the preference ordering for subject #13, based on Factors I and II. This subject thus prefers the unprocessed signal [indicated by the asterisk in Fig. 3(a)] to the other signals, and prefers all other signals to that with a slope of −12 dB/oct (open square with label ''−12''). The direction of the vector of subject #19, on the other hand, is pointed downward, and projection of all stimuli to this vector shows that the unprocessed signal also is frequently preferred, but not as often as the conditions with a little background noise (signal-to-noise ratio of 20 or 25 dB). For this subject, band-limited signals were among the least-preferred conditions. As a last example, subject #24's vector is pointed to the upper-right corner, and this subject clearly prefers all filtered signals, and dislikes all signals that are clipped or have background noise.

Overall, the results display a general agreement among subjects with respect to Factor I, and show that the between-subject differences were mainly due to differences in the weighting of Factor II.

2. Male speaker (retest)

The part of the experiment with normal-hearing subjects and a male voice as source signal was rerun about one year later utilizing the same 24 subjects. The correlation between the preference scores of the test and retest was calculated for each individual subject, and was found to be 0.81 on average (and ranged from 0.57 to 0.95). This number should be contrasted to the correlation *between* different subjects, which was 0.58 on average.

The results of a PCA on the retest data are given in Table I and also indicate four significant dimensions. The stimulus and subject space for Factors I and II are given in Fig. 4. Stimulus space in Fig. 4(a) resembles that of Fig. 3(a) very well (which is even true for the ordering of the five different hearing aids), perhaps not surprisingly, given the relatively large number of subjects. More remarkable is that the subject spaces of Figs. 3 and 4 are similar too, indicating that the individual subjects were consistent over time in their assignment of weights to the different factors. Thus the test–retest reliability is high.

With the retest, dimensions III and IV were not exclusively due to the deviant behavior of subjects #19 and #21, although these two subjects displayed the largest weight on Factor III. Inspection of the stimulus spaces revealed that, in the retest, Factor III subdivides the conditions into one group with clipped or compressed speech and one group with the remaining stimuli. Subject #21 again preferred the clipped or

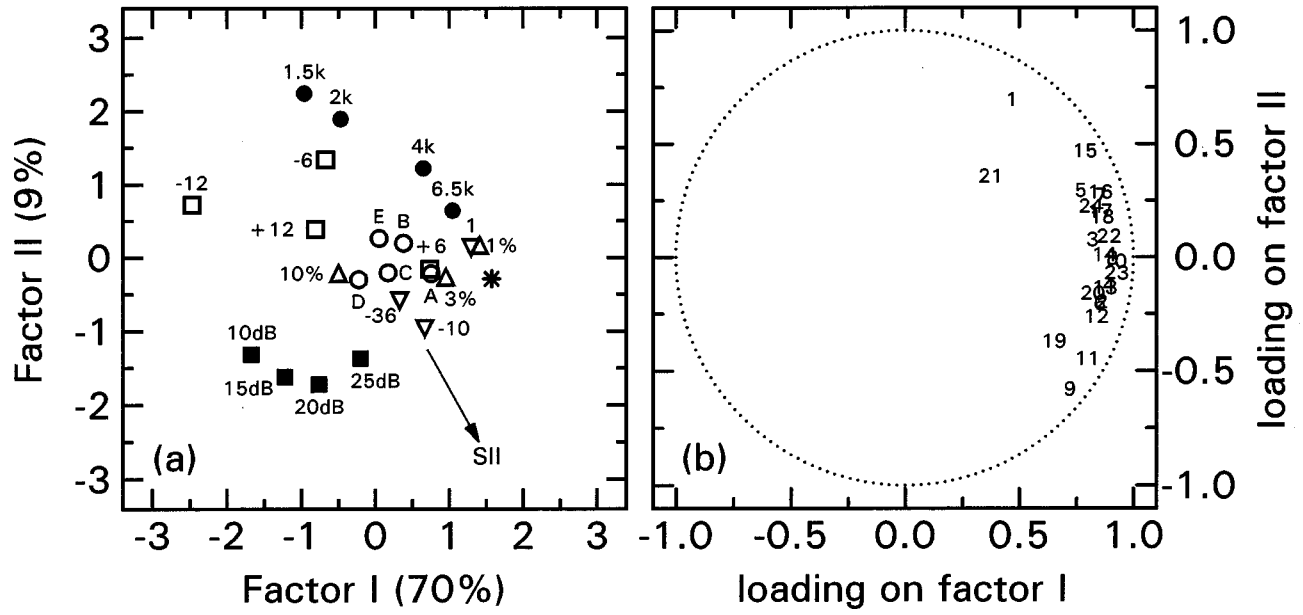


FIG. 5. Results of a principal components analysis for the group of normal-hearing subjects, listening to a female voice. See caption of Fig. 3 for details.

compressed stimuli. Furthermore, Factor IV distinguishes between clipped and compressed speech.

3. Female speaker

Principal Components Analysis on the data obtained with normal-hearing subjects for the female speaker resulted in three significant dimensions (see Table I). Figure 5 displays both stimulus and subject space for Factor I versus II. The stimulus configuration in Fig. 5(a) as well as the relative weights for the individual listeners in subject space in Fig. 5(b) are very similar to those obtained with the male speaker. This indicates that preference judgments are relatively unaffected by voice characteristics and sentence material, and mainly affected by the type and degree of processing. The position of the compressed and clipped stimuli is different for the male and female speaker, since the degree of processing was different. Despite the small positional differences, the ordering of the hearing aids was also similar for both speakers.

As was the case with the male speaker, Factor III was significant due mainly to two subjects. Subjects #21 and, to lesser degree, #19 loaded significantly on Factor III, indicating a preference for stimuli with relatively much low-frequency energy and the compressed and clipped stimuli.

II. EXPERIMENT II: HEARING-IMPAIRED SUBJECTS

A. Method

The setup of experiment II was much similar to that of experiment I. However, in order not to fatigue them too much, the hearing-impaired subjects received a subset of only 16, rather than 24, conditions. Furthermore, a preliminary experiment was conducted with the aim to equalize all stimuli with respect to loudness, such that loudness could not be a cue for preference.

1. Stimuli

A subset was formed out of the stimuli used in experiment I. The 16 conditions are divided in 5 subgroups (symbols relate to the figures presented and discussed below):

(1) *Unprocessed condition* (∗).
(2) *Reduction of the bandwidth* (●). The signals were band-pass filtered; filter slopes had a roll-off of approximately 72 dB/oct. The signals were band-pass filtered from 500 Hz to (a) 6500 Hz (the cutoff frequency of the anti-aliasing filter), (b) 4000 Hz, (c) 2000 Hz, or (d) 1500 Hz.
(3) *Spectral slope* (□). A spectral tilt of −6, +6, or +12 dB/oct was imposed on the speech signals.
(4) *Noise* (■). Noise with a spectral shape equal to the long-term average spectrum of either male or female speech was added to obtain a signal-to-noise (S/N) ratio of 15, 20, or 25 dB.
(5) *Hearing-aid recordings* (○). Five hearing aids (labeled A to E). The recordings were identical to those of experiment I.

2. Subjects

Sixteen hearing-impaired subjects (4 male and 12 female ranging in age from 22 to 57 years old) participated. All suffered from a (usually symmetric) mild-to-moderate high-frequency sensorineural hearing loss, and could in principle be considered as a candidate for the hearing aids used in this experiment. Half of the subjects were experienced hearing-aid users. The individual pure-tone thresholds (in dB HL) are given in Table II, and averages are plotted in Fig. 2 as open symbols for the hearing-aid users, and as filled symbols for those who did not use a hearing aid. Error bars indicate the between-subject standard deviation. Subjects participated on voluntary basis.

TABLE II. Pure-tone thresholds (dB HL) for the group of hearing-impaired subjects. Also given is the age, gender, and whether or not a hearing-aid user. The first row contains the average pure-tone threshold for the group of normal-hearing subjects.

				Frequency (Hz)					
Subject	Age	Gender	Aid	250	500	1000	2000	4000	8000
NH	--	--	--	0	2	1	1	2	7
HI01	47	♀	yes	45	50	50	50	75	80
HI02	50	♀	yes	25	35	45	60	75	80
HI03	47	♀	no	25	40	30	25	45	75
HI04	22	♀	no	20	30	45	45	45	70
HI05	23	♀	yes	25	40	30	50	30	50
HI06	52	♀	yes	15	20	35	60	45	30
HI07	49	♂	no	15	20	30	45	55	55
HI08	39	♂	no	35	25	30	30	30	45
HI09	24	♀	yes	5	15	40	45	45	50
HI10	42	♀	yes	20	35	50	45	60	45
HI11	41	♀	no	35	35	30	20	30	35
HI12	45	♀	no	20	30	25	30	40	45
HI13	44	♀	yes	20	30	25	65	80	95
HI14	44	♀	yes	35	40	35	35	45	55
HI15	49	♂	no	10	10	25	35	45	45
HI16	57	♂	no	0	15	25	55	65	90

3. Procedure

The procedure for hearing-impaired subjects was essentially the same as that described in experiment I. Subjects did not use their hearing aid during the test. The hearing-impaired subjects received the stimuli over their preferred ear, except for one subject (#HI14), who had normal hearing in one ear, and received the signals over the impaired ear. Since, for hearing-impaired listeners, the A-weighted level of a stimulus may be an unreliable predictor of loudness, a preliminary experiment was carried out to equalize all conditions with respect to their loudness. First, the subject was asked to adjust the intensity of unprocessed speech (one sentence) to its most comfortable level. To avoid a possible influence of intelligibility, the speech was time-reversed (i.e., played backward). Next, the sentence (still reversed in time) was partitioned in two halves, of which the first part remained unprocessed, but the other part was processed according to one of the processing techniques. The two parts were continuously repeated, and the subject was asked to adjust the intensity of the second (processed) part such that the two parts became equal in loudness. This was done once for each condition and for both speakers.

Pairs were formed with 16 stimuli, resulting in $16*15/2=120$ stimulus pairs. The conditions and the pairs were randomized and counterbalanced in a similar way. Two sessions were conducted, one with the male speaker and one with a female speaker. Each session lasted about 20 min. The hearing-impaired subjects also participated in the experiment dealing with the intelligibility of time-compressed speech. In total, the entire experiment lasted about 3 h.

B. Results

1. Male speaker

A PCA on the data obtained with hearing-impaired subjects indicated two significant dimensions (cf. Table III).

TABLE III. Results of two principal-components analyses on the data sets obtained with hearing-impaired subjects. (a) Male voice, (b) female voice. For each principal component the eigenvalue, proportion of variance accounted for, and cumulative proportion of variance is given.

	Male voice		Female voice		
Principal component	1	2	1	2	3
Eigenvalue	10.8	1.9	11.1	1.7	1.0
Variance	67.5	11.9	69.6	10.6	6.4
Cumulative variance	67.5	79.4	69.6	80.2	86.6

Figure 6(a) and (b) displays the stimulus space and the subject space for Factors I and II for the present subset, respectively. Indeed, it can be seen from the subject space that the two factors can account well for most of the variance with all subjects. In general, speech with high-frequency emphasis (+6 dB/oct spectral slope, high-pass filtering at 500 Hz, or the various hearing aids) were preferred to the unprocessed speech. Given the fact that most listeners had a high-frequency hearing loss, this is not a surprising result, but it clearly shows that sound quality judgments obtained with normally hearing subjects do not necessarily have to be similar to those of the hearing impaired. The data show that most hearing-impaired subjects, just as normal-hearing subjects, were rather annoyed by the presence of even a weak background noise. There is a clear distinction between the stimuli with added background noise and the other stimuli. The ordering of the five hearing aids was DCBEA, going from least to most preferred. In comparison with the ordering with normal-hearing subjects, hearing aid E was shifted to a more favorable position. The position of the band-pass filtered stimuli seemed somewhat capricious. They were logically ordered along Factor I, but not along Factor II.

2. Female speaker

A PCA on the data obtained with hearing-impaired subjects judging a female voice identified three significant factors. The stimulus space and the subject space for the first two factors are depicted in Fig. 7. Stimulus space obtained with a female speaker resembled that obtained with a male speaker. No clear interpretation could be given to Factor III. No systematic differences could be found between the subjects using hearing aids and those who did not.

III. CORRELATIONS WITH PHYSICAL MEASURES

The purpose of the experiment reported in this paper was to gain insight into the underlying factors that play a role in preference judgments of hearing aids. So far, only little attention has been paid to the interpretation of Factors I and II.

Gabrielsson and Sjögren (1979b) regarded ''clarity'' as one of the most important factors underlying sound quality. It is worth mentioning that with speech as a signal, this dimension often is said to be closely related to (Eisenberg *et al.*, 1997b) or even taken equivalent to (Preminger and van Tassel, 1995a) perceived intelligibility. Listeners seem to be fairly well able to rate intelligibility (Speaks *et al.*, 1972; Cox and McDaniel, 1984; Purdy and Pavlovic, 1992), thus it

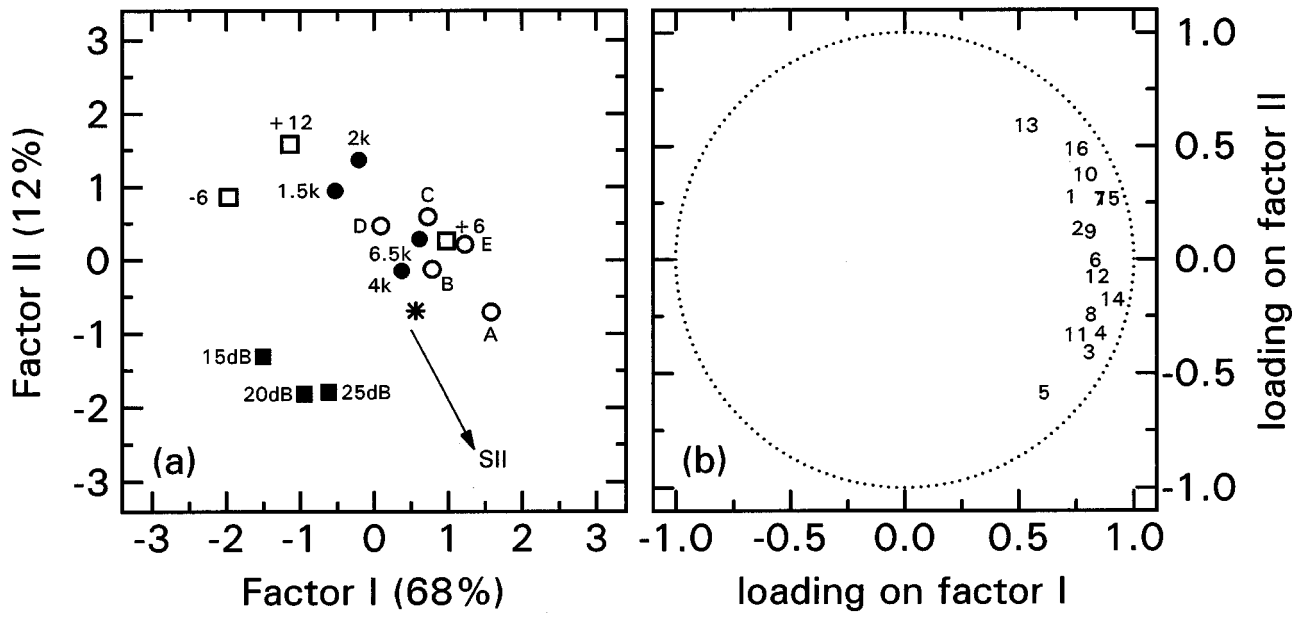


FIG. 6. Results of a principal components analysis for the group of hearing-impaired subjects, listening to a male voice. See caption of Fig. 3 for details.

is likely that intelligibility plays an important and probably dominating role in preference judgments of hearing-aid processed speech. Indeed, one of the principal dimensions from McDermott's (1969) results was interpreted as ''overall clarity.'' With the present results, obtained with normal-hearing subjects, the negative diagonal also seems to be related to perceived intelligibility. The results obtained with hearing-impaired subjects, displayed in Figs. 6 and 7, show the same trends, albeit less evident. In Sec. III A an attempt is made to relate this factor to the Speech Intelligibility Index (SII).

Inspection of Figs. 3, 4, and 5, obtained with normal-hearing subjects, shows that their stimulus spaces were rather similar. It can be seen that the conditions with band-pass filtering are opposed to those containing added noise. Thus subjects apparently make a distinction between those two types of signal processing. McDermott (1969), who conducted pairwise-comparison experiments to assess the sound quality of different electronic circuits, made similar observations. From her multidimensional analyses she concluded that this dimension could be interpreted as the distinction between ''distortion of the signal'' and ''added background distortion.'' In Sec. III B, an attempt is made to relate this dimension to the signal-to-distortion ratio (SDR), which is an objective measure for the amount of nonlinearity in a transmission channel.

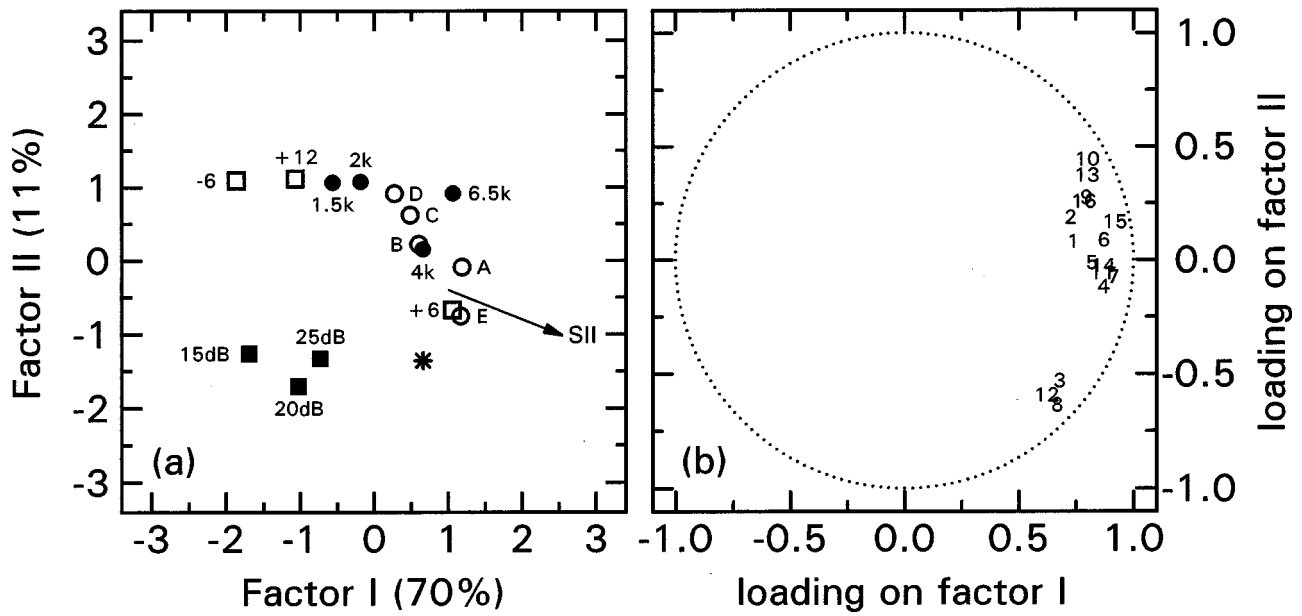


FIG. 7. Results of a principal components analysis for the group of hearing-impaired subjects, listening to a female voice. See caption of Fig. 3 for details.

TABLE IV. Speech Intelligibility Indices (SIIs) and Signal-to-Distortion Ratios (SDRs) for a subset of 16 conditions. SIIs values have been averaged across subjects.

	Normal hearing				Impaired hearing			
	Male voice		Female voice		Male voice		Female voice	
Condition	SII	SDR	SII	SDR	SII	SDR	SII	SDR
Unprocessed	0.92	10.3	0.96	7.2	0.63	10.3	0.65	7.2
500–6500 Hz	0.93	9.9	0.96	7.3	0.63	9.9	0.65	7.3
500–4000 Hz	0.93	3.5	0.92	2.6	0.63	3.5	0.65	2.6
500–2000 Hz	0.87	−0.7	0.83	−1.2	0.57	−0.7	0.56	−1.2
500–1500 Hz	0.81	−2.9	0.71	−2.8	0.51	−2.9	0.51	−2.8
−6 dB/oct	0.85	1.6	0.83	0.2	0.44	1.6	0.42	0.2
+6 dB/oct	0.97	13.3	0.97	10.6	0.71	13.3	0.75	10.6
+12 dB/oct	0.97	12.7	0.97	11.7	0.73	12.7	0.69	11.7
Noise 15 dB S/N	0.87	5.9	0.91	5.5	0.57	5.9	0.62	5.5
Noise 20 dB S/N	0.92	8.2	0.96	6.6	0.61	8.2	0.64	6.6
Noise 25 dB S/N	0.92	9.2	0.96	7.1	0.61	9.2	0.61	7.1
Hearing-Aid A	0.96	−5.9	0.96	−8.2	0.71	−5.9	0.73	−8.2
Hearing-Aid B	0.92	−6.2	0.94	−8.4	0.64	−6.2	0.68	−8.4
Hearing-Aid C	0.95	−6.1	0.96	−8.3	0.71	−6.1	0.73	−8.3
Hearing-Aid D	0.93	−6.0	0.96	−9.0	0.64	−6.0	0.71	−9.0
Hearing-Aid E	0.90	−6.8	0.91	−9.4	0.63	−6.8	0.65	−9.4

A. Speech intelligibility index

To investigate to what extent intelligibility played a role in the preference judgments, speech intelligibility indices (SII, ANSI, 1997) were calculated. The SII is determined by calculation of the audible part of the speech signal. To that end, the signal is split up into a number of frequency bands, and within each band the portion of the speech signal is determined that rises out above the absolute threshold or the possible background noise. The SII is equal to the sum of these portions weighted across the bands (while utilizing the standard importance function) and is said to be monotonically related to intelligibility. Completely audible speech results in an SII of unity; completely inaudible speech yields an SII of zero. The SII model cannot cope with nonlinear distortions, so only conditions where the speech spectrum was manipulated linearly (viz., band-pass filtering, imposing a spectral slope, and routing through a hearing aid), or those conditions with added noise could be taken into account. Since the subjects had different hearing losses, and since subjects were allowed to adjust the level of every single stimulus condition, SIIs needed to be calculated for each subject and condition separately. For the subset of 16 conditions, the SIIs averaged across subjects are given in Table IV. For comparison: The Speech Reception Threshold (SRT) for the sentence materials used in the present experiments (i.e., the signal-to-noise ratio for which 50% of the sentences is correctly perceived) corresponds to an SII of about 0.3. Thus the data in Table IV show that with the normal-hearing subjects, intelligibility for all conditions was well above threshold. For hearing-impaired subjects, SIIs were considerably lower, and especially the −6 dB/oct condition yielded poor intelligibility. It is interesting to compare the values in Table IV with the stimulus positions in Figs. 3–7. For example, the SII values may explain the somewhat capricious positions of the conditions with added noise in case of the group with hearing-impaired listeners. Due to the fact that the different conditions did not have the same intensity (for they were equated in loudness by the subject), SIIs display a nonmonotonic behavior.

For each subject, the preference scores were correlated with the SIIs. With the group of normal-hearing subjects, Spearman's rank-order correlation ranged from −0.5 to about 0.9, with a median of 0.4. With hearing-impaired subjects, the correlation ranged from 0.3 to 0.9, with a median of 0.5. The direction of the arrow in the stimulus space of Figs. 3–7 indicates the direction with the steepest increase in SII, therefore the steepest increase in intelligibility.

The SII calculations suggest that, notably for the hearing-impaired subjects, preference judgments were based to an important degree on the subjective judgment of the intelligibility, even in the range where the speech was clearly intelligible and SII differences were small. This is in accordance with the experimental findings of, for example, Purdy and Pavlovic (1992) who state that subjects are very well able to rate intelligibility.

B. Signal-to-distortion ratio

The second important dimension was labeled ''distortion of the signal versus addition of background distortion,'' which suggests a correlation between preference scores and the amount of nonlinear distortion. A number of investigators (Fortune and Preves, 1992a, b; Kates and Kozma-Spytek, 1994; Palmer *et al.*, 1995; Kozma-Spytek *et al.*, 1996) used coherence, or coherencelike measures to describe the preference ratings of hearing aids. The coherence between an input signal and an output signal is defined as the proportion of the output signal that is linearly related to the input signal. Thus if a signal is linearly transformed only (e.g., by filtering), in principle a coherence of unity is obtained. If the output signal is unrelated to the input signal, the coherence is zero. For all stimuli, the coherence between the unprocessed original signal (i.e., the stored waveform) and the processed signal at the output of the headphones was determined. These coherence values were converted to signal-to-distortion ratios or SDRs (Kates, 1992), expressed in decibels, and are given in Table IV. The SDR values for the different conditions seem odd at first, since one would expect large SDRs for filtered signals and small SDRs for signals with added noise. There are two main reasons for this counterintuitive result. First, coherence in the filtered part of the signal (i.e., the part below 500 Hz and above 1500, 2000, or 4000 Hz) is essentially zero due to background noise, whereas it is higher for signals with added noise. Second, the coherence is determined by averaging across frequency on a linear scale, thus emphasizing the higher frequencies.

The SDRs were correlated to the preference scores of each individual subject. Correlations were mainly positive for normal-hearing subjects (with a mean of 0.14 and a standard deviation of 0.16) and negative for hearing-impaired subjects (with a mean of −0.28 and a standard deviation of 0.18). Thus the SDR can account only very modestly for the obtained results. Although the idea was that the coherence could distinguish between signal distortions and background distortions, the calculations show that this is not the case, at least not for the present experiment, probably due to the two

reasons described above. It is conceivable that other coherencelike measures, as e.g., proposed by Kates and Kozma-Spytek (1994), are perceptually more relevant.

IV. DISCUSSION

The main results of the experiments reported in this paper indicate that both normal-hearing and hearing-impaired subjects are very well able to discriminate reliably between different stimulus conditions in a paired-comparison setting, showing that this experimental procedure is very well suited for these type of listening tests (cf. Eisenberg *et al.*, 1997a). Test–retest reliability was high with normal-hearing subjects, and for both groups essentially no differences were found between a male speaker and a female speaker, indicating that, at least for the present materials, speaker characteristics or speech materials have little influence on preference judgments. Differences between the normal-hearing group and the hearing-impaired group were evident: the latter group in general preferred signals with more high-frequency emphasis. An unexpected finding was that the majority of the subjects disliked even a low-level background noise. For all groups and speakers, the main factors underlying preference were interpreted as (1) intelligibility or clarity, and (2) distortion of the signal versus added background distortion. The speech intelligibility index (SII) could account for the first factor reasonably well, but attempts to relate the second factor to the signal-to-distortion ratio (SDR) were not very successful. Some subjects focused on intelligibility; some were inclined to pay attention to the nonlinearities in the signal. The majority of the subjects, however, took both factors into account, giving both factors about equal weight.

Removal of the low-frequency part of the speech spectrum (up to 500 Hz) had a considerable effect on preference judgments. This finding is well known in audiological practice and is in agreement with the results of Punch *et al.* (1980), Punch and Beck (1980), Punch and Beck (1986), and Tecca and Goldstein (1984). Removal of the upper part of the frequency spectrum (4000–6500 Hz) did not have a large effect on preference, presumably because speech contains little information in that region. However, if the low-pass cutoff frequency was shifted to 2000 or 1500 Hz, a substantial effect was observed. A possible explanation for the nonmonotonic behavior of the band-pass filtered stimuli along Factor II with the group of hearing-impaired subjects (Figs. 6 and 7) was offered by the speech intelligibility index, and probably was due to differences in intensity.

The stimulus spaces in Figs. 3–7 show that the spectral slope (open squares) had a strong effect on preference. Speech with low-frequency emphasis was disliked, as was the case with speech with too much high-frequency emphasis. Thus there seems to be an optimum, and to find the location of this optimum a parabola was fitted to the averaged preference scores. These calculations resulted in an optimum of 2.3 dB/oct for the group of normal-hearing subjects, and 3.9 dB/oct for the group of hearing-impaired subjects. There was no significant effect of speaker, and, surprisingly, there were no differences between the group of hearing-impaired subjects that used a hearing aid and those who did not. This is reassuring in that habituation to a hearing aid (especially with respect to the spectral slope) probably does not affect the relative preference. Since the headphones had an overall roll-off of −1.8 dB/oct, the net spectral slope for the normally hearing and hearing impaired is 0.5 dB/oct and 2.1 dB/oct, respectively. The latter slope is about 0.3 times the average slope in the hearing loss of this group of subjects (which was −7.5 dB/oct cf. Fig. 2), and corresponds well to one-third slope rule in the NAL procedure for hearing-aid gain (Byrne and Dillon, 1986). In agreement with the results in the literature (Crain and van Tassel, 1994; Hawkins and Naidoo, 1993) at comparable thresholds, compression was preferred to peak clipping, probably because the former produces less distortion products than the latter.

In spite of the small differences, the ordering of the hearing aids was similar for the normal-hearing group, viz. DECBA (from least to most preferred, cf. Figs. 3–5). With hearing-impaired subjects, this ordering was DCBEA, thus only type E changed place in comparison with the results for the normal-hearing subjects. The cutoff frequencies (defined by the points of intersection of the frequency response curve with a line 15 dB below the average gain at 500, 1000, and 2000 Hz) can completely account for this ordering. Hearing-impaired subjects take only the low-frequency cutoff into account (cf. Fig. 1), and preference increased with decreasing low-frequency cutoff. This finding has also been reported by Punch (1978), Punch *et al.* (1980), and Tecca and Goldstein (1984). The fact that hearing aid E has a relatively poor preference score with the normal-hearing group probably was caused by its low high-frequency cutoff. With music signals, it is known that normal-hearing listeners pay attention to both cutoff frequencies whereas hearing-impaired listeners only take the low-frequency cutoff into account (Franks, 1982), a finding that is completely consistent with the present results. The present results also do not contradict the findings of van Buuren and co-workers (1995, 1996), who found that the frequency response of a hearing aid could be varied to a large extent before the sound quality is significantly affected. In their experiments, one cannot speak of low- or high-frequency cutoff, since the speech signal was always placed within the dynamic range.

A direct consequence of the fact that different listeners used several factors which were weighted differently, is that differences between subjects can be substantial. This can account for why one person likes a given hearing aid, whereas another person does not. Although fitting rules might give satisfactory results for a group of people, it is important to know that, on an individual level, different settings may prove to be considerably better.

In the present experiment an attempt was made to determine the factors that underlie preference judgments for speech signals that were processed in divergent ways. It was found that just a few factors could describe the preference judgments very well. Next, an attempt was made to relate these factors to physical attributes of the signal. This was only partially successful. More research effort is required to find out whether, in general, more factors are involved, whether these factors are different for other types of signals (such as music), and whether subjects use the same weights

under different listening conditions. Knowledge of the underlying factors and, more importantly, their relation to physical parameters may significantly speed up the selection and adjustment process in the clinic. If the underlying factors are known beforehand, then one only needs to determine the factor weights of the individual patient (for example, by means of a preliminary listening experiment over headphones), such that it becomes possible to select and adjust the most appropriate hearing aid (i.e., the one that provides the highest sound quality), even before the patient has listened to it. This will increase user satisfaction and therefore the probability of a successful fit.

ACKNOWLEDGMENTS

Research financially supported by Philips Hearing Instruments BV. We wish to thank Gerard Kuijpers and Jos Leenen (PHI) for their assistance with the hearing-aid recordings, and Martin Boschman (Institute for Perception Research, Eindhoven) and W. Heiser (Leiden University) for the stimulating discussions on the analysis of paired-comparison data. The reviewers are acknowledged for their valuable comments.

[1]It was our intention to apply the same degree of peak clipping or compression for the male and female speech. Unfortunately, due to an additional signal-processing step with the female sentence (viz., high-pass filtering at 100 Hz to remove a 50-Hz hum) these levels turned out to be different.

[2]The factor $\sqrt{2}$ in Eq. (1) is due to the specific experimental paradigm. If this factor is included, the difference between two z-scores is equal to the sensitivity d' as defined by signal-detection theory (Macmillan and Creelman, 1991). For this paradigm, the probability of choosing stimulus 1 over stimulus 2 is given by

$$\Pr(''1''|\langle 1,2\rangle) = \Phi\left(\frac{z_1 - z_2}{\sqrt{2}}\right).$$

By definition, preference scores differing by an amount of unity are said to be just discriminable. (Whether or not this difference is *statistically* significant depends on the number of trials.)

Abramowitz, M., and Stegun, I. A. (**1970**). *Handbook of Mathematical Functions* (Dover, New York).

ANSI (**1997**). ANSI S3.5-1997, ''American national standard methods for calculation of the speech intelligibility index'' (American National Standards Institute, New York).

Bock, R. D., and Jones, L. V. (**1968**). *The Measurement and Prediction of Judgment and Choice* (Holden-Day, San Francisco).

Byrne, D., and Dillon, H. (**1986**). ''The National Acoustic Laboratories' (NAL) new procedure for selecting the gain and frequency response of a hearing aid,'' Ear Hear. **7**, 257–265.

Byrne, D., and Walker, G. (**1982**). ''The effects of multichannel compression and expansion amplification on perceived quality of speech,'' Aust. J. Audiol. **4**, 1–8.

Cox, R. M., and McDaniel, D. M. (**1984**). ''Intelligibility ratings of continuous discourse: Application to hearing aid selection,'' J. Acoust. Soc. Am. **76**, 758–766.

Crain, T. R., and van Tassel, D. J. (**1994**). ''Effect of peak clipping on speech recognition threshold,'' Ear Hear. **15**, 443–453.

Dillon, H. (**1997**). ''Converting insertion gain to and from headphone coupler responses,'' Ear Hear. **18**, 346–348.

Eisenberg, L. S., Dirks, D. D., and Gornbein, J. A. (**1997a**). ''Subjective judgments of speech clarity measured by paired comparison and category rating,'' Ear Hear. **18**, 294–306.

Eisenberg, L. S., Dirks, D. D., Takayanagi, S., and Schaefer Martinez, A. (**1997b**). ''Subjective judgment of clarity and intelligibility for filtered stimuli with equivalent Speech Intelligibility Index prediction,'' Poster presented at the 2nd Biennal Hearing Aid Research and Development Conference, 22–24 September 1997, Bethesda, MD.

Eisler, H. (**1966**). ''Measurement of perceived acoustic quality of sound-reproducing systems by means of factor analysis,'' J. Acoust. Soc. Am. **39**, 484–492.

Fortune, T. W., and Preves, D. A. (**1992a**). ''Hearing aid saturation and aided loudness discomfort,'' J. Speech Hear. Res. **35**, 175–185.

Fortune, T. W., and Preves, D. A. (**1992b**). ''Hearing aid saturation, coherence, and aided loudness discomfort,'' Journal of the American Academy of Audiology **3**, 81–93.

Franks, J. R. (**1982**). ''Judgments of hearing aid processed music,'' Ear Hear. **3**, 18–23.

Gabrielsson, A., Schenkman, B. N., and Hagerman, B. (**1988**). ''The effect of different frequency responses on sound quality judgments and speech intelligibility,'' J. Speech Hear. Res. **31**, 166–177.

Gabrielsson, A., and Sjögren, H. (**1979a**). ''Perceived sound quality of sound-reproducing systems,'' J. Acoust. Soc. Am. **65**, 1019–1033.

Gabrielsson, A., and Sjögren, H. (**1979b**). ''Perceived sound quality of hearing aids,'' Scand. Audiol. **8**, 159–169.

Harman, H. H. (**1970**). *Modern Factor Analysis* (The University of Chicago Press, Chicago).

Hawkins, D. B., and Naidoo, S. V. (**1993**). ''Comparison of sound quality and clarity with asymmetrical peak clipping and output limiting compression,'' Journal of the American Academy of Audiology **4**, 221–228.

Jeffers, J. (**1962**). ''Quality judgment in hearing aid selection,'' Journal of Speech and Hearing Disorders **25**, 259–266.

Kates, J. M. (**1992**). ''Optimal estimation of hearing-aid compression parameters,'' J. Acoust. Soc. Am. **94**, 1–12.

Kates, J. M., and Kozma-Spytek, L. (**1994**). ''Quality ratings for frequency-shaped peak-clipped speech,'' J. Acoust. Soc. Am. **95**, 3586–3594.

Killion, M. C., and Revit, L. (**1993**). ''CORFIG and GIFROC: Real ear to coupler and back,'' in *Acoustical Factors Affecting Hearing Aid Performance*, edited by G. A. Studebaker and I. Hochberg (Allyn and Bacon, Boston), pp. 65–85.

Kochkin, S. (**1992**). ''MarkeTrak III identifies key factors in determining consumer satisfaction,'' The Hearing Journal **45**, 39–44.

Kozma-Spytek, L., Kates, J. M., and Revoile, S. G. (**1996**). ''Quality ratings for frequency-shaped peak-clipped speech: Results for listeners with hearing loss,'' J. Speech Hear. Res. **39**, 1115–1123.

Kuk, F. K., Tyler, R. S., and Mims, L. (**1990**). ''Subjective rating of noise-reduction hearing aids,'' Scand. Audiol. **19**, 237–244.

Macmillian, N. A., and Creelman, C. D. (**1991**). *Detection Theory: A User's Guide* (Cambridge University Press, Cambridge).

McDermott, B. J. (**1969**). ''Multidimensional analyses of circuit quality judgments,'' J. Acoust. Soc. Am. **45**, 774–781.

Neuman, A. C., Bakke, M. H., Mackersie, C., Hellman, S., and Levitt, H. (**1995**). ''Effect of release time in compression hearing aids: Paired-comparison judgments of quality.'' J. Acoust. Soc. Am. **98**, 3182–3187.

Palmer, C. V., Killion, M. C., Wilber, L. A., and Ballad, W. J. (**1995**). ''Comparison of two hearing aid receiver-amplifier combinations using sound quality judgments,'' Ear Hear. **16**, 587–598.

Phillips, T. (**1964**). ''On the presentation of stimulus-objects in the method of paired comparisons,'' Am. J. Psychol. **77**, 660–664.

Plomp, R., and Mimpen, A. M. (**1979**). ''Improving the reliability of testing the speech reception threshold for sentences,'' Audiology **18**, 43–52.

Preminger, J. E., and van Tassell, D. J. (**1995a**). ''Quantifying the relation between speech quality and speech intelligibility,'' J. Speech Hear. Res. **38**, 714–725.

Preminger, J. E., and van Tassell, D. J. (**1995b**). ''Measurement of speech quality as a tool to optimize the fitting of a hearing aid,'' J. Speech Hear. Res. **38**, 726–736.

Punch, J. L. (**1978**). ''Quality judgments of hearing aid-processed speech and music by normal and otopathologic listeners,'' Journal of the American Audiology Society **3**, 179–188.

Punch, J. L., and Beck, E. L. (**1980**). ''Low-frequency response of hearing aids and judgments of aided speech quality,'' Journal of Speech and Hearing Disorders **45**, 325–335.

Punch, J. L., and Beck, L. B. (**1986**). ''Relative effects of low-frequency amplification on syllable recognition and speech quality,'' Ear Hear. **7**, 57–62.

Punch, J. L., Montgomery, A. A., Schwartz, D. M., Walden, B. E., Prosek, R. A., and Howard, M. T. (**1980**). ''Multidimensional scaling of quality

judgments of speech signals processed by hearing aids,'' J. Acoust. Soc. Am. **68**, 458–466.
Purdy, S. C., and Pavlovic, C. V. (**1992**). ''Reliability, sensitivity and validity of magnitude estimation, category scaling and paired-comparison judgments of speech intelligibility by older listeners,'' Audiology **31**, 254–271.
Smoorenburg, G. F. (**1986**). ''Speech reception in individuals with noise-induced hearing loss and its implication for hearing loss criteria,''in *Basic and Applied Aspects of Noise-Induced Hearing Loss*, edited by R. J. Salvi, D. Henderson, R. P. Hamernik *et al.* (Plenum, New York).
Speaks, C., Parker, B., Harris, C., and Kuhl, P. (**1972**). ''Intelligibility of connected discourse,'' J. Speech Hear. Res. **15**, 590–602.
Staffeldt, H. (**1974**). ''Correlation between subjective and objective data for quality loudspeakers,'' Journal of the Audio Engineering Society **22**, 402–415.
Studebaker, G. A., and Sherbecoe, R. L. (**1988**). ''Magnitude estimations of the intelligibility and quality of speech in noise,'' Ear Hear. **9**, 259–267.
Sullivan, J. A., Levitt, H., Hwang, J.-Y., and Hennessey, A.-M. (**1988**). ''An experimental comparison of four hearing aid prescription methods,'' Ear Hear. **9**, 22–32.
Tecca, J. E., and Goldstein, D. P. (**1984**). ''Effect of low-frequency hearing aid response on four measures of speech,'' Ear Hear. **5**, 22–29.
Thurstone, L. L. (**1927**). ''A law of comparative judgment,'' Psychol. Rev. **34**, 273–286. Reprinted in: Psychol. Rev. 101, 266–270.
van Buuren, R. A., Festen, J. M., and Plomp, R. (**1995**). ''Evaluation of a wide range of amplitude-frequency responses for the hearing-impaired,'' J. Speech Hear. Res. **38**, 211–221.
van Buuren, R. A., Festen, J. M., and Houtgast, T. (**1996**). ''Peaks in the frequency response of hearing aids: Evaluation of the effects on speech intelligibility and sound quality,'' J. Speech Hear. Res. **39**, 239–250.
Wagenaar, W. A. (**1969**). ''Note on the construction of Digram-Balanced Latin Squares,'' Psychol. Bull. **72**, 384–386.
Yonovitz, A., Bickford, B. J., Lozar, J., and Ferrell, D. R. (**1978**). ''Electroacoustic distortions: Multidimensional analysis of hearing aid transduced speech and music,'' IEEE International Conference of Acoustics, Speech and Signal Processing, 270–274.

Classification of dolphin echolocation clicks by energy and frequency distributions

D. S. Houser
Department of Biology, E&MS A316, University of California, Santa Cruz, California 95064

D. A. Helweg and P. W. Moore
SPAWAR Systems Center San Diego, Code 351, 49620 Beluga Road, San Diego, California 92152-6506

(Received 7 May 1998; accepted for publication 30 April 1999)

Dolphins demonstrate an adaptive control over echolocation click production, but little is known of the manner or degree with which control is exercised. Echolocation clicks ($N \sim 30\,000$) were collected from an Atlantic bottlenose dolphin (*Tursiops truncatus*) performing object discrimination tasks in order to investigate differential click production. Seven categories of clicks were identified using the spectral conformation and relative position of -3 and -10 dB peaks. A counterpropagation network utilizing 16 inputs, 50 hidden units, and 8 output units was trained to classify clicks using the same spectral variables. The network classified novel clicks with 92% success. Additional echolocation clicks ($N > 24\,000$) from two other dolphins were submitted to the network for classification. Classified echolocation clicks were analyzed for animal specific differences, changes in predominant click type within click trains, and task-related specificity. Differences in animal and task performance may influence click type and click train length. [S0001-4966(99)05109-7]

PACS numbers: 43.80.Ka [FD]

INTRODUCTION

Dolphins can inspect objects by emitting trains or sequences of impulsive sounds, termed clicks, with the interclick interval changing proportionally to target range (Penner, 1988). Bottlenose dolphins (*Tursiops truncatus*) display adaptive control over the emission of echolocation clicks both with respect to amplitude and frequency modulation, although the two are not fully independent of one another (Moore and Pawloski, 1990). Changes in the manner that click trains are utilized have been noted with respect to environmental noise, task specificity, and learning by both bottlenose dolphins and the false killer whale (*Pseudorca crassidens*) (Au *et al.*, 1995b; Brill *et al.*, 1992; Sigurdson, 1995). Furthermore, recent analysis of click train production through the use of chaos mathematics suggests underlying patterns may exist within click trains that are not detected by conventional mathematical techniques (Kremliovsky *et al.*, 1998). Evidence regarding the dynamic sound production system of these small odontocetes and its voluntary control prompts further investigation into click train structure as well as the structure of individual clicks.

Au *et al.* (1995b) classified clicks of the false killer whale into four categories based upon the frequency spectrum of collected echolocation signals. The distribution of categories suggested that the false killer whale utilized relatively broadband signals with peak frequencies between 46 and 100 kHz and that an association between source levels and frequency modulation indicated a physiological constraint on the sound production mechanism. Our goal was to investigate adaptive control over click structure in bottlenose dolphins. We expanded upon this classification technique by establishing criteria for more strictly defined categories, and applied it to a much larger set of data. To fully assess the utility of such a classification scheme and its applicability to the study of dolphin echolocation, comparisons of click utilization were made between bottlenose dolphins performing object detection tasks and between the different intervals of a three-alternative match-to-sample task performed by a single dolphin. The general stability of the scheme was further evaluated by submitting the same data set to an artificial neural network and comparing its classification to the user-defined system.

I. MATERIALS AND METHODS

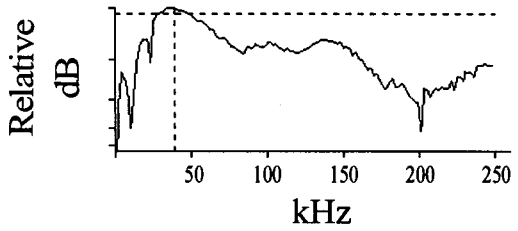
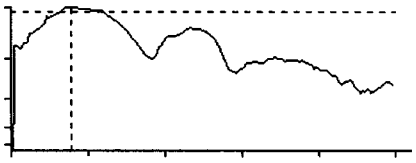
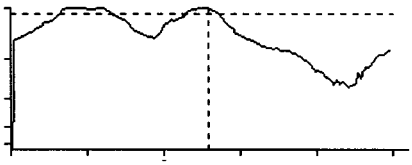
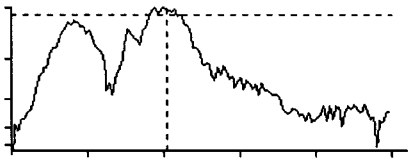
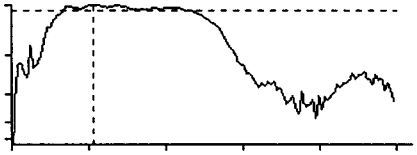
A. Subjects

Two Atlantic bottlenose dolphins (*Tursiops truncatus*: Tt751F, Tt018M) were trained to perform detection tasks in San Diego Bay, and another (Tt598M) performed a two-interval three-alternative match-to-sample (3A MTS) task in Kaneohe Bay, Hawaii. The object detection task was based upon a standard go/no-go paradigm in which the ''go'' response was emitted with the presence of the stimulus object. The two-interval paradigm required the animal to echo-inspect a sample target in the first interval, and echo-inspect a set of comparison targets in the second interval. Following the second interval the dolphin attempted to match the appropriate comparison target to the sample target (for details see Helweg *et al.*, 1996).

B. Three alternative match-to-sample (3A MTS)

Data were collected during match-to-sample tasks. Sample targets were placed 4.65 m in front of the subject and comparison targets were placed 3.65 m in front of the subject and 1.6 m to the left and right of center. The subject was required to echo-inspect objects and signal

TABLE I. Categories of click types, click type description, and a representative spectrum for each. The horizontal dotted line signifies −3 dB region and the vertical dotted line signifies peak frequency.

Click type	Description	Spectrum
A	unimodal, low frequency (< 70 kHz)	
B	unimodal, low frequency (<70 kHz); secondary high-frequency peak (>70 kHz) between −3 dB and −10 dB down	
C	bimodal	
D	unimodal, high frequency (>70 kHz); secondary low-frequency peak (<70 kHz) between the −3 dB and −10 dB down	
E	unimodal, high frequency (>70 kHz)	
W	wideband (single continuously bounded region within the −3-dB bandwidth with frequency bandwidth >85 kHz)	
M	3 or more distinctly bounded regions within the −3 dB	

''comparison'' choices by responding to foam rubber balls attached to flexible PVC rods located above and to the sides of the station (Helweg *et al.*, 1996).

A Bruel & Kjaer 8103 hydrophone (B&K) mounted 2 m from the subject and 1 m underwater was used to detect echolocation clicks. The B&K had a flat frequency response (±3 dB) up to 150 kHz, with a sensitivity of −211 dB at 100 kHz. Detection of a click triggered the computer to store the click after an appropriate delay. The trigger threshold was set at 150 dB which resulted in capture of all clicks in each click

train and clicks were amplified 20 dB (Hewlett-Packard 465 A). The clicks were digitized at 500 kHz with 12-bit resolution using an RC Electronics ICS-16 ComputerScope A/D board and 256 points per waveform were stored in a PC.

C. Object detection

During the object detection task, targets were placed 10 m in front of the subject. All targets were attached to monofilament line and lowered during target ''present'' trials. Targets were lowered to a depth of 1 m for presentation to the subject. A circular aperture was placed with the center 1 m underwater. The subject was required to echo-inspect objects through this ''window.'' An aluminum sheet was used to block attempts to echolocate on targets prior to and between trials. It was placed between the subject and targets and suspended from pulleys for lowering and raising from the shelter. Responses by the subject were made to a foam rubber ball attached to a flexible PVC rod located above and to the side of the station.

A Bruel & Kjaer 8103 hydrophone (B&K) mounted 1 m from the subject and 1 m underwater was used to detect echolocation clicks. The B&K had a flat frequency response (± 3 dB) up to 150 kHz, with a sensitivity of -211 dB at 100 kHz. Detection of a click triggered the computer to store the click after an appropriate delay. The trigger threshold was set at 150 dB which resulted in capture of all clicks in each click train and clicks were amplified 60 dB (Stanford Research Systems model SR560). The clicks were digitized at 500 kHz with 12-bit resolution using an RC Electronics ICS-16 ComputerScope A/D board and 256 points per waveform were stored in a PC.

D. Click classification

The frequency spectrum of $>30\,000$ clicks collected from Tt598M were visually inspected and seven categories of click types were developed from these observations (Table I). Each category was based upon Boolean characters that described the form of the spectrum. Clicks were classified according to: (1) peak frequency; (2) the number of distinctly bounded regions existing within the 3-dB bandwidth; (3) the secondary peak frequency of a region, if one exists within the 3-dB bandwidth; (4) the frequency bandwidth of distinctly bounded regions existing in the 3-dB bandwidth; (5) the 10-dB bandwidth; (6) the number and peak frequency of modal regions existing within -3-dB and -10-dB bandwidths; (7) and the drop in power of distinctly bounded regions existing between the -3-dB and -10-dB boundaries. A complete list of the rules utilized in the classification process is available upon request.

Type A clicks were defined by unimodal low-frequency (<70 kHz) spectral distributions (see Table I). Type B clicks were defined as unimodal low-frequency clicks with a secondary peak existing at a higher frequency between the -3 dB and -10 dB regions. Type E and type D clicks were, respectively, spectral mirror images of the previously described click types with the primary peak being high frequency (>70 kHz) and the secondary peak occurring at lower frequencies. Type C clicks contained a distinctly bounded bimodal distribution within the -3 dB bandwidth. Type W clicks were defined as wide-band clicks and contained a single bounded region of the spectrum within the -3-dB bandwidth with a frequency bandwidth of >85 kHz. All clicks that had three or more distinctly bounded regions within -3 dB of the peak frequency constituted type M, or multimodal, clicks.

TABLE II. Dolphin identification, gender, and number of echolocation clicks collected from each.

Dolphin ID	Gender	# of clicks	Task
Tt751F	Female	13 679	Detection
Tt018M	Male	11 043	Detection
Tt598M	Male	29 561	3A MTS

The classification process was automated by creating a computer program that used the same Boolean decisions employed by a human expert using frequency spectrum. The implementation of the Boolean rules eliminated the potential for error by human classifiers and, because of the dichotomous nature of the scheme, necessarily classified clicks into one of the defined categories. A threshold peak SPL of 150 dB was established for inclusion of clicks through the Boolean classification program (BCP) and the analysis of the frequency spectrum was restricted from 27 to 150 kHz. A total of 54 283 clicks collected from the three dolphins were classified with the BCP (Table II). Subsets of clicks were randomly chosen and visually compared to the automated scheme to verify the automated process. In all cases clicks were correctly classified. For each trial clicks were categorically summed and each trial was considered an observation for statistical analysis. A Mann–Whitney U-test on rank sums was used to test for differences in click type usage between dolphins that performed object discrimination and between the sample and comparison tasks completed by Tt598M. All tests were performed with $p<0.05$.

In order to test the intuitiveness of the classification scheme, i.e., without the implementation of Boolean rules, echolocation clicks were submitted to an artificial neural network (ANN) for classification. Because of its modular construction relevant to biological models and its potential and historical success at pattern recognition (Dayhoff, 1990; Roitblat *et al.*, 1989; Rojas, 1996), a counterpropagation network was created that utilized 16 inputs, 50 hidden units, and 8 output units. The learning sequence consisted of submitting spectra of a given category, as defined by the Boolean scheme, as input. Training sets of 25 ideal (easily categorized) click spectra representing each click type were submitted to the network during the learning sequence. Additional sets of 125 novel click spectra, termed the ''generalization sets,'' were selected from the remaining and most ideal spectral forms and submitted to the neural network for the initial testing of its categorization ability. After initial testing, the entire click data set was submitted to the ANN and the output was compared to the automated classification program to determine the percentage of overall agreement between the two methods.

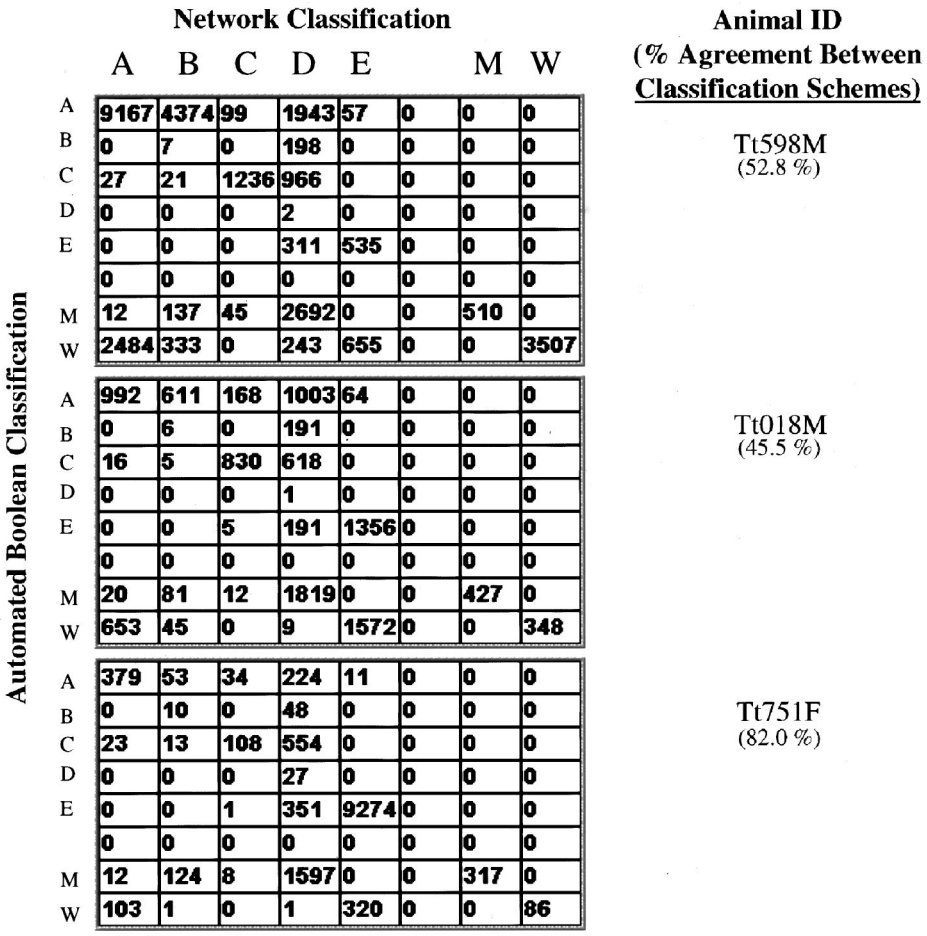


FIG. 1. Distribution of clicks as categorized by the counterpropagation neural network. Clicks are distributed against the Boolean classification scheme and the overall agreement between the neural network and the classification program is given for each animal.

II. RESULTS

A. Neural network

The counterpropagation network achieved a 92% success rate classifying the ''generalization sets'' of selected click types when compared to the automated Boolean classification scheme. The categorization of all of the click data by the ANN is presented in Fig. 1. Agreement between the network and the automated BCP for the entire click data set was variable; the percentage of agreement ranged from 45.5% to 82.0%. ANN classification of click types was most variable for Tt018M. In contrast, the ANN predominantly classified clicks produced by Tt751F as type E, or unimodal high-frequency clicks, while predominantly classifying clicks produced by Tt598M as type A clicks. The ANN classified 12 989 clicks as type D clicks, even though type D clicks, defined by Boolean category rules, were highly underrepresented within the data set (30 of ~54 000 clicks).

B. Click usage

1. Object detection task

Significant differences between Tt751F and Tt018M in click type usage were observed for all categories (Table III). Except for unimodal high-frequency clicks (type E), Tt018M produced a greater number of clicks of each given category. Tt751F produced the lowest mean number of clicks per click train ($\text{mean}_{751}=12.1+11.0$). Most of her clicks were type E, with spectra that had only one peak existing above 70 kHz and within the −3-dB band (Fig. 2). Tt018M also produced relatively few clicks per train ($\text{mean}_{018}=15.8\pm 8.8$) and displayed a broader use of click types (Fig. 2). Tt018M emitted type W clicks during early portions of click trains but did not persist as the click train progressed (Fig. 2). In both cases, the contribution of type B and type D clicks to click train composition was minimal.

Observations of position specific click type proportions indicated a change in click type production for Tt751F as the click train lengthened (Fig. 2). As the mean click train length was exceeded the production of type E clicks switched to that of type A, unimodal low-frequency clicks, and type M clicks which had multiple peak regions within −3 dB of peak amplitude and across the frequency range. In contrast, Tt018M demonstrated no changes in click type across click trains, but produced stable proportions of type A, type E, and

TABLE III. Comparison of click types used by animals performing an object detection task (Tt751F and Tt018M) and a comparison of click types used by Tt598M in both sample and comparison segments of a two-interval match-to-sample task. Asterisks designate significant differences between animals or task interval.

Animal ID	A	B	C	D	E	M	W
Object detection Tt751F vs Tt018M	*	*	*	*	*	*	*
Matching to sample Sample vs comparison	**	**	**		**	**	**

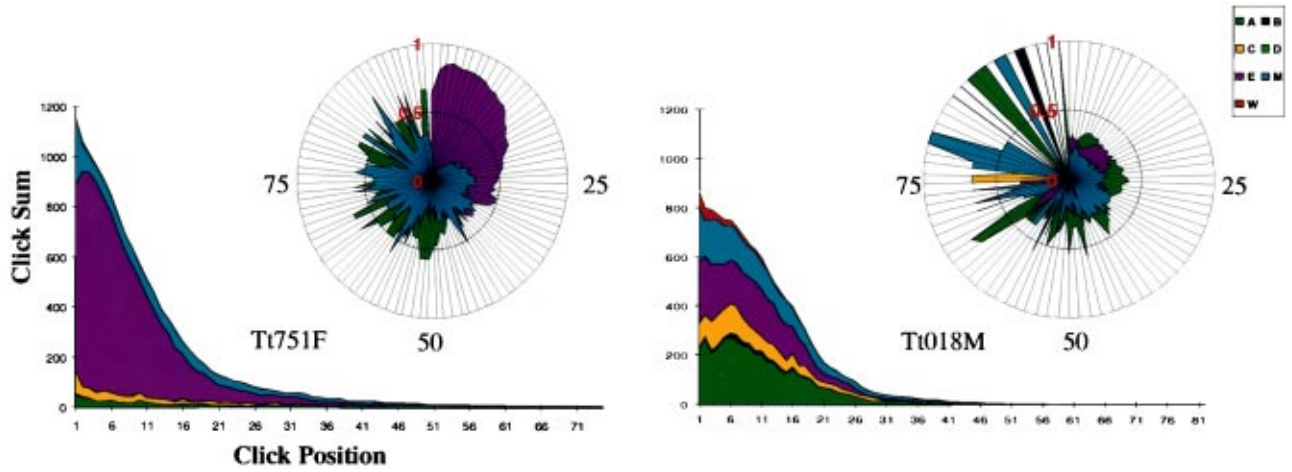


FIG. 2. Rolling sum of click types according to position within the click train for Tt751F and Tt018M performing object detection tasks. Polar plots represent the proportion of click types utilized by position within the click train for the same. Position within the click train is labeled on the periphery of the polar plot. Click types are color coded for identification.

type M when click train lengths of less than 60 clicks were considered (Fig. 2). In the few click trains exceeding 60 clicks in length, the clicks appeared erratic and without uniformity.

2. Matching-to-sample task

Statistical analysis of sample and comparison intervals performed by Tt598M indicated significant differences between all click categories except type D clicks (Table III). During the sample interval a greater number of clicks of all types except A and D were produced. In contrast, the production of type A clicks during the comparison interval was overwhelming (see Fig. 3). Proportional click usage throughout the click train for the other categories appeared to be stable but minimal during this interval. Unfortunately, a true mean click train length could not be determined for Tt598M because emitted click trains were often of a length that exceeded the capacity of the recording system ($\leqslant$99 clicks).

In both sample and comparison intervals, Tt598M produced type A clicks more often than any other click type. During sample intervals, type W and type M clicks were produced at a near-constant proportion across the duration of the train and, early in the click train, even comprised a greater portion of the clicks than type A clicks. In contrast, type A clicks always comprised ~70%–80% of the clicks for the comparison interval, regardless of position within the click train.

III. DISCUSSION

The click classification scheme described here demonstrated qualitative differences between click production by dolphins in similar echolocation tasks, as well as between the intervals of a task performed by the same dolphin (sample inspection versus comparison). Future comparisons should continue to focus on dolphins of different ages and sexes performing identical tasks with identical targets, as well as on the same dolphin performing multiple tasks. If the classification scheme continues to demonstrate qualitative differences in dolphin-dependent and task-dependent echolocation strategies, it should further our understanding of the adaptive control of echolocation as well as the ecological and physiological influences over it.

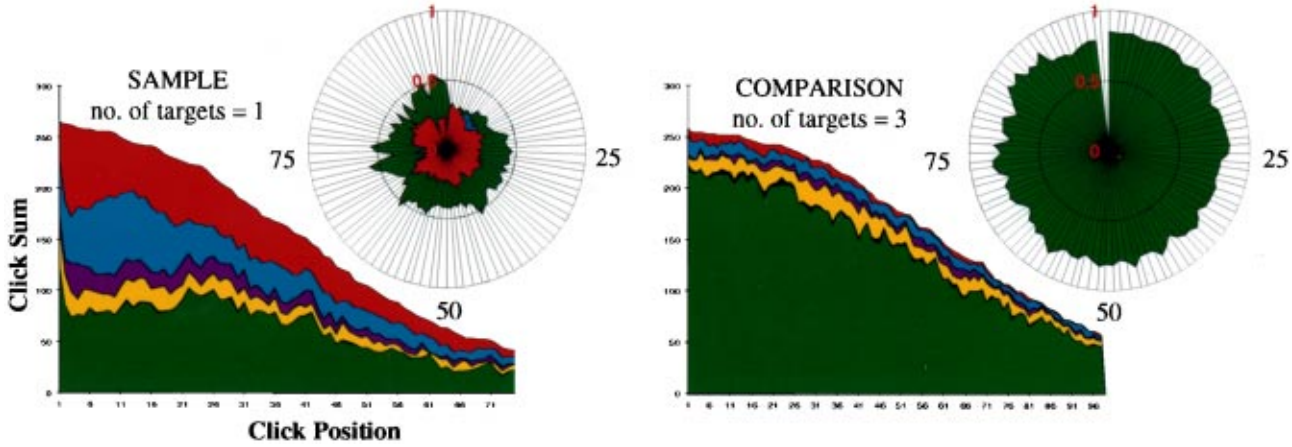


FIG. 3. Rolling sum of click types according to position within the click train for Tt598M performing both sample and comparison intervals of a two-interval match-to-sample task. Polar plots represent the proportion of click types utilized by position within the click train for the same. Position within the click train is labeled on the periphery of the polar plot. Click types are color coded for identification.

Consideration of the relevance of type B and type D categories to the overall classification system may need to be taken into consideration. Both click types contributed relatively little to the overall click production of any animal and type D clicks occurred only 30 times out of >54 000 clicks produced. The correlation between frequency and amplitude of emitted clicks observed in this species and other odontocetes supports the existence of mechanistic constraints on the sound production system (Au *et al.*, 1995b; Moore and Pawloski, 1990; Thomas and Turl, 1990; Thomas *et al.*, 1988). Lack of production of type D clicks, and possibly type B clicks, may relate to these constraints. Alternatively, given that type B and D clicks are similar in spectrum shape to type A and E clicks, respectively, they may represent transitional states away from or toward one of the latter click types. In either case, the classification scheme appears to be emphasizing characteristics that are of little use in determining differences in click production. Further comparisons between dolphins and tasks will determine whether type B clicks should be merged with type A clicks and whether type D clicks should be merged with type E clicks.

Artificial neural networks have been utilized in dolphin bioacoustics but have primarily focused upon aspects of target discrimination (Au *et al.*, 1995a; Au, 1994; Helweg and Moore, 1997; Moore *et al.*, 1991; Roitblat *et al.*, 1992, 1989). These studies utilized various ANN schemes to address the importance of echo features to the discrimination task as well as the biological relevance of the neural processes involved. This study took a different approach by using a counterpropagation network to assess spectral properties of emitted clicks, specifically addressing characteristics that may be used to distinguish between variable types of click production. The ANN performed well when given ideal frequency spectra for a given click type but performance declined when submitted with the entire data set. This suggests that the ANN was capable of learning patterns that were distinctive and of ideal spectral shape for a category, but that performance deteriorated as spectral distributions drifted from the ideal shape. The ANN classification of a large number of clicks as type D is worth particular note. The type D category appeared to act as a attractor in this system, trapping energy states which did not readily settle into other categorical states. The underlying cause of this was likely the under-representation of type D categories within the data set such that a more defined energy state for that category was not learned. Still, the ANN provided general agreement about overall click distributions as compared to the echolocation click classification program; i.e., gross differences between dolphins and intervals of a task were still notable.

The three dolphins demonstrated different degrees of production of specific click types with regard to a given task. The variables which influenced these preferences are potentially numerous and may have included such things as environmental noise, the physiological condition of the animal and the demands of the task. For example, Tt598M performed his echolocation tasks in Kaneohe Bay, a noisy environment when compared to the relatively quiet waters of San Diego Bay (Au *et al.*, 1985). This may have impacted decisions as to which frequencies and amplitudes the animal utilized, presumably adapting sound production to optimize useful echo returns (Au *et al.*, 1985; Moore and Pawloski, 1990). Evidence presented here suggests Tt598M had a preference for low-frequency clicks, which contrasts reports given for other animals performing similar tasks within Kaneohe Bay (Au *et al.*, 1985).

Senescence of the auditory system may result in the alteration of click production in order to accommodate the loss of sensitivity at certain frequencies. For example, Tt018M is 33 years of age and recent audiograms indicate a bilateral decrease in sensitivity above 50 kHz, possibly as a result of age-related retrograde neural loss (Ketten *et al.*, 1997; Brill *et al.*, submitted). Alterations of click train structure in response to the attenuation of returning echoes has been previously demonstrated by Brill and Harder (1991). Senescence of the sound reception mechanism may create an analogous scenario with regard to the sensitivity of both the frequency and amplitude of returning echoes. This may explain the increased production of lower-frequency clicks by Tt018M relative to Tt751F, since clicks of this type should attenuate less rapidly in water than those of higher frequency and would better match his hearing profile (Brill *et al.*, submitted). The strategy used by Tt751F, a 14-year-old female, contrasts that of the older male dolphin. Tt751F's production of high-frequency unimodal clicks may indicate a greater sensitivity to higher-frequency echo returns and a sound reception mechanisms as of yet unaffected by age- or sex-related senescence.

Tt598M produced low-frequency clicks in a similar manner to Tt018M. No audiograms currently exist for Tt598M, but at 17 years of age, he was approaching the age range for which decreases in high-frequency sensitivity have been noted for males of this species (Ridgway and Carder, 1993, 1997). Unfortunately, it is impossible to differentiate between the environmental influences, the demands of the task, and possible physiological influences that may impact click production without knowing the hearing sensitivity of the animal or without comparing to other dolphins performing the same task.

Differences in click production by interval suggest that variable strategies may be employed to optimize success at a task. In the matching-to-sample task performed by Tt598M there was a strong shift to type A click usage when performing the comparison interval. Although it is beyond the capacity of this study to address the specifics underlying the change in echolocation strategy, we can speculate that use of low-frequency clicks may have related to the evocation of salient echo features utilized by Tt598M in the decision making process.

IV. CONCLUSIONS

The click classification scheme presented here demonstrated differences between the types of clicks produced by individual dolphins performing similar tasks and by a single dolphin within a task. Further comparisons need to be made between dolphins performing the same task and across a variety of tasks in order to fully evaluate the utility of this classification scheme. If it continues to prove useful, the scheme may provide another tool with which to study dol-

phin echolocation from the point of click production and help to further understand the function of the dolphin sonar system.

ACKNOWLEDGMENTS

The Office of Naval Research graduate fellowship program made this research possible through its summer internship program. We thank L. Dankiewicz and M. Xitco for assisting in tests of the click classification program.

Au, W. W. L. (**1994**). ''Comparison of sonar discrimination: Dolphin and an artificial neural network,'' J. Acoust. Soc. Am. **95**, 2728–2735.

Au, W. W. L., Andersen, L. N., Rasmussen, A. R., Roitblat, H. L., and Nachtigall, P. E. (**1995a**). ''Neural network modeling of a dolphin's sonar discrimination capabilities,'' J. Acoust. Soc. Am. **98**, 43–50.

Au, W. W. L., Carder, C. A., Penner, R. H., and Scronce, B. L. (**1985**). ''Demonstration of adaptation in beluga whale echolocation signals,'' J. Acoust. Soc. Am. **77**, 726–730.

Au, W. W. L., Pawloski, J. L., Nachtigall, P. E., Bonz, M., and Gisiner, R. C. (**1995b**). ''Echolocation signals and transmission beam pattern of a false killer whale (*Pseudorca crassidens*),'' J. Acoust. Soc. Am. **98**, 51–59.

Brill, R. L., and Harder, P. J. (**1991**). ''The effects of attenuating the returning echolocation signals at the lower jaw of a dolphin (*Tursiops truncatus*),'' J. Acoust. Soc. Am. **89**, 2851–2857.

Brill, R. L., Moore, P. W. B., Dankiewicz, L. A., and Ketten, D. R. (**submitted**). ''Dolphin (*Tursiops truncatus*) auditory thresholds measured with jawphones,'' J. Acoust. Soc. Am.

Brill, R. L., Pawloski, J. L., Helweg, D. A., Au, W. W., and Moore, P. W. B. (**1992**). ''Target detection, shape discrimination, and signal characteristics of an echolocating false killer whale (*Pseudorca crassidens*),'' J. Acoust. Soc. Am. **92**, 1324–1330.

Dayhoff, J. E. (**1990**). *Neural Network Architectures: An Introduction* (Van Nostrand Reinhold, New York), pp. 192–216.

Helweg, D. A., Au, W. W. L., Roitblat, H. L., and Nachtigall, P. E. (**1996**). ''Acoustic basis for recognition of aspect-dependent three-dimensional targets by an echolocating bottlenose dolphin,'' J. Acoust. Soc. Am. **99**, 2409–2420.

Helweg, D. A., and Moore, P. W. B. (**1997**). ''Classification of aspect-dependent targets by a biomimetic neutral network,'' NRaD Technical Report 1747. DTIS.

Ketten, D. K., Moore, P. W. B., Dankiewicz, L. A., Brill, R. L., and Van Bonn, W. (**1997**). ''The slippery slope of a Johnsonian ear: Natural variability versus natural loss,'' J. Acoust. Soc. Am. **102**, 3101.

Kremliovsky, M., Kadtke, J., Inchiosa, M., and Moore, P. (**1998**). ''Characterization of dolphin echo-location data using a dynamical classification method,'' Int. J. Bifurcation Chaos. Appl. Sci. Eng. **8**, 813–823.

Moore, P. W. B., and Pawloski, D. A. (**1990**). ''Investigations on the control of echolocation pulses,'' in *Sensory Abilities of Cetaceans: Laboratory and Field Evidence*, edited by J. Thomas and R. Kastelein (Plenum, New York), pp. 305–316.

Moore, P. W. B., Roitblat, H. L., Penner, R. H., and Nachtigall, P. E. (**1991**). ''Recognizing successive dolphin echoes with an integrator gateway network,'' Neural Networks **4**, 701–709.

Penner, R. H. (**1988**). ''Attention and detection in dolphin echolocation,'' in *Animal Sonar: Processes and Performance*, edited by P. Nachtigall and P. Moore (Plenum, New York), pp. 707–714.

Ridgway, S. H., and Carder, D. A. (**1993**). ''High-frequency hearing loss in old (25+years-old) male dolphins,'' J. Acoust. Soc. Am. **94**, 1830.

Ridgway, S. H., and Carder, D. A. (**1997**). ''Hearing deficits measured in some *Tursiops truncatus*, and discovery of a deaf/mute dolphin,'' J. Acoust. Soc. Am. **101**, 590–594.

Roitblat, H. L., Moore, P. W. B., Nachtigall, P. E., Penner, R. H., and Au, W. W. L. (**1989**). ''Natural echolocation with an artificial neural network,'' Int. J. Neural Net. **1**, 239–248.

Roitblat, H. L., Moore, P. W. B., Helweg, D. A., and Nachtigall, P. E. (**1992**). ''Representation and Processing of acoustic information in a biomimetic neural network,'' in *Animals to Animals 2, Proceedings of the Second International Conference on Simulation of Adaptive Behavior*, edited by J. A. Meyer, H. L. Roitblat, and S. W. Wilson (MIT Press, Cambridge), pp. 90–99.

Rojas, R. (**1996**). *Neural Networks: A Systematic Approach* (Springer-Verlag, New York), pp. 411–426.

Sigurdson, J. E. (**1995**). ''Open-water echolocation of bottom objects by dolphins (*Tursiops truncatus*),'' J. Acoust. Soc. Am. **100**, 2610.

Thomas, J. A., and Turl, C. W. (**1990**). ''Echolocation characteristics and range detection threshold of a false killer whale (*Pseudorca crassidens*), in *Sensory Abilities of Cetaceans: Laboratory and Field Investigations*, edited by J. Thomas and R. Kastelein (Plenum, New York), pp. 321–334.

Thomas J., Stoermer, M., Bower, C., Anderson, L., and Garver, A. (**1988**). ''Detection abilities and signal characteristics of echolocating false killer whale (*Pseudorca crassidens*),'' in *Animal Sonar: Processes and Performance*, edited by P. Nachtigall and P. Moore (Plenum, New York), pp. 323–328.

LETTERS TO THE EDITOR

This Letters section is for publishing (a) brief acoustical research or applied acoustical reports, (b) comments on articles or letters previously published in this Journal, and (c) a reply by the article author to criticism by the Letter author in (b). Extensive reports should be submitted as articles, not in a letter series. Letters are peer-reviewed on the same basis as articles, but usually require less review time before acceptance. Letters cannot exceed four printed pages (approximately 3000–4000 words) including figures, tables, references, and a required abstract of about 100 words.

On singular surfaces in isotropic linear thermoelasticity with initial stress[a)]

A. Montanaro[b)]
Dipartimento di Metodi e Modelli Matematici per le Scienze Applicate, via Belzoni 7, Padova, Italy

(Received 19 October 1998; revised 5 May 1999; accepted 5 June 1999)

It will be shown that the theory of a fundamental paper of Chadwick and Powdrill on singular surfaces, propagating in a linear thermoelastic body which is stress-free, homogeneous, and isotropic, also holds when the medium is subjected to hydrostatic initial stress provided the two characteristic speeds are suitably changed. The result is obtained by using Biot's linearization of the constitutive law for the stress. © *1999 Acoustical Society of America.* [S0001-4966(99)03409-8]

PACS numbers: 43.20.Jr [AN]

INTRODUCTION

In Ref. 1 the authors give a detailed account of the propagation of singular surfaces in a homogeneous, isotropic, linearly thermoelastic solid in a natural configuration which is not subjected to body forces or heat sources. They show that the wavefronts associated with the characteristic surfaces of the linear thermoelastic equations propagate with one of the constant speeds

$$v_T = \sqrt{\frac{\lambda+2\mu}{\rho_0}}, \quad v_S = \sqrt{\frac{\mu}{\rho_0}}, \tag{1}$$

where λ and μ are the Lamé moduli of the material and ρ_0 is the mass-density in its reference state.

A singular surface or discontinuity surface of order r in the displacement and temperature fields is referred to as a weak (thermoelastic) wave if $r \geqslant 2$ and a strong (thermoelastic) wave if $r=0$ or 1. Weak waves and strong waves of order 1 are characteristic and may be described as dilatational or rotational according as their propagation speed is v_T or v_S. Decay laws are written down for the strength of each kind of wave and it is pointed out that for waves of order 0 the strength is not completely determined by its distribution on an initial wavefront. Studies of acceleration waves in heat-conducting elastic materials antecedent to Ref. 1 are reviewed in Ref. 2; studies subsequent to Ref. 1 appear in Refs. 3–5. In particular, Ref. 3 studies the effects of a hydrostatic prestress on acceleration waves in isotropic nonlinear thermoelasticity. These papers show that the unified treatment[1] of weak and strong waves in linear thermoelasticity referred to a configuration where the material is isotropic, homogeneous, and unstressed, cannot be extended to nonlinear thermoelasticity. Formulae (1) for the wave speeds have been known since 1939 in the context of harmonic waves in non-heat-conducting isotropic elastic media, and certainly since 1971 (cf. section 4 of Ref. 4) in connection with acceleration waves.

In the present paper we consider an isotropic, linearly thermoelastic body B that in the initial configuration is in a state of hydrostatic stress and show that in the homogeneous case the linear thermoelastic equations for B coincide with the linear thermoelastic equations in Ref. 1 for the unstressed case provided speeds (1) be respectively replaced with

$$\bar{v}_T = \sqrt{\frac{\bar{\lambda}+2\bar{\mu}}{\rho_0}}, \tag{2a}$$

$$\bar{v}_S = \sqrt{\frac{\bar{\mu}-\overset{0}{p}/2}{\rho_0}}, \tag{2b}$$

where $\overset{0}{p}$ is the initial pressure and $\bar{\lambda}$ and $\bar{\mu}$ are the Lamé moduli in Biot's linearization[6] of the constitutive law for the stress.

Hence, by using this coincidence of equations, Ref. 7 shows that any result of Ref. 1 remains true also when the medium is subjected to homogeneous hydrostatic initial stress. Of course $\bar{\lambda}$ and $\bar{\mu}$ in (2) may depend on $\overset{0}{p}$ (cf. Ref. 8, p. 164); when $\bar{\mu}-\overset{0}{p}/2<0$ the characteristic speed (2b) is not real and thus there is no wave propagating with this speed.

[a)]This paper has been performed within the activity of the Consiglio Nazionale delle Ricerche, Group No. 3, in the academic years 1996–97 and 1997–98.

[b)]Electronic mail: montanaro@dmsa.unipd.it

The present paper shows that the assertion (Ref. 9, p. 523) ''... a homogeneous hydrostatic pressure does not affect the laws of wave propagation,'' which holds for purely mechanical linear elasticity, is true for isotropic linear thermoelasticity, too.

Note that speeds (2) coincide with those found in Ref. 10 in the purely mechanical case for discontinuity waves and, in the homogeneous case, for plane progressive waves.

I. ISOTROPIC LINEAR THERMOELASTICITY WITH HYDROSTATIC INITIAL STRESS

Let $\mathbf{B}$ be the closure of a regular, open, and connected subset of the three-dimensional Euclidean ambient space E^3, which is referred to a Cartesian inertial frame of reference. We consider a body B which in its isotropic equilibrium configuration κ occupies the region $\mathbf{B}=\kappa(B)$ under the external actions of the (possibly zero) body force field $\overset{0}{\mathbf{b}}$ on its interior and of the contact force field $\overset{0}{\mathbf{c}}$ on its boundary. We suppose that in $\mathbf{B}$ the body has hydrostatic initial stress

$$\overset{0}{\mathbf{T}}=-\overset{0}{p}\mathbf{I}, \quad \overset{0}{p}=\overset{0}{p}(\mathbf{X}), \tag{3}$$

mass-density $\rho_0=\rho_0(\mathbf{X})$, and uniform absolute temperature T_0, where $\mathbf{X}=\kappa(X)$, $X\in B$. The fields $\overset{0}{\mathbf{T}}$ and $\overset{0}{\mathbf{b}}$ satisfy the equilibrium conditions $\rho_0\overset{0}{b}_i+\partial\overset{0}{T}^{ji}/\partial x_j=0$ on $\mathbf{B}$, which by (3) become $\rho_0\overset{0}{b}_i=\partial\overset{0}{p}^{ji}/\partial x_j$. Moreover, in a given time interval B is subjected to thermoelastic perturbations of small amplitude,

$$\mathbf{u}=\mathbf{x}-\mathbf{X}, \quad \theta=T-T_0, \tag{4}$$

respectively, of position and temperature, about this reference state. Assuming that incremental body forces and heat sources vanish, the linearized field equations for small thermoelastic disturbances (4) read

$$\frac{\partial\rho}{\partial t}+\rho_0\frac{\partial v_j}{\partial x_j}=0, \tag{5a}$$

$$\rho_0\frac{\partial v_i}{\partial t}=\frac{\partial(\sigma^{ji}-\overset{0}{T}^{ji})}{\partial x_j}, \tag{5b}$$

$$\rho_0\frac{\partial u}{\partial t}=\frac{\partial h_j}{\partial x_j} \quad \text{on} \quad \mathcal{B}':=I\times\operatorname{int}\mathbf{B}, \tag{5c}$$

where ρ_0, u and v_i, h_i, and σ^{ji} are the mass-density, the specific internal energy measured from the reference state and the components of the velocity vector, the heat flux vector, and the stress-tensor. These quantities will be referred to as auxiliary variables. The components of the velocity, small strain, and small rotation respectively are

$$v_i=\frac{\partial u_i}{\partial t}, \tag{6a}$$

$$e_{ij}=\frac{1}{2}\left(\frac{\partial u_i}{\partial x_j}+\frac{\partial u_j}{\partial x_i}\right), \tag{6b}$$

$$\omega_{ij}=\frac{1}{2}\left(\frac{\partial u_j}{\partial x_i}-\frac{\partial u_i}{\partial x_j}\right), \tag{6c}$$

where u_i are the components of the displacement vector $\mathbf{u}$. The stress-tensor σ in Eq. (5b) is the transpose $\mathbf{S}^T$ of the first Piola stress-tensor $\mathbf{S}$ used, e.g., in Ref. 11; σ is usually referred to as nominal stress (see, e.g., Ref. 12). Hence the linearization for $\mathbf{S}$ written in Ref. 1 must be understood as referred to σ^T. Thus for isothermal disturbances in B the nominal stress σ admits the linear approximation $\sigma=\mathbf{S}^T=\{\overset{0}{\mathbf{T}}+\mathbf{C}[\nabla\mathbf{u}]\}^T$, where $\overset{0}{\mathbf{T}}=\mathbf{S}(\mathbf{I})$ and $\mathbf{C}$ is the elasticity tensor, that is the linear map $\mathbf{C}:\mathrm{Lin}\to\mathrm{Lin}$ defined by $\mathbf{C}[\mathbf{A}]=D_{\mathbf{F}}\mathbf{S}(\mathbf{I})[\mathbf{A}]$. The above linear law is equivalent to[8,11]

$$\sigma=\{\overset{0}{\mathbf{T}}+\overset{0}{\omega}^T\overset{0}{\mathbf{T}}+\tfrac{1}{2}(\overset{0}{\mathbf{e}}\overset{0}{\mathbf{T}}-\overset{0}{\mathbf{T}}\overset{0}{\mathbf{e}})+\mathbf{L}[\mathbf{e}]\}^T, \quad \mathbf{L}[\mathbf{e}]=\operatorname{sym}\mathbf{C}[\mathbf{e}], \tag{7}$$

where $\mathbf{L}$ is Biot's incremental elasticity tensor.[6] By (3) the constitutive linear law (7) becomes $\sigma=-\overset{0}{p}(\mathbf{I}+\omega)+\mathbf{L}[\mathbf{e}]^T$. We assume that B is isotropic, hence that $\mathbf{L}$ has the isotropic structure

$$L^{pqhk}=\bar{\lambda}\delta^{pq}\delta^{hk}+\bar{\mu}(\delta^{ph}\delta^{qk}+\delta^{pk}\delta^{qh}), \quad \text{with} \quad \bar{\lambda}=\bar{\lambda}(\mathbf{X}), \quad \bar{\mu}=\bar{\mu}(\mathbf{X}), \tag{8}$$

where $\bar{\lambda}$ and $\bar{\mu}$ are the counterparts of the Lamé parameters (see Ref. 11, p. 304). Thus the linearized thermoelasticity theory provides the laws

$$\begin{aligned}&\sigma^{ij}=-\overset{0}{p}(\delta^{ij}+\omega^{ij})+\bar{\lambda}e_{pp}\delta^{ij}+2\bar{\mu}e_{ij}-\frac{\alpha}{\kappa_T}\theta\delta_{ij} \quad (\theta=T-T_0),\\ &h_i=k\frac{\partial\theta}{\partial x_i}, \quad \text{and} \quad u=\frac{\alpha T_0}{\rho_0\kappa_T}e_{pp}+c_e\theta\end{aligned} \tag{9}$$

for the stress-tensor, the heat flux vector, and the internal energy from the reference state, where α, κ_T, k $(\geqslant 0)$, and c_e are respectively the volume coefficient of thermal expansion, the isothermal compressibility, the thermal conductivity, and the specific heat at constant strain of the material in its reference state. Now suppose that the u_i and θ are each of class C^2 on $\mathcal{B}'$. Elimination of the auxiliary variables in favor of the u_i and θ between Eqs. (5b), (5c), (6a)–(6c), and (9) yields the linear thermoelastic equations

$$\begin{aligned}\rho_0\frac{\partial^2u_i}{\partial t^2}=&\left(\bar{\mu}-\frac{\overset{0}{p}}{2}\right)\frac{\partial^2u_i}{\partial x_p\partial x_p}+\left(\bar{\lambda}+\bar{\mu}+\frac{\overset{0}{p}}{2}\right)\frac{\partial^2u_p}{\partial x_i\partial x_p}\\ &-\frac{\alpha}{\kappa_T}\frac{\partial\theta}{\partial x_i}-\frac{\partial\overset{0}{p}}{\partial x_p}\omega_{pi}+\frac{\partial\bar{\lambda}}{\partial x_i}e_{pp}\\ &+2\frac{\partial\bar{\mu}}{\partial x_p}e_{ip}-\theta\frac{\partial}{\partial x_i}\left(\frac{\alpha}{\kappa_T}\right),\end{aligned} \tag{10a}$$

$$\rho_0c_e\frac{\partial\theta}{\partial t}+\frac{\alpha T_o}{\kappa_T}\frac{\partial^2u_p}{\partial t\partial x_p}=k\frac{\partial^2\theta}{\partial x_p\partial x_p}+\frac{\partial k}{\partial x_p}\frac{\partial\theta}{\partial x_p} \tag{10b}$$

on $\mathcal{B}'$. By using Eq. (2), Eq. (10a) becomes

$$\rho_0 \frac{\partial^2 u_i}{\partial t^2} = \rho_0 \overline{v_S^2} \frac{\partial^2 u_i}{\partial x_p \partial x_p} + \rho_0 (\overline{v_T^2} - \overline{v_S^2}) \frac{\partial^2 u_p}{\partial x_i \partial x_p} - \frac{\alpha}{\kappa_T} \frac{\partial \theta}{\partial x_i} - \frac{\partial \overset{0}{p}}{\partial x_p} \omega_{pi} + \frac{\partial \overline{\lambda}}{\partial x_i} e_{pp} + 2 \frac{\partial \overline{\mu}}{\partial x_p} e_{ip} - \theta \frac{\partial}{\partial x_i} \left(\frac{\alpha}{\kappa_T} \right), \tag{11}$$

and thus in the homogeneous case Eqs. (10) reduce to Eqs. (31) of Ref. 1. Reference 7 shows that by this coincidence any result of Ref. 1 remains true also when the medium is subjected to a homogeneous hydrostatic initial stress; moreover, Ref. 7 shows that some results of Ref. 1 remain true in the inhomogeneous case, too; for instance, (*i*) the characteristic equations formally coincide with the ones in the unstressed homogeneous case provided the characteristic speeds (1) be replaced with speeds (2); and (*ii*) weak thermoelastic singular hypersurfaces are characteristic and propagate at speeds (2).

[1] P. Chadwick and B. Powdrill, ''Singular surfaces in linear thermoelasticity,'' Int. J. Eng. Sci. **3**, 561–595 (1965).

[2] C. Truesdell, ''General and exact theory of waves in finite elastic strain,'' Arch. Ration. Mech. Anal. **8**, 263–296 (1961).

[3] P. J. Chen, ''Thermodynamic influences on the propagation and the growth of acceleration waves in elastic materials,'' Arch. Ration. Mech. Anal. **31**, 228–254 (1968).

[4] P. Chadwick and R. W. Ogden, ''A theorem of tensor calculus and its application to isotropic elasticity,'' Arch. Ration. Mech. Anal. **44**, 54–68 (1971).

[5] P. Chadwick and P. K. Currie, ''The propagation of acceleration waves in heat-conducting elastic materials,'' Arch. Ration. Mech. Anal. **49**, 137–158 (1972).

[6] M. A. Biot, ''Nonlinear theory of elasticity and the linearized case for a body under initial stress,'' Philos. Mag. **27**, 468–489 ()1939.

[7] A. Montanaro, ''Details on singular surfaces in isotropic linear thermoelasticity with initial stress,'' Forthcoming in Atti Istituto Veneto Sci. Lett. Arti.

[8] C.-S. Man and W. L. Lu, ''Towards an acoustoelastic theory for measurement of residual stress,'' J. Elast. **17**, 159–182 (1987).

[9] M. A. Biot, ''The influence of initial stress on elastic waves,'' J. Appl. Phys. **11**, 522–530 (1940).

[10] A. Montanaro, ''On small-displacement waves in a prestressed body with isotropic incremental elasticity tensor,'' Meccanica **32**, 505–514 (1997).

[11] A. Hoger, ''On the determination of residual stress in an elastic body,'' J. Elast. **16**, 303–324 (1986).

[12] R. W. Ogden, ''Non-linear Elasticity: Incremental Equations and Bifurcation Phenomena,'' in *Nonlinear Equations in the Applied Sciences*, edited by W. F. Ames and C. Rogers (Academic, New York, 1992), Vol. 185, pp. 437–468.

Three apparently paradoxical results of sound radiation theory

Miguel C. Junger
Cambridge Acoustical Associates/ETC, 84 Sherman Street, Cambridge, Massachusetts 02140-3261

(Received 12 January 1999; accepted for publication 18 May 1999)

Three important generally accepted theoretical results challenge physical interpretation either by being apparently paradoxical or by embodying a surprising coincidence. The first two results bear on the radiation loading of effectively infinite force- or line-excited elastic plates; the third result bears on the radiation loading of an unbaffled rigid piston. © *1999 Acoustical Society of America.* [S0001-4966(99)00609-8]

PACS numbers: 43.20.Rz, 43.20.Tb, 43.40.Dx [DEC]

List of symbols

a = piston radius
ρ = fluid density
ρ_s = density of plate material
c = sound velocity of the fluid
c_p = low-frequency compressional wave velocity in plates
D = flexural rigidity, $h^3\rho_s c_p^2/12$
h = plate thickness
k = wave number, ω/c
m = plate mass per unit area, $\rho_s h$

INTRODUCTION

After contributing nearly half a century to the journal in the hope of providing some readers with useful information, this author submits one final contribution in the hope of obtaining a physical interpretation of three theoretical results in the area of radiation loading. Even though not challenged in the literature, these results appear either paradoxical or seem surprising coincidences.

I. HIGH-FREQUENCY POWER RADIATED BY POINT-EXCITED, EFFECTIVELY INFINITE ELASTIC PLATES

A familiar result of thin plate analysis is that the *in vacuo* point impedance of effectively infinite thin plates in units of force over velocity is purely resistive and frequency-independent:[1]

$$Z_v = R_v = 8(mD)^{1/2} = \frac{4}{3^{1/2}}\rho_s c_p h^2. \tag{1}$$

The corresponding time-averaged power input by a concentrated force is

$$P_v = \frac{F^2}{2R_v} = \frac{F^2}{16(mD)^{1/2}}. \tag{2}$$

The power radiated by this plate when the plate is exposed to fluid on one side was computed by Gutin.[2] Specifically, in Table I of his paper, he tabulates the power radiated by a point-excited thin plate exposed to fluid on one side, normalized to the structure-borne power in Eq. (2) above. His numerical results are, respectively, for an aluminum and a steel plate radiation loaded by water. He notes that this power ratio tends asymptotically to unity as frequencies increase above the plate's coincidence frequency, viz., that the acoustic power radiated into the fluid medium tends to the structure-borne power injected by the point force into the plate *in vacuo*. One might question whether rotary inertia and shear deformations do not invalidate Gutin's thin plate theory at these frequencies. However, since this asymptotic trend depends on frequency normalized to coincidence frequency, it applies to a fluid medium where the sound velocity is low, e.g., to the atmosphere, and where consequently the coincidence frequency is small compared to the plate's fundamental thickness resonance frequency. For these situations, the asymptotic limit falls in the low-frequency range where thin plate theory is adequate. In the frequency range above the coincidence frequency flexural waves are supersonic. Consequently one does indeed anticipate that all the propagating structure-borne sound is eventually radiated into the fluid. However, in view of radiation loading, one anticipates intuitively that the drive-point impedance exceeds the *in vacuo* impedance in Eq. (1) thereby reducing the power input, Eq. (2). This, however, is not the case, because the ratio of the point impedance of the plate *in vacuo* normalized to the impedance of the radiation loaded plate rises asymptotically to unity as the ratio of frequency to coincidence frequency increases (see Gutin's Table II, which tabulates this ratio). In fact, even when the frequency equals only half the coincidence frequency, this impedance ratio has already risen to an absolute value of 0.96. The reactance has dropped to 9.5% of the impedance amplitude, a result consistent with the trend toward the purely resistive nature of the *in vacuo* impedance, Eq. (1).

Since fluid parameters do not bear on the structure-borne power input *in vacuo*, Eq. (2) above, Gutin's asymptotic result implies that the power radiated by the plate exposed to a low impedance fluid, viz., a gas equals the power radiated by a plate exposed to a high-impedance fluid, viz., a liquid. This is contrary to this writer's intuition. Admittedly, Gutin's results, being based on thin-plate theory, viz., on a low coincidence frequency, imply that the high-impedance fluid owes this property to its large density rather than to a high bulk modulus. Such properties are found, e.g., in water containing gas bubbles.

II. LOW-FREQUENCY IMPEDANCE OF INFINITE RADIATION-LOADED PLATES DRIVEN RESPECTIVELY BY LINE- AND POINT-FORCES

Crighton derives the low-frequency admittances of radiation-loaded plates driven by a line force [Ref. 3, Eq. (21)] and by a point force [Ref. 4, Eq. (3.3)]. It is recalled that the line force result had been derived by Nayak.[5] Crighton's results are formulated for a frequency dependence of the form $\exp(-i\omega t)$. It is noted that Fahy, in his text book,[6] tabulates these results in a misleading manner, in that the imaginary components have the same sign as in Crighton's paper, even though Fahy assumes a time dependence of the form $\exp(j\omega t)$.

The reader may find it convenient to consult Crighton's more recent, comprehensive Structural Acoustics paper[7] based on his 1988 Rayleigh lecture, where both the line drive admittance [Eq. (2.9)] and point drive admittance [Eq. (2.10)] are presented. Dropping, for the sake of simplicity, the real factors multiplying these results, Smith[8] expresses the low-frequency admittances as being proportional to

$$Z^{-1} \propto 1 - i \tan\frac{\pi}{10} \quad \text{for the point-excited plate,} \tag{3}$$

$$\propto 1 + i \tan\frac{\pi}{10} \quad \text{for the line-excited plate.} \tag{4}$$

The corresponding impedances in polar form are

$$Z = |Z| e^{i\theta}, \quad \theta = \tan^{-1}\frac{X}{R}, \tag{5}$$

where R and X are, respectively, the resistance and the reactance. These impedance components, as well as the impedance magnitude $|Z|$ differ for the point and line-driven excitations. The factors multiplying the expressions in Eqs. (3) and (4) being real, one concludes that

$$\theta = \frac{\pi}{10} \quad \text{for the point-excited plate,} \tag{6}$$

$$\theta = -\frac{\pi}{10} \quad \text{for the line-excited plate.} \tag{7}$$

With Crighton's time dependence convention $\exp(-i\omega t)$, the former identifies a springlike reactance component, the latter is a masslike reactance. Recalling that the point-excited plate displays a purely resistive impedance *in vacuo*, Crighton expressed surprise (Ref. 4, p. 391) that radiation loading introduces a springlike reactance. Considering the effect of radiation loading on the structural wave number spectrum, Smith[8] proposes a convincing explanation of this admittedly surprising result, an explanation accepted by Crighton in his Rayleigh lecture (Ref. 7, p. 5). However, neither author sheds light on the equally surprising conclusion whereby the expressions in Eqs. (3) and (4) display a phase angle of identical magnitude for two fundamentally different physical situations.

III. LOW-FREQUENCY ENTRAINED MASS OF CIRCULAR PISTON RADIATORS

The entrained mass acting on an unbaffled, flat circular piston, submerged in an acoustic medium whose diameter is small in terms of acoustic wavelengths equals the entrained mass acting on this same piston when it is baffled and exposed to fluid on one side only.[9,10] Specifically, using the $\exp(-i\omega t)$ convention for time dependence, the reactance in units of force over velocity is

$$X = -\frac{i8}{3} k a^3 \rho c. \tag{8}$$

The corresponding entrained mass is

$$M = \frac{X}{-i\omega} = \frac{8}{3} a^3 \rho.$$

The identity of the entrained mass acting on two radically different sound sources is a surprising result, the small baffled piston being effectively a monopole, while the small free piston radiates as a dipole.

This author is grateful to the reviewer who drew his attention to a similarly simple result derived by Strasberg[11] for the impedance of an unbaffled piston one side of which is vibrating while the other side remains stationary. Modeling the baffled piston as a free piston whose two faces vibrate in opposite directions, elementary physical considerations yield an expression for Strasberg's hybrid piston's impedance acting on one side of respectively the baffled and the unbaffled piston. The long-wavelength impedance being effectively reactive, the hybrid piston's impedance is therefore the half-sum of X and of $X/2$, i.e., $3X/4$, where X is given in Eq. (8), the resulting entrained mass is $2a^3\rho$.

[1] See, for example, M. C. Junger and D. Feit, *Sound, Structures and Their Interaction*, 2nd ed., ASA Publication, 1993; item #0-262-10034; originally published by MIT Press (1986), p. 212.
[2] L. Ya. Gutin, Sov. Phys. Acoust. **10**, 369–371 (1965).
[3] D. G. Crighton, J. Sound Vib. **20**, 209–218 (1972).
[4] D. G. Crighton, J. Sound Vib. **54**, 389–391 (1977).
[5] P. R. Nayak, J. Acoust. Soc. Am. **47**, 191–201 (1970); and **49**, 330 (1971).
[6] F. Fahy, *Sound and Structural Vibration* (Academic, London, 1985), p. 141.
[7] D. Crighton, J. Sound Vib. **133**, 1–27 (1989).
[8] P. W. Smith, Jr., J. Sound Vib. **60**, 213–216 (1978).
[9] A. J. Silbiger, J. Acoust. Soc. Am. **33**, 1515–1592 (1961).
[10] L. L. Beranek, *Acoustics* (McGraw-Hill, New York, 1954), pp. 124 and 126. There is a slight difference in the last digit of the numerical coefficients, presumably as a result of a rounding off error. It is clear from the analysis in Ref. 9 that the two coefficients are identical.
[11] M. Strasberg, J. Acoust. Soc. Am. **34**, 679 (1962).

Experimental observation of pseudocapillary and Rayleigh modes on soft gels

P.-K. Choi, E. Jyounouti, K. Yuuki, and Y. Onodera
Department of Physics, Meiji University, 1-1-1 Higashimita, Tama-ku, Kawasaki 214-8571, Japan

(Received 21 April 1999; accepted for publication 25 May 1999)

Surface wave propagation on agarose gels was studied by measuring the velocity of mechanically excited pulsed waves around sol–gel transition temperature. Two independent modes were clearly observed in the gel state. The velocities of the two modes were successfully fitted to the theoretical calculation of pseudocapillary and Rayleigh modes which were predicted by Onodera and Choi.
 [S0001-4966(99)01309-0]

PACS numbers: 43.35.Pt [HEB]

INTRODUCTION

Surface wave propagation on gels shows interesting behavior around the sol to gel transition point. In sols, surface tension dominates the surface waves and dispersive capillary waves are observed. In the course of gelation, elasticity emerges and surface elastic waves become dominant. A crossover from the capillary mode to the elastic mode has been observed in gelatin[1–3] and tungstic acid.[4] Monroy and Langevin[5] reported the coexistence of the capillary and elastic modes in the crossover region in agarose gels, and explained their results by the hydrodynamic theory of Harden *et al.*[6] Recently, Onodera and Choi[7] closely analyzed the dispersion relation and pointed out that two modes of pseudocapillary waves and surface elastic waves (Rayleigh waves) can propagate on gels. The pseudocapillary mode propagates with the phase velocity twice as large as that of the Rayleigh mode, radiating shear waves into the bulk of the medium. This phase velocity should be affected by shear elasticity, in contrast to the velocity of the capillary mode that was observed by Monroy and Langevin, which is determined only by the surface tension.

This paper describes a direct experimental observation of the two surface modes proposed by Onodera and Choi. Short-pulsed surface waves are mechanically excited on agarose gels, and temporal profiles of the two surface modes are separately detected at a distance. The phase velocities measured as a function of temperature are found to agree with the theoretical prediction.

I. EXPERIMENT

The experimental system used was described elsewhere.[2] Surface waves were excited on gels with a piezoelectric (PZT) bimorph transducer having a width of 10 mm. Pulsed surface waves were measured using a laser beam deflection from a sample surface. A position sensitive photodiode detects a deflection angle which gives a waveform proportional to the surface gradient. The waveform is stored in a digital oscilloscope. The phase shift of the detected waveform as a function of propagation distance gives the velocity of the surface waves. A one-cycle signal of 500 Hz was applied to the transducer so as to gain a high resolution in time. In our previous works, burst waves of ten cycles were used, and accordingly we could not isolate the two modes which would possibly overlap.

Agarose solutions with the concentration of 0.5 wt% were prepared by dissolving the powder (purchased from Oncor) in deionized water at 90 °C. The sol–gel transition temperature was about 38 °C. The experiment was done with decreasing temperature from 41.7 °C, at which the agarose is in a sol state, to 18.2 °C, at a rate of 10 °C/hr. The temperature of the sample was controlled by circulating thermoregulated water in a metal jacket which surrounded the sample. The temperature was monitored by immersing a thermocouple in the sample.

II. RESULTS AND DISCUSSION

Typical oscilloscope tracings observed at a propagation distance of 8 mm from the transducer are shown in Fig. 1. Figure 1(a) represents a waveform obtained at 35.0 °C, at which temperature the gelation process has already started but surface tension rather than elasticity dominates the restoring force of the surface waves. Only one surface mode was observed at this temperature, and this mode is named Wave C. We noticed that a maximum peak of Wave C transfers to a neighboring peak one after another when changing temperature. This indicates that the phase velocity differs from the group velocity; in other words, that Wave C is dispersive. As the temperature decreases, the amplitude of Wave C decreases and a new mode is revealed behind Wave C in the time axis at temperatures below about 30 °C. Figure 1(b) shows the two surface modes obtained at 27.5 °C. We names the new mode Wave R, as indicated in Fig. 1(b). As the temperature decreases further, Wave C disappears at about 25 °C and the amplitude of Wave R increases rapidly. Figure 1(c) shows Wave R obtained at 18.2 °C. The amplitude is about four times larger than that at 27.5 °C. We noticed no waveform change of Wave R, except the amplitude, when changing temperature.

Thus the temperature ranges investigated are divided into three on the basis of the observed types of surface modes. In the range 30–41.7 °C, only Wave C was observed. In the range 25.5–30 °C, where surface tension competes with shear elasticity, Waves C and R were observed. In the

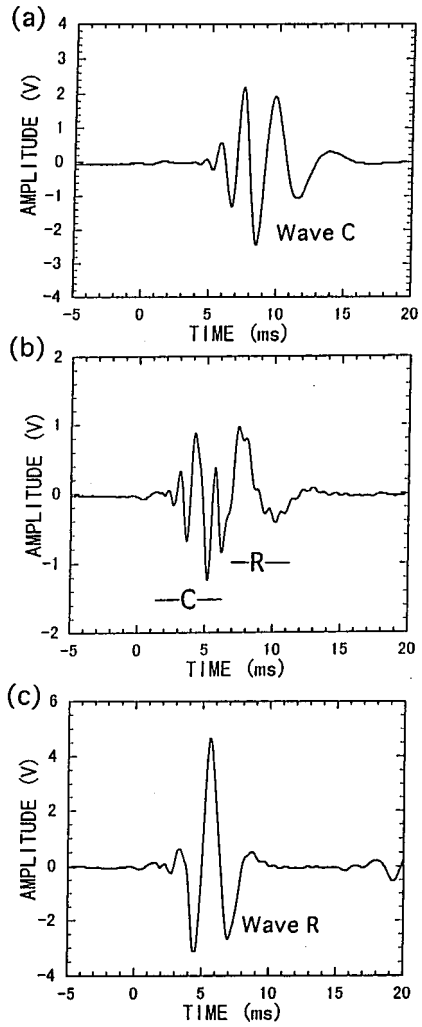


FIG. 1. Oscilloscope tracings of surface waves on agarose gels. (a) Only Wave C was observed at 35.0 °C. (b) Wave C and Wave R were observed at 27.5 °C where surface tension competes with shear elasticity. (c) Only Wave R was observed at 18.2 °C where shear elasticity exceeds surface tension greatly.

range 18.2–25.5 °C, where shear elasticity exceeds surface tension greatly, only Wave R was observed.

The phase velocity measured is shown in Fig. 2. The solid and open circles denote the phase velocities of Wave C and Wave R, respectively. The velocities at 41.7 and 18.2 °C were obtained from the dependence of the time position of a specific peak on propagation distance at constant temperature, and the velocities at other temperatures were obtained from relative measurements of the time position of the specific peak. At temperatures above 38 °C the velocity of Wave C is almost constant, indicating that the surface mode in this range has a pure capillary nature. The velocity increases with decreasing temperature below 38 °C, the sol–gel transition temperature. The increase in velocity results from the emergence of shear elasticity due to gelation.[2] The velocity of Wave R, as indicated by the open circles, also increases with decreasing temperature, though the degree of the increase is small compared with that of Wave C.

Onodera and Choi[7] predicted that two surface modes, the pseudocapillary mode and the Rayleigh mode, can propagate on gels. The pseudocapillary mode is not localized on

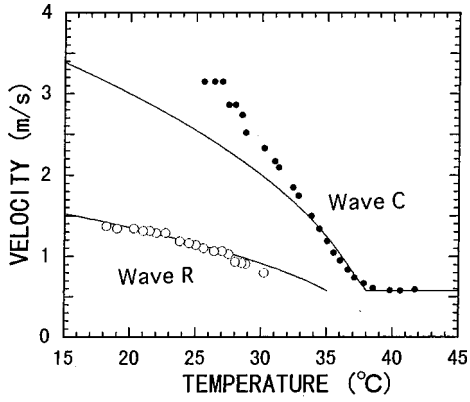


FIG. 2. The phase velocity as a function of temperature for Wave C (the solid circles) and Wave R (the open circles). The solid lines represent the theoretical curves calculated on the basis of Onodera and Choi's theory. Wave C and Wave R were successfully assigned to the pseudocapillary mode and Rayleigh mode, respectively.

the surface, and radiates shear waves into the bulk of the medium. The velocity of this mode is continuously connected to that of the pure capillary mode in sols, and is approximately given by the following relation:[7]

$$v_C = \sqrt{(\gamma k + 4G)/\rho}. \tag{1}$$

Here γ is surface tension, k is wavenumber, G is shear elasticity, and ρ is density. The Rayleigh mode emerges when shear elasticity develops and exceeds the critical value $(\rho\gamma^2\omega^2)^{1/3}$, where ω is angular frequency. The velocity v_R of the emerging Rayleigh mode is equal to the shear wave velocity $v_s(=\sqrt{G/\rho})$. As the gelation proceeds, it approaches the well-known Rayleigh wave velocity $0.955v_s$ with the Poisson ratio 0.5.

According to this theoretical consideration, it is reasonable to assign the observed Wave C and Wave R to the pseudocapillary mode and Rayleigh mode, respectively. We have fitted the theoretical curve of the Rayleigh mode velocity to the experimental points, and obtained the values of shear elasticity. Then we have calculated the velocity of the pseudocapillary mode using the values of the shear elasticity obtained. The calculated curves are represented by the solid lines in Fig. 2. The agreement with the experiment is fairly good. In the calculation we used the exact dispersion relation given by Eqs. (17) and (18) in Ref. 7 instead of the approximate one given by Eq. (1). The values of shear elasticity obtained from the fitting are

$$G = 1.1\times 10^2 (T_0 - T)[\mathrm{Pa}], \tag{2}$$

where T is the temperature in °C and $T_0 = 38$ °C. The parameters used are $\rho = 1.0\times 10^3$ kg/m^3, $\gamma = 6.0\times 10^{-2}$ N/m, and $\omega = 2\pi\times 500$ s^{-1}. The experimental data of Wave C below 33 °C deviate from the theoretical curve. There may be two reasons for the deviation. The amplitude of Wave C decreases while that of Wave R increases with decreasing temperature. The interference between these modes causes a phase shift of Wave C. Second, the central frequency of the spectrum of Wave C may be changed, which resulted in the velocity change because Wave C is dispersive.

The pseudocapillary mode should show a dispersion as indicated by Eq. (1). This is certainly what we noticed from the waveform change in the temperature range of 30–41.7 °C. On the other hand, the Rayleigh mode should show little dispersion, which agreed with the experimental observation in the range 18.2–25.5 °C.

Monroy and Langevin[5] observed the coexistence of capillary and Rayleigh modes in 0.2% agarose gels by measuring spatial profiles of excited surface waves. A striking contrast between their study and the present one lies in the interpretation of the capillary mode. Previous studies,[6,8,9] including Monroy and Langevin's, consider that the capillary mode on gels has fundamentally the same character as that on sols, and that the velocity of the capillary mode is given by $\sqrt{\gamma k/\rho}$. The present study demonstrated that the velocity of the capillary mode is determined by not only the surface tension but also the shear elasticity, as given approximately by Eq. (1). This mode is not localized on the surface and radiates shear waves with a canted-wave vector into the bulk of the medium. Therefore the term ''pseudocapillary'' or ''leaky-capillary'' mode is appropriate.

We have detected another wave pulse at the time position of 19 ms in Fig. 1(c). It is possible that this wave pulse is due to the shear waves that are radiated from the pseudocapillary mode and are reflected back from the bottom of the sample container. This is supported by the observation that the time position of the wave pulse depended on the sample depth while those of Wave C and Wave R did not depend on the depth.

In conclusion, we have demonstrated the existence of two surface-wave modes, the pseudocapillary mode and the Rayleigh mode, in agarose gels. The phase velocities of the two modes were measured using short-pulse excitation, and were found to agree with the prediction by Onodera and Choi. The existence of shear waves which leaked from the pseudocapillary mode has not been confirmed, and should be investigated in the future.

[1] P.-K. Choi, ''Measurements of gel rigidity using surface waves,'' Jpn. J. Appl. Phys., Suppl. **31**, 54–56 (1992).

[2] H. Takahashi and P.-K. Choi, ''Sol-gel transition in gelatin observed with surface waves'' Jpn. J. Appl. Phys. **35**, 2939–2943 (1996).

[3] H. Kikuchi, K. Sakai, and K. Takagi, ''Complex propagation of surface waves on soft gels,'' Phys. Rev. B **49**, 3061–3065 (1994).

[4] H. Okabe, K. Kuboyama, K. Hara, and S. Kai, ''Propagating property of surface waves and viscoelasticity near the gelation point,'' Jpn. J. Appl. Phys. **37**, 2815–2817 (1998).

[5] F. Monroy and D. Langevin, ''Direct experimental observation of the crossover from capillary to elastic surface waves on soft gels,'' Phys. Rev. Lett. **81**, 3167–3170 (1998).

[6] J. L. Harden, H. Pleiner, and P. A. Pincus, ''Hydrodynamic surface modes on concentrated polymer solutions and gels,'' J. Chem. Phys. **94**, 5208–5221 (1991).

[7] Y. Onodera and P.-K. Choi, ''Surface-wave modes on soft gels,'' J. Acoust. Soc. Am. **104**, 3358–3363 (1998).

[8] B. H. Cao, M. W. Kim, and H. Z. Cummins, ''Surface waves on polymer solutions: Complete capillary wave-elastic wave crossover,'' J. Chem. Phys. **102**, 9375–9379 (1995).

[9] C. H. Wang and Q. R. Huang, ''Hydrodynamic surface waves in concentrated polymer solutions in the presence of surface adsorption,'' J. Chem. Phys. **107**, 5898–5906 (1997).

Comments on "A field survey on the annoyance caused by sounds from large firearms and road traffic" [J. Acoust. Soc. Am. 104, 2890–2902 (1998)]

Paul D. Schomer
US Army Construction Engineering Research Laboratory, Champaign, Illinois 61821

(Received 2 November 1998; revised 17 February 1999; accepted 9 June 1999)

Buchta and Vos [J. Acoust. Soc. Am. **104**, 2890–2902 (1998)] have recently analyzed one set of blast-sound attitudinal survey data. From this survey, they conclude that annoyance increases *less* rapidly with level for high-energy impulsive sounds than it does for conventional sounds like road traffic noise, a result that is at odds with a significant number of research results. This paper discusses the analysis by Buchta and Vos and points to several areas in their analysis that might cause their results to depart from the results found by others. Specifically, it appears that the communities in their study exhibit abnormally high levels of annoyance to noise. Also, the great majority of the data that they analyze are below the threshold at which people in the community will notice these sounds. These factors may account for the results found. © *1999 Acoustical Society of America.* [S0001-4966(99)03509-2]

PACS numbers: 43.50.Qp, 43.50.Ba, 43.50.Pn [MRS]

INTRODUCTION

Assessing the noise from high-energy impulsive sounds has been the subject of much research during the last 35 years. High-energy impulsive sounds primarily include military weapons noise from tanks, artillery, bombs and demolition, quarry and mining explosions, and sonic booms. This research has been performed through attitudinal surveys (e.g., Bullen and Hede, 1984; Borsky, 1965), controlled-room laboratory studies (e.g., Broadbent and Robinson, 1964; Leatherwood and Sullivan, 1994), and *in situ* laboratory tests (e.g., Kryter *et al.*, 1967; Schomer, 1994).

This combined literature was recently reviewed by the U.S. National Research Council Committee on Hearing Bioacoustics and Bio-mechanics (CHABA, 1996). The CHABA working group report, after reviewing and analyzing all the data, concluded with two methods to assess high-energy impulsive sound. For one method, the C-weighted sound exposure levels (CSEL) of high-energy impulsive sounds are converted to equivalently annoying A-weighted sound exposure levels (ASEL). The equation used to relate impulsive sound CSEL, L_{CE}, to an equivalently annoying ASEL, L_{AR}, is

$$L_{AR}=103+2(L_{CE}-103)=2L_{CE}-103. \quad (1)$$

In the United States, this procedure using Eq. (1) has been incorporated into an American National Standard—ANSI S12.9, Part 4 (1997).

Equation (1) reflects the general finding of many researchers on this topic—annoyance increases more rapidly with level for high-energy impulsive sounds than it does for conventional sounds like road traffic noise. Some of the researchers finding this conclusion include Borsky (1965), Kryter *et al.* (1967), and Schomer (1994).

The general form to Eq.(1) is

$$L_{AR}=\mathrm{PNSE}+(1/\beta)(L_{CE}-\mathrm{PNSE}), \quad (2)$$

where PNSE is the point of subjective equivalency between the CSEL of a high-energy impulsive sound and the ASEL of an equivalently annoying conventional sound, and $1/\beta$ is the slope to this relation.

In Eq. (1), the slope $1/\beta$ is equal to 2. This factor of 2 for $1/\beta$ quantitatively provides for the conclusion that annoyance grows more rapidly with level to high-energy impulsive noise than to conventional noise. Any value of $1/\beta$ greater than 1 implies that the growth in annoyance with level is greater for high-energy impulsive sounds than is the growth in annoyance with level to conventional sounds. A factor of $1/\beta$ equal to 1 implies that the growths in annoyance with level are equal for high-energy impulsive sounds and conventional sounds, and a factor $1/\beta$ that is less than 1 implies that the growth in annoyance with level is less for high-energy impulsive sounds than it is for conventional sounds.

Buchta and Vos (1998) have recently analyzed one set of attitudinal survey data. From this survey, they conclude that the coefficient in Eq. (1) is less than 1. This paper discusses the analysis by Buchta and Vos and points to several areas in their analysis that might cause their results to depart from the results found by others.

I. DISCUSSION

The Buchta and Vos study included both ''artillery'' noise and road traffic noise. Because their study directly includes these two noise sources, it should provide clear evidence for the relation between blast noise and equivalently annoying conventional noise. However, the Buchta and Vos data and their analysis raise several questions.

The Buchta and Vos paper contains data for the percent of subjects highly annoyed versus artillery noise C-weighted day–night level (CDNL) and road-traffic noise A-weighted day–night level (ADNL). Figure 1 plots the Buchta and Vos data for road traffic compared to the original Schultz (1978) data. Figure 2 shows the data plotted against the recent meta-

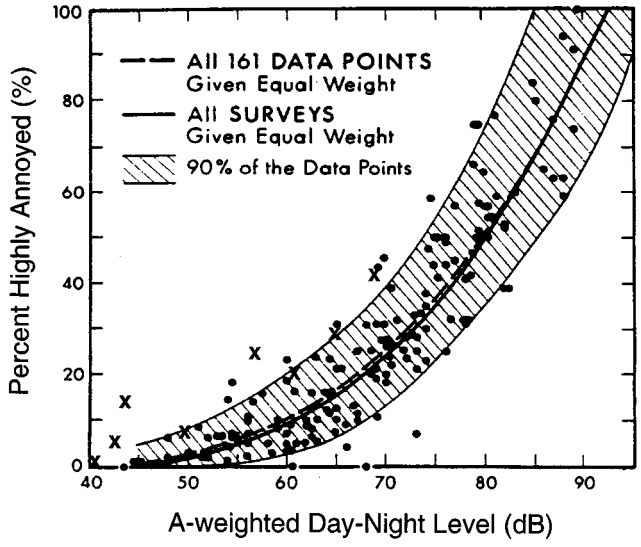


FIG. 1. Buchta and Vos vehicle data (×'s) against the Schultz data. Most of these new data lie outside the indicated 90% prediction interval and many are actually outside the 95% prediction interval.

analysis data by Finegold *et al.* (1994), and Fig. 3 shows the data plotted against the recent curve by Miedema and Vos (1998). All three figures show the same result. Virtually all of the Buchta and Vos data indicate much higher prevalences of annoyance than are typically found. Much of the Buchta and Vos data lie outside the 95% prediction intervals for these other data sets.

There are many possible reasons why the Buchta and Vos data exhibit such large levels of annoyance. First, the physical predictions may be in error. The basic predictions made were for ''downwind'' conditions as required by German law. These ''downwind'' predictions were converted to annual average conditions through ''simplified'' corrections. These simplified corrections may not be sufficiently accurate and/or it is possible that the ''downwind'' noise levels have been underpredicted. Second, for the adjectival scale, the actual adjectives in German may not express the same degree

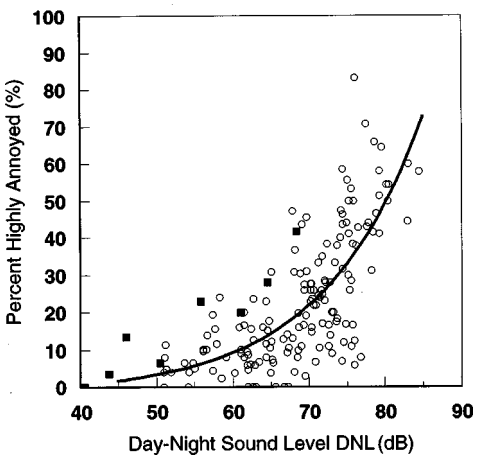


FIG. 2. Buchta and Vos vehicle data (dark squares) against the Finegold data (open diamonds). Most of these new data lie outside the 90% to 95% prediction interval.

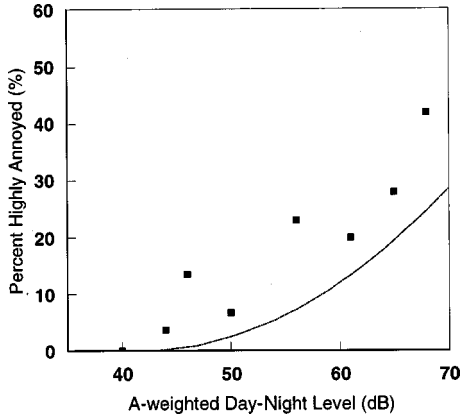


FIG. 3. Buchta and Vos vehicle data (dark squares) against the New Miedema curve for traffic noise (dashed line).

of annoyance as is expressed by similar adjectives in English. Also, for the numerical scale, 33% is higher than the 28% value that is commonly chosen for the threshold of high annoyance (Miedema and Vos, 1998). Third, Fidell *et al.* (1988) has suggested that ''response bias'' may contribute to community responses that differ markedly from the norm. It may be that there are large response biases in this community. As such, this community may not represent typical community responses to noise. In any event, the reported annoyance in this study significantly exceeds the norm for road traffic noise. Buchta and Vos conclude that earlier relations for high-energy impulsive sound ''underestimate annoyance by 12 dB.'' One possible reason for this conclusion may be the generally high levels of response in this community as compared with the norm.

Buchta and Vos devote much of their paper to determining new coefficients for Eq. (2). They perform repeated analyses on their data using various combinations of PNSE and β to find the combination that yields the smallest mean square error between predicted annoyance and the actual measured annoyance.

The data in Table IV of their paper are used for this analysis of PNSE and $1/\beta$ values. This table (Table I herein) contains information as to the distribution of ''single-event CSEL levels'' for all residential areas. The data in Table I raise at least two major questions. The first question regards the small range to the data. Normally, high-energy impulsive sounds exhibit a large scatter to the data. Even in a single day, at 5 km the typical variation from the lowest to highest level is 40 dB. Over an extended time period, this range is 60 dB (Schomer *et al.*, 1978).

The second question raised by Table I concerns the actual levels reported therein. At distances such as 5 km, artillery noise has its spectral peak at about 25 Hz with little energy above about 50 Hz. However, at 25 Hz the threshold of hearing is about 60 dB (ANSI S12.2, 1995). Since most of the energy in an artillery sound arrives during a 250-ms time period, the equivalent level during this $\frac{1}{4}$-s period will be the artillery sound CSEL plus 6 dB. Given the threshold of hearing, artillery sounds for which the CSEL is less than 54 dB will be virtually inaudible. According to Sneddon *et al.* (1996), subjects notice about 50% of events when they ex-

TABLE I. From Buchta and Vos (1998), Table IV. Grouped cumulative frequency distributions (in %) of single-event CSEL for all residential areas. The column headed 77.5 CSEL shows that nearly all of the data were below this level. In fact, except for three sites, about 80% of the remaining data are below this level.

No.	Residential area	47.5	52.5	57.5	62.5	67.5	72.5	77.5	82.5	87.5	92.5	97.5	102.5
1	Eschede	30.3	39.2	65.3	100	100	100	100	100	100	100	100	100
2	Schneverdingen	41.3	61.2	73.4	100	100	100	100	100	100	100	100	100
3	Neuenkirchen	22.4	38.1	65.3	92.0	97.5	99.3	100	100	100	100	100	100
4a	Unterlüss-Nord	0.0	15.6	40.9	53.3	69.6	100	100	100	100	100	100	100
5a	Unterlüss-Mitte	0.0	7.7	35.1	49.2	67.8	100	100	100	100	100	100	100
6a	Unterlüss-Süd	0.0	17.5	43.9	58.2	71.6	100	100	100	100	100	100	100
7	Sülze	0.0	18.6	26.1	55.2	67.7	90.7	100	100	100	100	100	100
8	Soltau	0.0	0.0	0.0	53.0	59.0	75.0	95.5	100	100	100	100	100
9	Bomlitz	0.0	0.0	24.0	63.1	81.5	86.7	97.4	99.1	100	100	100	100
10	Bergen	0.0	0.0	0.0	32.3	50.8	65.5	89.9	97.0	99.9	100	100	100
11	Fallingbostel	0.0	0.0	0.0	35.0	50.7	74.8	86.4	94.2	97.8	99.0	100	100
12	Weitzendorf	0.0	0.0	0.0	0.0	21.3	21.3	27.9	31.7	54.9	99.2	100	100
13	Munster-Nord	0.0	0.0	0.0	0.0	0.0	17.3	21.8	29.6	43.6	97.4	100	100
14	Munster-Sud	0.0	0.0	0.0	0.0	0.0	0.0	0.0	18.3	25.5	45.5	99.4	100
15	Meissendorf	0.0	0.0	6.5	55.5	65.2	76.3	80.3	88.9	93.7	99.9	100	100
16	Westenholz	0.0	0.0	1.1	46.8	55.9	71.0	76.5	87.8	91.1	92.2	99.9	100
17	Dorfmark	0.0	0.0	0.0	11.1	30.0	71.0	91.4	95.0	96.8	99.3	100	100

ceed the background by 11 dB and 90% are noticed when they exceed the background by 20 dB. Less that 10% are noticed when they fail to exceed the ambient by about 8 dB. Therefore, we can expect that people will usually notice blast sounds when they exceed the threshold of hearing by 10–15 dB, when the CSEL is greater than 64–69 dB.

A similar result for blast noise is noted by Cook *et al.* (1994) who states

> ''A lower limit should be set, at which impulses would (for all practical purposes) be regarded as being inaudible. This limit was found to be at 95 dB linear peak sound pressure level. A maximum of 1000 such impulses should be permitted in a day.''

From Schomer *et al.* (1978), the difference between the flat-weighted peak of a blast and the CSEL is about 25 dB. So Cook's limit of 95-dB peak can be converted to about 70-dB CSEL. Similarly, the EPA (1974) levels document states ''...certainly for sonic booms much less than 4.79 to 9.58 pascals (peak pressure), a large percentage of the population is not even expected to sense the boom.'' Eight pascals corresponds to 108-dB peak level, and the difference between peak level and CSEL for sonic booms also is about 25 dB (CHABA, 1981), so this peak level criteria for not sensing booms corresponds to a CSEL of about 83 dB. Hence, the calculation herein that shows subjects will not notice a blast, outdoors, when the CSEL is less than 64 dB is quite conservative. Cook suggests a 70-dB threshold and the EPA levels document suggests over an 80-dB threshold (EPA, 1974).

However, the 64-dB CSEL criteria for noticing blasts is for subjects that are situated outdoors. People spend most of their time indoors. The typical attenuation of artillery sound by a house is about 13 dB (Schomer *et al.*, 1997). Therefore, for an artillery sound to be usually noticed indoors, the CSEL should exceed 77 dB. Examination of Table I shows that about 80% of the data represent levels that are less than 77 dB. It is therefore likely that the analysis by Buchta and Vos includes much data not usually noticed by the respondents.

The Buchta and Vos calculations use free-field blast sound measurements; the ANSI procedure uses pressure-doubled measurements. ANSI requires that the CSEL be measured for a ''pressure double condition'' such as a microphone placed against a hard surface like a wall. Although the theoretical value for the difference between pressure-doubled and free-field blast sound measurements should be near to 6 dB, field measurements have shown that this difference is actually about 4 to 5 dB. Table II compares the Buchta and Vos results with the ANSI results. In this com-

TABLE II. Calculation of equivalent ASEL from CSEL for the β and PNSE indicated using the two Buchta and Vos relations and the ANSI 12.9 Part 4 relation. Pressure-doubled (PD) CSEL are used by ANSI. Buchta and Vos use free-field (FF) measured CSEL. The difference between these is about 4.5 dB.

	Buchta 1	Buchta 2	ANSI	
$1/\beta$	0.79	1.125	2	
β	1.27	0.89	0.5	
PNSE	103	53	103	
CSEL				
PD	FF	Equivalent ASEL		
77	72.5	78.9	74.9	51.0
80	75.5	81.3	78.3	57.0
83	78.5	83.6	81.7	63.0
86	81.5	86.0	85.1	69.0
89	84.5	88.4	88.4	75.0
92	87.5	90.8	91.8	81.0
95	90.5	93.1	95.2	87.0
98	93.5	95.5	98.6	93.0
101	96.5	97.9	101.9	99.0
104	99.5	100.2	105.3	105.0
105	100.5	101.0	106.4	107.0
107	102.5	102.6	108.7	111.0
110	105.5	105.0	112.1	117.0
113	108.5	107.3	115.4	123.0

parison, pressure-doubled levels are used for the ANSI calculations and free-field levels are used for the Buchta and Vos calculations. The free-field CSEL are calculated from the pressure-doubled CSEL by subtracting 4.5 dB.

As noted earlier, Buchta and Vos conclude that the value for $1/\beta$ in Eq. (2) is 0.79 with PNSE equal to 103. They also note that $1/\beta$ may be 1.125 with PNSE equal to 53. As described above, Table II gives results for these two relations between CSEL (free-field levels) and equivalently annoying ASEL. Table II also includes results for the ANSI S12.9 Part 4 relation using pressure-doubled CSEL levels [Eq. (1)]. The data are given for pressure-doubled CSEL between 77 (noticeability) and 113 dB. The value of 113 CSEL is chosen for the high end because this value is known to be a totally unacceptable (prohibited) level for quarry explosions at one per day (Siskind *et al.*, 1980).[1] At the high end where CSEL equals 113 dB, Buchta's two relations predict equivalent ASEL of 107 and 115 dB, respectively. The day-night average sound level (DNL) one would calculate for these two values of ASEL are 57 and 65, respectively. Neither of these is considered to be totally unacceptable. In fact, 57 dB could be considered almost acceptable. In contrast, the equivalent DNL calculated from the ANSI relation is 73 dB—a level that is clearly unacceptable for transportation noise sources in virtually all countries. Similarly, eight sonic booms per day with a peak level of 130 dB are known to be unacceptable. The 130-dB peak level corresponds to a CSEL of about 105 dB. The Buchta relations predict equivalent ASELs of about 101 and 106 dB, respectively. In comparison, the ANSI relation predicts an equivalent ASEL of 107 dB. Eight events per day that each produce an ASEL of 107 calculates to a DNL of 67, which is a level at which sound is considered to be unacceptable. In contrast, eight events per day that each produces an ASEL of 101 dB calculates to a DNL of 60, a level sometimes termed marginally unacceptable. So at this point, Buchta's relation 2 and ANSI agree, but Buchta's relation 1 still appears to be somewhat low.

II. CONCLUSIONS

The high prevalence of annoyance in the Buchta and Vos data, the use of data that are virtually inaudible, and a seemingly small blast noise range all appear to be contributing to a result that may overpredict annoyance at low levels but clearly underpredicts annoyance for high level blasts and booms. In addition, another source of difference is the use of ''free-field'' data rather than ''pressure-doubled'' data.

[1]Siskind actually gives a limit of 133-dB peak with a 2-Hz low-frequency measurement limit. The Siskind report shows that the C-slow level for 107 blasts is, on average, 22.7 dB less than the peak level. For this type of quarry blast event, the duration is about 2 s, so the CSEL is about 3 dB greater than the C-slow level. Hence, the CSEL will be about 20 dB below the peak level, so the unacceptable CSEL is about 113 dB.

ANSI (**1995**). *American National Standard Criteria for Evaluating Room Noise—ANSI S12.2*, American National Standards Institute (ANSI), Table 4.

ANSI (**1997**). *American National Standard Quantities and Procedure for Description and Measurement of Environmental Sound, Part 4: Noise Assessment and Long-term Community Response—ANSI S12.9 Part 4*, American National Standards Institute (ANSI).

Borsky, P. N. (**1965**). ''Community Reactions to Sonic Booms in the Oklahoma City Area,'' National Opinion Research Center, AMRL-TR-65-37.

Broadbent, E. E., and Robinson, D. W. (**1964**). ''Subjective measures of the relative annoyance of simulated sonic bangs and aircraft noise,'' J. Sound Vib. **1**, 162–174.

Buchta, E., and Vos, J. (**1998**). ''A field survey on the annoyance caused by sounds from large firearms and road traffic,'' J. Acoust. Soc. Am. **104**, 2890–2902.

Bullen, R. B., and Hede, A. J. (**1984**). ''Community Response to Impulsive Noise: A Survey Around Holsworthy Army Range,'' Socio-Acoustic Research Section, Australian Government Publishing Service, Canberra.

CHABA (**1981**). *Community Response to High-Energy Impulsive Sounds*, National Research Council, Committee on Hearing Bio-acoustics and Biomechanics (CHABA), edited by W. Galloway (National Academy, Washington, DC).

CHABA (**1996**). *Community Response to High-Energy Impulsive Sounds: An Assessment of the Field Since 1981*, National Research Council, Committee on Hearing Bio-acoustics and Bio-mechanics (CHABA), edited by S. Fidell (National Academy, Washington, DC).

Cook, R., and Job, D. R. S. (**1994**). ''A Criterion for Community Response to Impulse Noise from Large Calibre Weapons and Explosions,'' National Acoustic Laboratories, Commissioned Report No. 79, Chatswood NSW, Australia.

EPA (**1974**). ''Information on Levels of Environmental Noise Requisite to Protect Public Health and Welfare with an Adequate Margin of Safety,'' US Environmental Protection Agency, Office of Noise Abatement and Control, Washington, DC.

Fidell, S., Shultz, T. J., and Green, D. M. (**1988**). ''A theoretical interpretation of the prevalence rate of noise-induced annoyance in residential populations,'' J. Acoust. Soc. Am. **84**, 2109–2113.

Finegold, L. S., Harris, C. S., and von Gierke, H. E. (**1994**). ''Community annoyance and sleep disturbance: Updated criteria for assessment of the impacts of general transportation noise on people,'' Noise Control Eng. J. **42**(1), 25–30.

Kryter, K. D., Johnson, P. J., and Young, J. R. (**1967**). ''Sonic Boom Experiments at Edwards Air Force Base, Annex B: Psychological Experiments on Sonic Booms,'' National Sonic Boom Evaluation Office, USAF Contract AF 49(638)-1758, Stanford Research Institute.

Leatherwood, J. D., and Sullivan, B. M. (**1994**). ''Laboratory study of effects of sonic boom shaping on judged loudness and acceptability,'' Noise Control Eng. J. **42**, 59–69.

Miedema, H. M. E., and Vos, H. (**1998**). ''Exposure-response relationship for transportation noise,'' J. Acoust. Soc. Am. **104**, 3432–3445.

Schomer, P. D., Goff, R. J., and Little, L. M. (**1978**). ''Statistics of amplitude and spectrum of blasts propagated in the atmosphere,'' J. Acoust. Soc. Am. **63**, 1431–1442.

Schomer, P. D. (**1994**). ''New descriptor for high-energy impulsive sounds,'' Noise Control Eng. J. **42**(5), 179–191.

Schomer, P. D., Sias, J. W., and Maglieri, D. (**1997**). ''A comparative study of human response, indoors, to blast noise and sonic booms,'' Noise Control Eng. J. **45**(4), 169–182.

Schultz, T. J. (**1978**). ''Synthesis of social survey on noise annoyance,'' J. Acoust. Soc. Am. **64**, 377–405.

Siskind, D. E. *et al.* (**1980**). ''Structure Response and Damage Produced by Airblast for Surface Mining,'' US Bureau of Mine Report of Investigations, RI 8485, US Department of the Interior.

Sneddon, M., Howe, R., Pearsons, K., and Fidell, S. (**1996**). ''Laboratory Study of the Noticeability and Annoyance on Sounds of Low Signal-to-Noise Ratio,'' NASA Contractor Report 201613, Contract NAS1-20101, Fig. 11.

Response to "Comments on 'A field survey on the annoyance caused by sounds from large firearms and road traffic'" [J. Acoust. Soc. Am. 106, 1594–1597 (1999)]

Edmund Buchta
Institut für Lärmschutz, Düsseldorf, Germany

Joos Vos
TNO Human Factors Research Institute, Soesterberg, The Netherlands

(Received 15 April 1999; accepted for publication 9 June 1999)

Percentages of ''highly annoyed'' respondents were determined by utilizing the more commonly adopted cutoff point at 72% instead of at 63% of the rating scales. The results showed that (1) the community response to the road-traffic sounds was not different from that obtained in many other field surveys, and (2) the overall adjustment to the C-weighted day–night level of the artillery sounds was equal to the previously derived adjustment. The estimated optimal parameter values in Schomer's new rating procedure for high-energy impulsive sounds were hardly different from the previously obtained values. It was shown that analyses in which (inaudible or faintly audible) single events with relatively low CSEL values were excluded yielded essentially the same results as reported earlier. [S0001-4966(99)03609-7]

PACS numbers: 43.50.Qp, 43.50.Ba, 43.50.Pn [MRS]

INTRODUCTION

A. General

As a method of rating high-energy impulsive sounds, Schomer (1994, 1999) recommends application of a level-dependent conversion of the C-weighted sound exposure level (CSEL) of the impulses into the A-weighted sound exposure level (ASEL) of equally annoying vehicle passby sounds. For single events, this rating sound level is equal to $(1/\beta)(L_{CE}-\mathrm{PNSE})+\mathrm{PNSE}$. PNSE represents the point at which the impulse and vehicle sounds with numerically equal levels are also equally annoying.

In spite of the results both from a previous reanalysis of field data collected for residential areas around the artillery ranges of Fort Lewis and Fort Bragg (Vos, 1997), and from a recent analysis of field data collected for residential areas around the military training fields of Bergen and Munster (Buchta and Vos, 1998), Schomer still seems to adhere to the use of his previously recommended equation with $1/\beta=2$ and $\mathrm{PNSE}=103$ dB.

Three issues are relevant here. The first issue is related to the degree in which application of Schomer's procedure results in a proper estimation of the rating sound level (LR) for artillery sounds. The second issue is related to the way in which CSEL is measured or calculated. In many European countries, including Germany and the Netherlands, CSEL is determined in a free-field setting, i.e., level increments caused by facade reflections are explicitly avoided, whereas the American National Standard (ANSI, 1997) seems to require inclusion of such reflections.

With free-field CSELs, application of Schomer's model (with $1/\beta=2$ and $\mathrm{PNSE}=103$ dB) results in an underestimation of LR for artillery sounds of 14 dB in the case of the Fort Bragg and Fort Lewis data (Vos, 1997), and of 18 dB in the case of the Bergen/Munster data (not shown in Buchta and Vos, 1998). With ''double pressure'' CSELs, the underestimation equals 5 and 7 dB in the two respective data sets.

The third issue is related to results obtained in laboratory or quasi-field experiments, showing that for high-energy impulsive single events, the annoyance increases more rapidly with level than the annoyance caused by vehicle sounds (i.e., $1/\beta>1$). The present authors never attempted to disprove or invalidate such results. The point raised both by Buchta and Vos (1998) and by Vos (1997) was that with $\mathrm{PNSE}=103$ dB and $1/\beta=2$, LR was moderately underestimated in the case of ''double pressure'' CSELs and considerably underestimated in the case of free-field CSELs. With free-field CSELs, and satisfying the requirement that $1/\beta$ should be greater than 1, it was shown that a proper LR could be obtained with $\mathrm{PNSE}=63$ dB and $1/\beta$ equal to about 1.2.

B. Specific questions addressed

Schomer (1999) suggested that our previous conclusions about the adequacy of his rating procedure were, among other things, affected by (1) the abnormally high levels of noise annoyance exhibited by the respondents and (2) inclusion in the analyses of many inaudible sound events. Responses to these questions will be given in Secs. I and II, respectively. Although in Buchta and Vos (1998) it was the primary aim to explore the usefulness of Schomer's model for free-field CSELs, some results for ''double pressure'' CSELs will be presented in Sec. II C.

I. DOSE–RESPONSE RELATIONSHIPS

Schomer (1999) noted that the percentages of ''highly annoyed'' respondents reported in Buchta and Vos (1998) were much higher than found in other studies. It will be shown here that the higher percentages were the result of using a relatively low threshold of high annoyance: by uti-

lizing a more commonly adopted cutoff point at 72% instead of at 63% of the rating scales, the community response to the road-traffic sounds in the Bergen/Munster study was not different from that obtained in many other field surveys. Moreover, with the cutoff point at 72%, the adjustment for shooting noise was equal to the previously derived adjustment.

A. Dose–response relationship for shooting sounds

Two questions informed about the degree of annoyance caused by shooting sounds. In one of these questions (Question A1) the respondents had to select their answers from a verbally labeled 5-point rating scale [(1) ''not disturbing,'' (2) ''weakly disturbing,'' (3) ''clearly disturbing,'' (4) ''strongly disturbing,'' and (5) ''very strongly disturbing'']. In the other question (Question A2) a 9-point rating scale with, at the extremes, the labels ''not annoying'' and ''extremely strongly annoying'' was used.

In Buchta and Vos (1998) the subjective effects, as determined with the help of Question A1, had, among other things, been expressed as the percentage of respondents who indicated being ''strongly'' or ''very strongly disturbed.'' In this case the percentages were based on the respondents with scores that exceeded 60% of the rating scale: the adopted cutoff point, x, was equal to 60%. It is more common, however, to determine the percentages of ''highly annoyed'' respondents by utilizing the cutoff point $x=72\%$ (e.g., see the review reported by Miedema and Vos, 1998). The community response resulting from this criterion will be labeled as $\%HA_{72}$. For Question A1, $\%HA_{72}$ is based on the sum of all respondents who had indicated being ''very strongly disturbed,'' plus $(80-72)/(80-60)=40\%$ of the respondents who had indicated being ''strongly disturbed.''

The effects, determined with the help of Question A2, had been expressed as the percentage of respondents who had selected a score $\geqslant 7$, implying adoption of the cutoff point $x=66.6\%$. For Question A2, $\%HA_{72}$ is based on the sum of all respondents who had selected a score of 8 or 9, plus $(77.7-72)/(77.7-66.6)=51\%$ of the respondents who had selected a score of 7.

The data were stratified in exactly the same way as in Buchta and Vos (1998). Figure 1 shows the mean percentages (y) obtained in the various class intervals as a function of CDNL. The equation of the cumulative distribution function (dashed line) fitted to the data in Fig. 1 is given by $y=\Phi[(L_{\mathrm{Cdn}}-78.0)/23.2]$. With $x=72\%$, the mean percentages were 2–12 percent points lower than the previously determined mean percentages (column 7 in Table II given in Buchta and Vos, 1998). Overall, the difference increased with increasing mean CDNL.

B. Dose–response relationship for road-traffic sounds

On the basis of their corresponding A-weighted day–night levels (ADNLs) of the road-traffic sounds, the annoyance ratings of the respondents were assigned to one of the eight classes that were also adopted in Buchta and Vos (1998). The community response to the road-traffic sounds was determined for two questions, B1 and B2, which were identical to Questions A1 and A2 (Sec. I A) with respect to both the wording of the questions and with respect to the response alternatives.

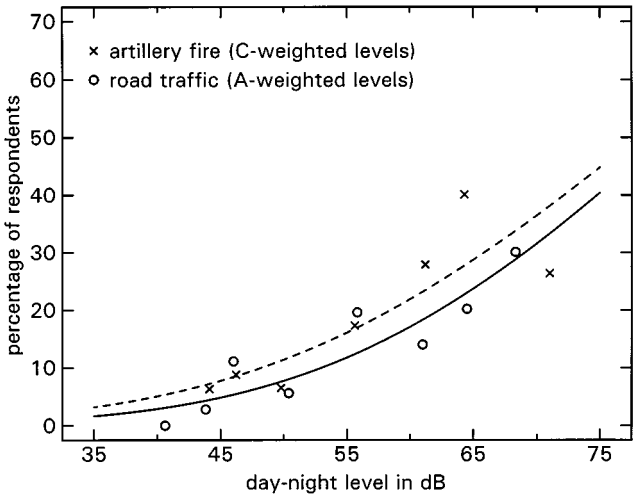


FIG. 1. Percentage of ''highly annoyed'' respondents ($x=72\%$) as a function of the yearly average C-weighted day–night level for the shooting sounds, and as a function of the A-weighted day–night level for the road-traffic sounds. For each sound source, the percentages are based on the responses to two questions. The inserted dashed and solid lines are regression functions for the shooting and road-traffic data, respectively.

As in Sec. I A, the percentages for Questions B1 and B2 were determined with the cutoff point $x=72\%$. The mean percentages are included in Fig. 1. The equation of the function (solid line) fitted to the data in Fig. 1 is given by $y=\Phi[(L_{\mathrm{Adn}}-80.1)/21.2]$. With $x=72\%$, the mean percentages were 0–12 percent points lower than the previously determined percentages (column 7 in Table III from Buchta and Vos, 1998). Overall, the difference increased with mean ADNL.

Figure 2 provides a comparison between the newly de-

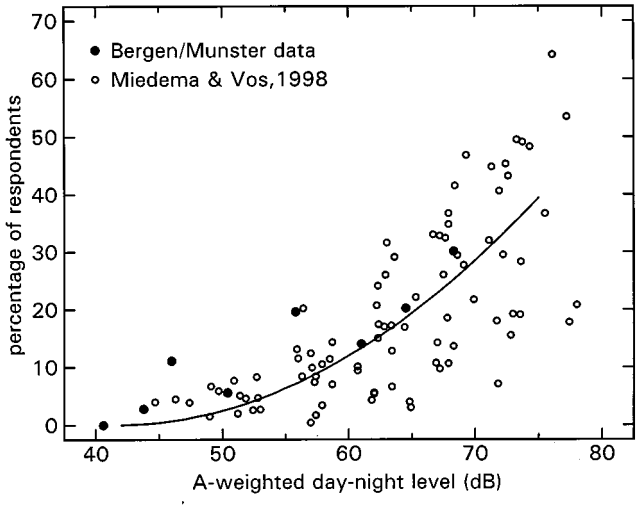


FIG. 2. Percentage of ''highly annoyed'' respondents ($x=72\%$) as a function of the yearly average A-weighted day–night level for the road-traffic sounds (filled dots). As a reference, the dose–response relationship for road-traffic sounds synthesized by Miedema and Vos (1998), as well as all data points used for this synthesis, are given.

termined data points from the present study and the dose–response relationship synthesized by Miedema and Vos (1998). As an indication of the spread in the data, all data used by Miedema and Vos [see their Fig. 2(b)] are reproduced in Fig. 2 as well. Figure 2 shows that six of the eight data points almost coincide with the general dose–response relationship that was based on the data from 26 different field surveys. In fact, there was just one data point (mean $L_{\mathrm{Adn}}=46$ dB, $\%\mathrm{HA}_{72}$ equal to 11) for which the community response was about 7 percent points higher than found in previously investigated residential areas with about the same ADNL. Schomer's hypothesis, that the respondents in the present field survey exhibited abnormally high levels of annoyance to noise, must therefore be rejected.

C. Adjustment for shooting noise

From the two dose–response relationships shown in Fig. 1, it can be concluded that a correction is needed. For a criterion of, for example, 15% HA_{72} respondents, an adjustment of 4.2 dB has to be added to obtain ADNL of equally annoying road-traffic sounds; for a criterion of 30% HA_{72} respondents, the adjustment is 3.1 dB. These results imply that the adjustment is not significantly dependent on the overall level of the shooting sounds, and that application of a 4-dB adjustment would be adequate.

From the two relationships (Fig. 1 in Buchta and Vos, 1998) in which the community response was determined with cutoff points of $x=60\%$ (Questions A1 and B1) and $x=67\%$ (Questions A2 and B2), essentially the same overall adjustment of 4 dB was derived.

II. OPTIMIZATION OF A LEVEL-DEPENDENT CONVERSION PROCEDURE

The conversion of the equivalent-energy-average CSEL of the impulsive sounds into the ASEL of equally annoying vehicle sounds was performed for various combinations of PNSE and $1/\beta$ in exactly the same way as in Schomer (1994) and in Buchta and Vos (1998). As in Buchta and Vos (1998), single events with (free-field) $L_{\mathrm{CE}}\leqslant 50$ dB were excluded from the analysis (Sec. II A). In Sec. II B, the effects of various higher CSEL thresholds of the single events will be reported.

From the ASEL values, the yearly ADNLs were determined for each combination of PNSE and $1/\beta$. Next, the differences between these transformed day–night levels and the original day–night levels of equally annoying road-traffic sounds were calculated for each of 20 conditions separately. The utilized dose-response relationship for road-traffic sound was described in Sec. I B and was shown in Fig. 1.

A. Calculations for single events with (free-field) $L_{\mathrm{CE}}>50$ dB

For all combinations of PNSE and $1/\beta$, the mean differences between the transformed day–night levels for the impulse sounds and the day–night levels for the equally annoying road-traffic sounds were only 1.0–1.2 dB higher than the differences given in Table V from Buchta and Vos (1998). The root-mean-square (rms) values of these differences were typically -0.5 to 1.5 dB higher than found previously.

TABLE I. Values for $1/\beta$ which in combination with the specific PNSE values yielded unbiased fits between the predicted and the originally obtained A-weighted day–night levels of equally annoying road-traffic sounds (mean differences equal to 0 dB). The $1/\beta$-values are given for three threshold conditions: single events were included in the calculations only if their corresponding CSELs exceeded the threshold indicated. Root-mean-squares (rms) of the differences are expressed as decibels.

	CSEL threshold of the single events for inclusion in the calculations					
	$L_{\mathrm{CE}}>50$ dB		$L_{\mathrm{CE}}>65$ dB		$L_{\mathrm{CE}}>70$ dB	
PNSE	$1/\beta$	rms	$1/\beta$	rms	$1/\beta$	rms
53	1.08	8.5	1.10	9.1	1.08	9.1
63	1.12	8.6	1.14	9.2	1.11	9.1
73	1.20	8.8	1.23	9.4	1.18	9.2
83	1.56	10.4	1.60	11.1	1.43	9.8
93	0.70	8.4	0.62	8.8	0.57	9.1
103	0.85	8.3	0.82	8.8	0.82	9.0
113	0.90	8.3	0.88	8.8	0.89	9.0
123	0.93	8.4	0.91	8.8	0.92	9.0
133	0.94	8.4	0.93	8.8	0.94	9.0
143	0.95	8.4	0.94	8.8	0.95	9.0

Again, there were two clusters of $1/\beta$ and PNSE combinations for which the (absolute) mean of the differences was lower than the values obtained for the neutral situation ($1/\beta=1$). The ranges of the PNSE and $1/\beta$ values in these clusters were similar to those found in Buchta and Vos (1998).

For the combination of PNSE=103 dB and $1/\beta=1.5$, which was previously indicated to be an optimal combination for the prediction of LR for artillery sounds (Schomer, 1994), LR for the impulsive sounds in the present study would be underestimated by 10 dB. For high-energy impulsive sounds in general, including artillery fire, demolition, mining explosion, and sonic booms, Schomer recommended utilization of PNSE=103 dB in combination with $1/\beta=2.0$. With these parameter values, and by using free-field CSELs as the input, LR would be underestimated by 16 dB.

In order to find, for the set of relevant PNSE values, those $1/\beta$ values for which the mean differences between the predicted and the obtained day–night levels were equal to 0 dB, additional computations were performed. From the results shown in Table I (section with $L_{\mathrm{CE}}>50$ dB), it can be concluded that such unbiased solutions with reasonably low rms values between 8 and 9 dB were obtained for $53\leqslant\mathrm{PNSE}\leqslant 73$ dB with $1/\beta>1$, and for $93\leqslant\mathrm{PNSE}\leqslant 143$ dB with $1/\beta<1$. An adequate prediction of ADNL of equally annoying road-traffic sound was obtained with, for example, PNSE equal to 103 dB and $1/\beta$ equal to 0.85.

B. The effect of excluding single events with relatively low (free-field) CSEL values

Schomer (1999) suggested that the results of the previous analyses might have been influenced by inclusion of many inaudible events. The present authors agree that this issue deserves further attention. To determine the degree of potential unwanted artificial effects resulting from inclusion of inaudible or faintly audible sound events, the calculations

TABLE II. Mean differences (in dB) between the transformed A-weighted day–night levels for the impulse sounds and the A-weighted day–night levels for the equally annoying road-traffic sounds, determined for combinations in which PNSE equals 103 dB, $1/\beta$ equals 1, 1.5, or 2, and the (free-field) CSEL threshold for inclusion of the single events in the analyses ranges from 50 to 70 dB. The calculations were performed for CSELs determined in free-field conditions, and for ''double pressure'' levels, in which CSELs were increased by 4.5 dB to include facade reflections.

	''Free field''			''Double pressure''		
Threshold	$1/\beta=1$	$1/\beta=1.5$	$1/\beta=2$	$1/\beta=1$	$1/\beta=1.5$	$1/\beta=2$
50	−2.7	−10.4	−16.4	1.8	−3.6	−7.4
55	−2.7	−10.4	−16.4	1.8	−3.6	−7.4
60	−2.8	−10.4	−16.4	1.7	−3.7	−7.4
65	−3.2	−10.7	−16.6	1.3	−3.9	−7.6
70	−2.8	−9.4	−14.4	1.7	−2.7	−5.4

described in Sec. II A will be supplemented for four additional CSEL thresholds, ranging from 55 up to 70 dB. The effects of higher thresholds were not explored: especially in ''downwind'' propagation conditions, with the listeners outdoors, or with the listeners indoors with the windows opened, the impulsive sounds produced by 25–35-mm cannons are clearly audible for CSELs higher than 55 or 60 dB, and the impulsive sounds produced by 155 mm howitzers are clearly audible for CSELs higher than 60 or 65 dB.

Since the community response is determined both by the annoyance experienced indoors and outdoors (e.g., see Birnie *et al.*, 1980; Myncke *et al.*, 1977), it would be unjustified to adopt indoor listening with the windows closed as a sole criterion for audibility.

In columns 2–4 of Table II, it can be seen that for combinations in which PNSE=103 dB and $1/\beta$ equals 1, 1.5, or 2, exclusion of single events with CSELs lower than 55, 60, or 65 dB yielded essentially the same degree of underestimation of LR as found with a threshold of 50 dB. With a relatively high threshold of 70 dB, the underestimation of LR was about 0.5–2 dB lower. Since in the residential areas of Eschede and Schneverdingen the received single event levels were all lower than 70 dB, the calculations with a threshold of 70 dB were based on 18 instead of on 20 conditions.

For the set of relevant PNSE values, the $1/\beta$ values for which the mean differences between the predicted and the obtained day–night levels were equal to 0 dB were computed also. The results for thresholds of 55 and 60 dB were identical to those obtained with a threshold of 50 dB. Table I shows that the effects of excluding single event levels lower than 65 or 70 dB on the optimal $1/\beta$ values and corresponding rms values are negligibly small.

C. Calculations for ''double pressure'' CSELs

The last three columns in Table II summarize results from calculations in which the equivalent-energy-average CELs of the impulsive sounds were increased by 4.5 dB to include facade reflections. For the neutral situation, LR was slightly overestimated by 1.5–2 dB.

For $1/\beta=1.5$ combined with PNSE=103 dB, exclusion of single events with CSELs lower than 50, 55, 60, or 65 dB yielded LRs that were underestimated by about 4 dB. With the relatively high threshold of 70 dB, the underestimation was equal to about 3 dB. For $1/\beta=2$ combined with PNSE =103 dB, exclusion of events with CSELs lower than 50–65 dB yielded LRs that were underestimated by about 7.5 dB. With the threshold of 70 dB, the underestimation was equal to about 5 dB. For the threshold levels between 50 and 70 dB, unbiased solutions were obtained with PNSE=103 dB and $1/\beta$ equal to about 1.15.

III. GENERAL CONCLUSIONS

From the determination of the percentages of ''highly annoyed'' respondents, in which a cutoff point at 72% of the rating scales was used, it was concluded that (1) the community response to the road-traffic sounds was not different from that obtained in many other field surveys, and (2) the overall adjustment to the C-weighted day–night level of the artillery sounds was equal to the previously derived adjustment.

With the help of the newly determined dose–response relationships for the artillery and road-traffic sounds, the parameter values in Schomer's level-dependent conversion procedure were optimized. The results showed that (3) the estimated optimal parameter values in Schomer's rating procedure for high-energy impulsive sounds were hardly different from our previously obtained values, and (4) exclusion of inaudible or faintly audible single events with relatively low CSEL values yielded essentially the same results as reported earlier.

ANSI (**1997**). *American National Standard Quantities and Procedure for Description and Measurement of Environmental Sound, Part 4: Noise Assessment and Long-term Community Response - ANSI S12.9 Part 4* (American National Standards Institute).

Birnie, S. E., Hall, F. L., and Taylor, S. M. (**1980**). ''The contribution of indoor and outdoor effects to annoyance at noise in residential areas,'' in *Proceedings Internoise '80* (Miami, FL), pp. 975–978.

Buchta, E., and Vos, J. (**1998**). ''A field survey on the annoyance caused by sounds from large firearms and road traffic,'' J. Acoust. Soc. Am. **104**, 2890–2902.

Miedema, H. M. E., and Vos, H. (**1998**). ''Exposure-response relationships for transportation noise,'' J. Acoust. Soc. Am. **104**, 3432–3445.

Myncke, H., Cops, A., and Gambart, R. (**1977**). ''Traffic noise measurements in Antwerp and Brussels. Part II: Enquiry concerning annoyance,'' in *Proceedings 9th I.C.A.-Congress* (Madrid), Paper E19.

Schomer, P. D. (**1994**). ''New descriptor for high-energy impulsive sounds,'' Noise Control Eng. J. **42**(5), 179–191.

Schomer, P. D. (**1999**). ''Comments on 'A field survey on the annoyance caused by sounds from large firearms and road traffic' [J. Acoust. Soc. Am. **104**, 2890–2902 (1998)],'' J. Acoust. Soc. Am. **106**, 1594–1597.

Vos, J. (**1997**). ''A re-analysis of the relationship between the results obtained in laboratory and field studies on the annoyance caused by high-energy impulsive sounds,'' Noise Control Eng. J. **45**(3), 123–131.

Wave number frequency (k-ω) analysis using a wavelet transform and an eigenvalue decomposition

John I. Salisbury
Echo Technology, P.O. Box 527, Chepachet, Rhode Island 02814

(Received 3 July 1997; revised 3 May 1999; accepted 12 May 1999)

In a previous paper [Salisbury, J. Acoust. Soc. Am. **96** (1994)], the spacial resolution of a conventional KW (2D FFT) was improved by the use of an eigendecomposition in the spatial domain. In this letter, a new method for generating four-dimensional wave number frequency-time (k-ω-t) vs level plots is presented. It combines eigendecomposition in the space domain and a wavelet transform in the time domain. This technique uses the true broadband continuous Wavelet Transform (WT). This combination generates k-ω plots that are functions of time, and is a significant advance over existing methods for analyzing broadband impulse type and nonstationary signals. The WT-ED technique has wide ranging applicability from diagnosis of acoustic and nonacoustic source generating mechanism to medical imaging technology. [S0001-4966(99)00809-7]

PACS numbers: 43.60.Gk, 43.30.Xm, 43.80.Ev [JCB]

INTRODUCTION

A normal k-ω plot, calculated using a two-dimensional fast Fourier transform (FFT), is a three-dimensional (3D) snapshot in time of the energy being received by a group of N sensors, usually located at equal distances along a line. Energy can be located within the medium surrounding the array (*acoustic*) or within the array structure itself (*non-acoustic*).

The hybrid technique (WT-ED) addressed herein is implemented by taking a one-dimensional continuous wavelet transform (WT) on the sampled time series data recorded by each array sensor and using these results to obtain a frequency dependent cross spectral density matrix (CSD). Eigendecomposition is then applied to this CSD matrix to obtain a set of eigenvalues and eigenvectors.[1]

I. THEORY OF THE WT-ED TECHNIQUE FOR k-ω ANALYSIS

The conventional k-ω takes a FFT in both domains, time and space. By using an FFT, the assumptions are made that the signals are stationary and narrow band. In most real world situations, these assumptions are not met, especially in the time domain and may lead to erroneous conclusions.

To rectify these limitations and make the k-ω analysis more general, a WT is performed on the temporal data of each of the array elements. This transform enables one to process wide-band, nonstationary, dispersive signals as a function of time. The continuous WT is defined as follows, Eq. (1):

$$W_{gf}(s,\tau) = 1/\sqrt{(s)} * \int \left[f(t) g\left(\frac{t-\tau}{s} \right) \right] dt, \tag{1}$$

where $f(t)$ is the received signal, and $g(s,\tau)$ is the mother wavelet with time translation τ and scale (dilation) s.

The WT is a functional dot product between the sensor's signal and another function known as a wavelet which is time compressed, dilated, and shifted. Thus the WT defines an integral transform kernel and, unlike some other transforms, the WT does not possess severe interference or cross terms in the time-frequency plane.[2] In order for $g(t)$ to be an admissible wavelet, it must obey the regularity condition:[3]

$$c_g = \int_{-\infty}^{\infty} \frac{|G(\omega)|^2}{|\omega|} d\omega < \infty,$$

where $G(\omega)$ is the Fourier transform of the Mother Wavelet, $g(t)$. The constant c_g is the admissibility constant of the function, $g(t)$. Any admissible function can be used as a mother wavelet. This allows much leeway in the selection of a mother wavelet. One can be selected to resemble or match characteristics within the received signal in order to enhance classification capabilities or have a large BT to improve resolution.

The spacial cross spectral density function (CSD) of the array, $S(\boldsymbol{\delta},\boldsymbol{\omega},t)$, is obtained by the wavelet transform (WT) method and is given by:

$$S(\boldsymbol{\delta},\boldsymbol{\omega},t) = \langle E_i(x,\boldsymbol{\omega},t) E_i^*(x-\boldsymbol{\delta},\boldsymbol{\omega},t) \rangle, \tag{2}$$

where $E(x,\boldsymbol{\omega},t)$ is the wavelet transform of the entire segment of time series data to be processed, and is given by:

$$E(x,\boldsymbol{\omega},\tau) = 1/\sqrt{(s)} \int_{-\infty}^{\infty} \boldsymbol{\xi}(x,t) g\left(\frac{t-\tau}{s} \right) dt. \tag{3}$$

Operating on each sensor's time domain data results in the time-frequency domain relationship of Eq. (3). We next operate in the spatial domain with an eigendecomposition (ED) technique that obtains the wave number spectrum.[1] This is accomplished by solving for the eigenvalues and eigenvectors of the Hermitian matrix formed from the spatial cross spectral density function (CSD) Eq. (2). $S(t)$ can be represented in matrix form as

$$S(t) = H(\boldsymbol{\delta},\boldsymbol{\omega},t) \boldsymbol{\lambda} H^*(\boldsymbol{\delta},\boldsymbol{\omega},t), \tag{4}$$

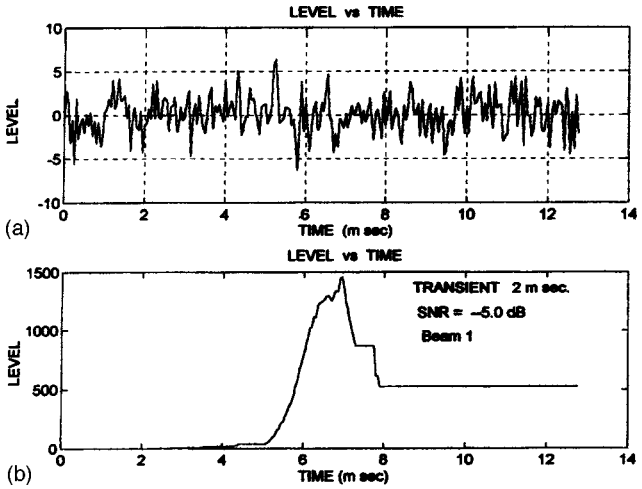


FIG. 1. (a) Time history of acoustic energy present at one array element ($F=3000$ Hz, SNR$=-5.0$ dB). (b) Array processor output level versus time, eigenbeam 1.

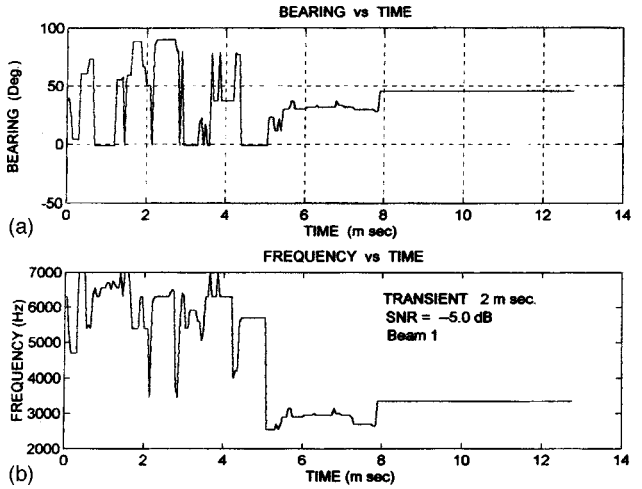


FIG. 2. (a) Target bearing versus time. (b) Target radiate frequency versus time.

where λ_N and H_N are the N eigenvectors and eigenvalues of $S(t)$, and N is the number of array elements. The k-ω spectrum can now be represented as;

$$G(k,\boldsymbol{\omega},t)=\Sigma^* S(t)\Sigma, \tag{5}$$

where $\Sigma=[1,e^{(-jkx)},\ldots,e^{(-jk(N-1))x}]^T$, $k=\boldsymbol{\omega}/c^*\cos(\theta)$ =wave number, $\boldsymbol{\theta}$=spatial direction of energy, x=array elements, constant spacing, $\boldsymbol{\omega}$=angular radian rate$=2\pi f$, c=propagation velocity, either acoustic or nonacoustic. Combining Eq. (4) and Eq. (5) results in

$$G_{p_i}(k,\boldsymbol{\omega},t)=\sum_{l=1}^{p}\lambda_1|\Sigma^* H_{i_1}|^2 \tag{6}$$

or

$$G_{p_i}(k,\boldsymbol{\omega},t)=\sum_{l=1}^{p}\lambda_1(\mathrm{FFT}(H_{i_1}))^2, \tag{7}$$

where $p=1,2,...,N$ and $\boldsymbol{\lambda_1}>\boldsymbol{\lambda_2}>\ldots\boldsymbol{\lambda_N}$. The number of dominate energy sources p can be determined by looking at the N eigenvalues.

The hybrid (WT-ED,t) spectrum is computed by solving Eq. (7). A Blackman–Harris window and zero padding is applied to the pth eigenvector for frequency bin j, time i prior to taking the FFT in order to reduce leakage and to produce a smoother spatial spectrum. In this method, the individual eigenbeam $G_i(k_p,\boldsymbol{\omega}_j)$ is computed for time i and the value at the maximum response axis (MRA) is selected. A k-ω plot for selected eigenvalues can be generated and analyzed without having undesirable energy sources from the other eigenbeams interfering and at the desired time interval. This ability to evaluate each eigenbeam separately and selecting the MRA is what enhances the spatial resolution.

II. DATA ANALYSIS

As test examples, two scenarios are set up. In both cases the Morlet wavelet, Eq. (8), is the Mother wavelet used;

$$g(t)=\exp\left(-\frac{1}{(2\boldsymbol{\alpha}^2)}t^2+j\boldsymbol{\omega}_0 t\right). \tag{8}$$

This complex wavelet allows us to tailor the wavelet to match the signal more accurately by adjusting α and the ω_0. The value of the parameters α, f_0, and the sampling frequency used in both scenarios are 0.0015 s, 25 000 Hz, and 20 000 Hz, respectively. Also similar in both scenarios are the number of samples, 256 (12.75 ms), and a line array composed of 22 equally spaced elements is used as the receiver. Independent wide-band Gaussian noise is added at each element to the generated signal. The frequency range of 1000–7000 Hz, in 50-Hz steps, was performed by the wavelet transform. This frequency sweep is accomplished by specifying the range and increment and at each frequency, f, the scale value s is computed using the relationship $s=f_0/f$. This scale is then used in the wavelet transform.

Scenario 1:

The first case simulates a target generating a 2-ms acoustic transient having an SNR level of -5 dB and propa-

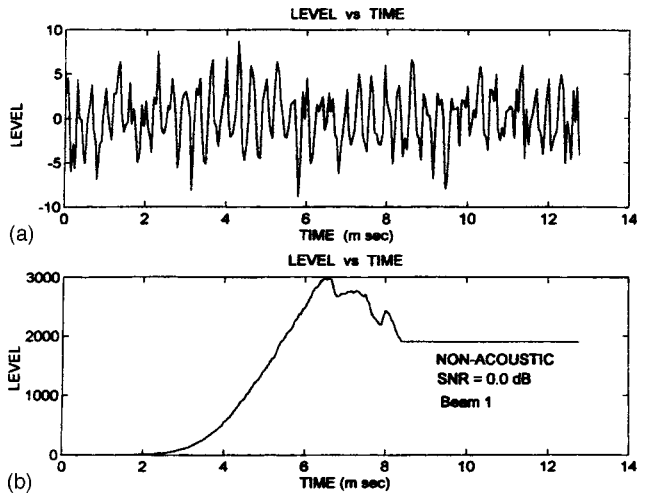


FIG. 3. (a) Time history of nonacoustic energy present at one array element ($F=3000$ Hz, SNR$=0.0$ dB). (b) Array processor output level versus time for eigenbeam 1.

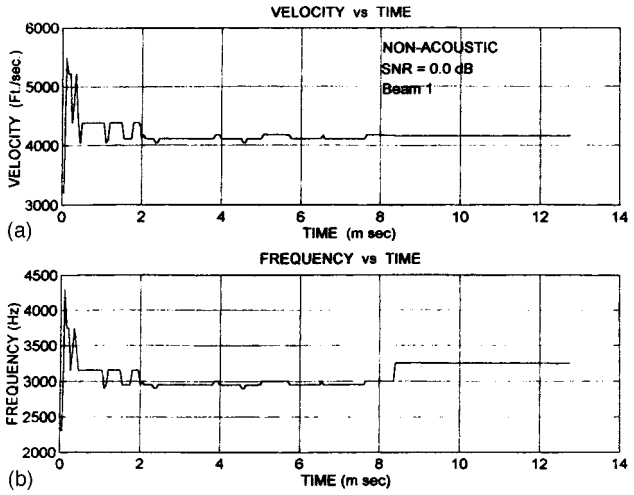


FIG. 4. (a) Propagation velocity versus time of medium, eigenbeam 1. (b) Frequency of nonacoustic interference versus time, eigenbeam 1.

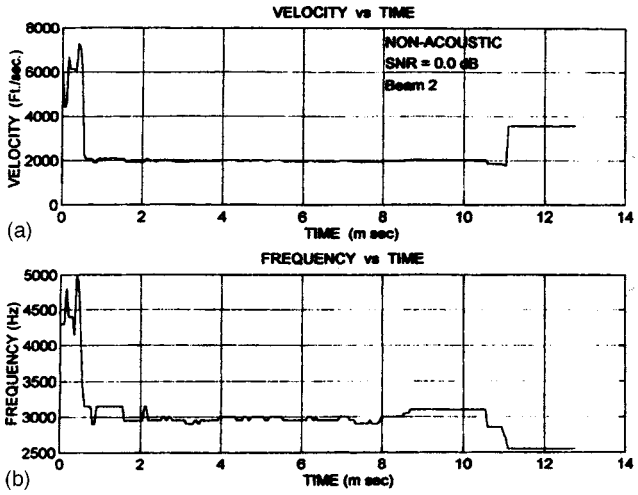


FIG. 5. (a) Propagation velocity versus time of medium, eigenbeam 2. (b) Frequency of nonacoustic interference versus time, eigenbeam 2.

gating within the medium at a velocity of 5000 ft/s. The target is at a bearing of 35 deg. The bearings are relative to the line array, i.e. ($\theta=0.0$, end fire, and $\theta=90$ deg broad side). The frequency of the radiated transient is $f=3000$ Hz. Equation (7) was evaluated at each time increment i. The maximum (k-ω) level at time i is recorded along with the associated k, ω values. The bearing, velocity, and frequency are then computed as a function of time. Figures 1 and 2 depict the results for case 1 using only eigenbeam 1, since all the other beams were random noise. Figure 1(a) shows the received signal in one element. No signal averaging was performed. The transient starts at 5 ms and is 2 ms in duration. It is difficult to detect the transient in the raw data. In Fig. 1(b), the maximum k-ω levels of the WT-ED processor are displayed as a function of time. As can be seen, the transient signal can now be detected at the correct time interval. In Fig. 2(a) and (b), the target bearing and radiated frequency are displayed, 35 degrees and 300 Hz, respectively, over the period at which the transient is present. The WT-ED technique determines the target bearing automatically using the eigenvector which is directed at the energy source and the eigenvalues determine the level of the sources.

Scenario 2:

In case 2, two CW signals are generated at a frequency of 3000 Hz in the nonacoustic region (i.e., $c<5000$ ft/s). They represent interfering energy generated by the array itself. These signals are propagating down the array at a 0 degree bearing, and velocities of 4000 and 2000 ft/s, respectively. The SNR of each CW signal is 0.0 dB. Figures 3, 4, and 5 depict the results of the second scenario. Figure 3(a) shows the generated time signal received at one array element and Fig. 3(b) shows the calculated detection level. Figure 4(a) and (b), respectively, depicts the propagation velocity (4000 ft/s) and frequency (3000 Hz) computed using eigenbeam 1 of the first interference source. Figure 5(a) and (b), respectively, depicts the propagation velocity (2000 ft/s) and frequency (3000 Hz) of the second interference in eigenbeam 2. As can be seen, the WT-ED technique is able to detect signal sources, both transient and CW types, as a function of time and also separate them so they can be analyzed individually to determine their cause.

III. CONCLUSION

The WT-ED processor is capable of generating true wide-band k-ω information as a function of time. This enables us to detect nonstationary, transient signals and determine their spectra and bearings. The processor is also able to determine the propagation velocity on nonacoustic interference and separate the desired signals from noise as well as other types of interfering energy.

[1] J. I. Salisbury, ''Wave-number-frequency (k-ω) analysis using an eigenvalue decomposition technique,'' J. Acoust. Soc. Am. **96**, 838–844 (1994).

[2] F. Hlawatsch and G. F. Boudreaux-Bartels, ''Linear and quadratic time-frequency signal representations,'' IEEE Signal Process. Mag. **April**, 21–67 (1992).

[3] R. K. Young, *Wavelet Theory and its Applications* (Kluwer Academic, New York, 1993).

The effects of differences in burst duration and rise/decay time on auditory apparent motion in the free field

Sherry A. Span
University of Southern California, Department of Psychology, Los Angeles, California 90089-1061

Thomas Z. Strybel
Department of Psychology, California State University Long Beach, Long Beach, California 90840

(Received 29 June 1998; revised 15 December 1998; accepted 19 May 1999)

This study examined the effects of differences in burst duration and rise/decay time between the lead and lag stimuli on auditory apparent motion (AAM: the illusion of motion produced by the proper timing between two stimuli). Three standard burst durations (25, 100, and 400 ms) with 0-ms rise/decay times were each employed with three stimulus onset asynchronies (SOAs) that produce continuous motion for lead and lag stimuli of equal durations. During a trial block, five variable burst durations (25, 50, 100, 200, and 400 ms) and two rise/decay proportions (.00 and .15) were tested. On half the trials the standard burst was presented first. The signals were high-pass noise, and they were presented from two loudspeakers at ±20° azimuth. AAM was heard most often when the duration of the lead and lag stimuli were equivalent. The SOAs that produced AAM depended on the duration of the lead source only. No significant effects of rise/decay proportion were observed.
 [S0001-4966(99)00509-3]

PACS numbers: 43.66.Mk, 43.66.Pn, 43.66.Qp [DWG]

INTRODUCTION

Auditory apparent motion (AAM) is an illusion of motion produced when two sounds are presented in succession. If the timing is appropriate, a single sound is heard to move continuously between the loudspeakers. Timing refers to burst duration and stimulus onset asynchrony (SOA), the onset-onset time difference. Normally the lead and lag stimuli are presented from two loudspeakers separated in space. However, cues to the location of the lead and lag stimuli affect only discrimination of AAM direction, not whether AAM *per se* is heard (for a review of AAM, see Perrott and Strybel, 1997). The timing requirements for AAM are summarized as follows: First, a minimum burst duration of roughly 20–25 ms is required for AAM (Strybel *et al.*, 1992; Strybel and Neale, 1994). Second, the SOAs for motion depend on burst duration. Korte (1915) first described the relationship between burst duration and SOA for visual apparent motion. This relationship is called Korte's second law of apparent motion, and it states that SOA must increase with burst duration, although at a slower rate, for motion to be heard (Kolers, 1972).

Given the importance of SOA and burst duration in AAM, we elaborated on Korte's second law under free-field binaural listening conditions in the present experiment. Specifically, we determined whether differences in the duration of the lead and lag stimuli would affect the timing requirements for AAM. In most previous investigations of AAM, the lead and lag stimuli were identical in all respects. Nevertheless, Strybel and Menges (1998) did find that the timing requirements for AAM were affected by differences in frequency. At frequencies less than 1000 Hz, AAM is heard only when the difference in frequency between the lead and lag stimuli is less than 100 Hz. At high frequencies (e.g., above 2000 Hz), AAM can be heard with greater frequency differences if the SOA is increased. In the current study, we tested AAM with unequal burst durations to determine whether AAM is heard with disparate lead–lag durations and if the effect of burst duration on SOA is caused by the lead stimulus, lag stimulus, or some combination. We also examined differences in the rise/decay time of the stimuli because the rise/decay time would affect the audible portion of each burst and possibly the timing requirements for AAM.

I. METHOD

A. Participants

Four adults were tested, including one author (SS). All of the participants had normal hearing and were experienced with auditory apparent motion experiments.

B. Apparatus

The testing was conducted in a 2.7×3.7×2.1 m audiometric chamber. Each participant sat in a chair in the center of the room, with his or her head fixed by a chin rest. Two Realistic Minimus-7 loudspeakers were mounted 1 m in front of the participant at an elevation corresponding to ear level. The separation between the two speakers was 40°, with the midpoint between them at 0° azimuth. High-pass noise (lower cutoff 2 kHz) was produced by two Quatech WSB-100 waveform synthesizers and WSB-100 stimulus generators mounted in the expansion slots of an IBM PC/AT compatible computer. The output of each synthesizer was led through separate channels of a custom-made amplifier into one of the two speakers in the chamber. The amplitude was set at 55 dB, A-weighted for each burst duration. The computer controlled all timing and stimulus presentations. The participants' responses were collected by a response box attached to the game port of the computer. This box contained three, three-position spring-loaded switches.

C. Procedure

Prior to the actual experiment, all participants completed a practice session. Two 100-ms bursts and five SOAs (5, 45, 70, 110, and 200 ms) were used. Each SOA was presented 16 times in random order for a total of 80 practice trials. For each practice trial, it was randomly determined whether the stimulus originated to the left or the right of the participant. Each participant was instructed to listen to the stimulus sequence and classify the stimulus into one of three categories: *Continuous motion*—a sound starting at one side and moving continuously to the other side; *Broken motion*—a sound starting at one side and moving toward the other side, but the movement was discontinuous, broken, or interrupted; and *No Motion*—one or more nonmoving sounds (e.g., two stationary sounds heard in sequence, two sounds heard simultaneously). Besides classifying his/her perception of the stimuli, each participant also indicated whether the first sound originated on the left or right. This judgment involved determining the apparent direction of the motion on trials that produced apparent motion, or the proper temporal sequence on trials that did not. Each category selection was indicated with an appropriately labeled switch, which the participants threw either to the left or right to indicate the direction of origination. When they could not determine direction, the participants were instructed to guess.

After the practice session, the experiment began. Three standard burst durations (25, 100, and 400 ms) were paired with three SOAs. At a standard duration of 25 ms, the SOAs tested were 20, 25, and 30 ms. At 100 ms, the SOAs tested were 20, 60, and 100 ms, and at 400 ms, 60, 100, and 140 ms were used. The SOAs at 25- and 100-ms standard burst durations were chosen to span the range of SOAs that produce continuous motion when identical burst durations are used for the lead and lag stimuli (Strybel and Neale, 1994). The SOAs for the 400-ms standard burst duration were chosen from research on Korte's second law because AAM has not been previously tested at this duration.[1] Within a session, the standard burst duration and SOA were constant, but the variable burst duration and rise/decay time varied. Five variable burst durations (25, 50, 100, 200, and 400 ms) and two rise/decay proportions (.00 or .15) were tested. Rise/decay proportion was defined as the proportion of the variable burst duration required for signal amplitude to rise from 10% to 90% of the final amplitude, and the time required for signal amplitude to decrease from 90% to 10%. A rise/decay proportion of .15 therefore corresponded to rise decay times of 7.5 ms at 50-ms burst duration, 15 ms at 100-ms burst duration, 30 ms at 200-ms burst duration, and 60 ms at 400-ms burst duration. The rise/decay proportion of the standard stimulus was always .00, or less than 1 ms. Within each session these variables created ten unique pairings between the standard and variable sound stimuli. For each unique pairing, on half the trials the standard stimulus was presented first, and on the other half, the variable stimulus was presented first. Each variable burst duration, rise/decay proportion, and presentation order combination was presented randomly 15 times for a total of 300 trials per session. Three sessions were run at each of the nine standard duration-SOA combinations. The order of standard-SOA combinations was determined randomly for each participant.

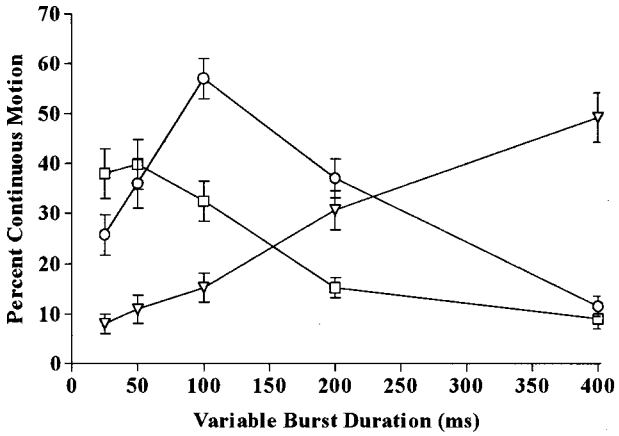


FIG. 1. Percent continuous motion as a function of standard and variable burst durations (squares: standard burst duration=25 ms, circles: standard burst duration=100 ms, triangles: standard burst duration=400 ms).

II. RESULTS

Overall, continuous motion was reported most often at a standard duration of 100 ms (mean=34%) compared with either 25 ms (mean=27%) or 400 ms (mean=23%). For each standard burst duration, a four-way repeated measures analysis of variance (ANOVA) was performed using the factors of variable burst duration, SOA, rise time, and order of presentation. No significant effects of rise/decay proportion were obtained.

Figure 1 illustrates that continuous motion was heard most often with equivalent lead and lag burst durations. In addition, the percentage of continuous motion reports decreased as the disparity between the burst durations increased. The main effect of variable burst duration was significant only at the 100- and 400-ms standards [$F(4,12)=12.23$, $p<0.01$ and $F(4,12)=4.72$, $p<0.05$, respectively]. Nevertheless, a similar trend for the 25-ms standard was present.

The significant variable burst duration-order-SOA interaction observed at the 25- and 100-ms standards [$F(8,24)=2.44$ $p<0.05$ and $F(8,24)=2.78$, $p<0.05$, respectively] depicted in Fig. 2 suggests that the SOAs that produce motion may depend on the duration of the lead stimulus only. At the 25-ms standard, continuous motion was reported most often at variable burst durations between 25 and 100 ms with SOA=20 ms. When the variable burst duration was also 25 ms, continuous motion was not affected by order. Nevertheless, at variable burst durations of 50 and 100 ms, continuous motion was reported more frequently when the standard was the lead stimulus.

At 100-ms standard, motion was reported most often when SOA=60 ms with variable burst durations of 50 and 100 ms. At the 50-ms variable burst duration, motion was heard more frequently when the standard led the stimulus sequence. At variable burst duration=100 ms, order was less important, although it appears that more motion was reported when the standard was the lag stimulus.[2] When the variable burst duration was 200 ms, the optimal SOA for continuous

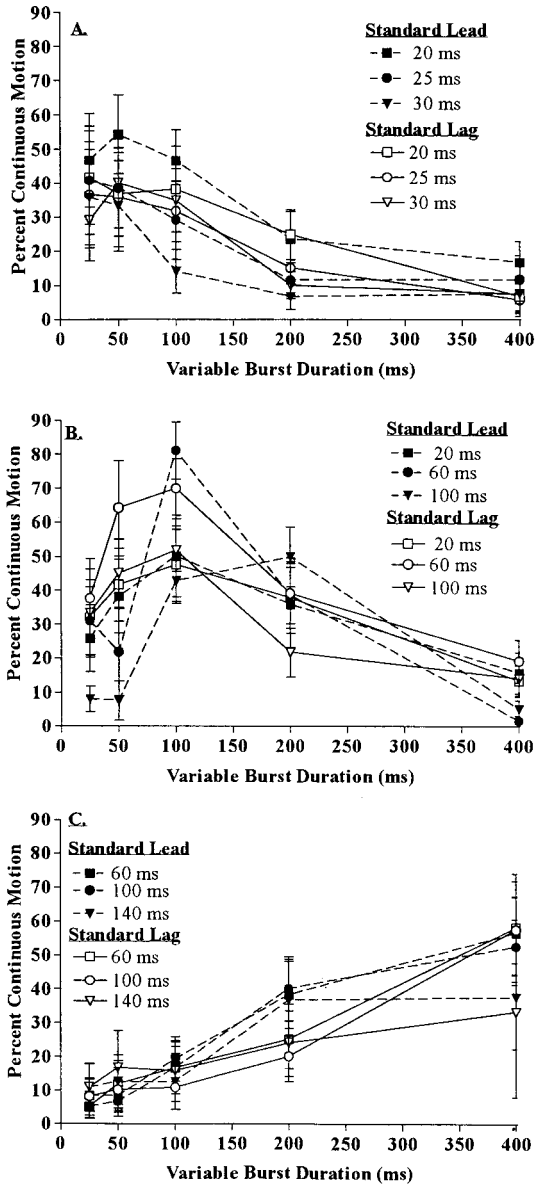


FIG. 2. Percent continuous motion as a function of order of stimulus presentation, variable burst duration and SOA at each standard burst duration tested: (a) standard burst duration=25 ms, (b) standard burst duration=100 ms, and (c) standard burst duration=400 ms.

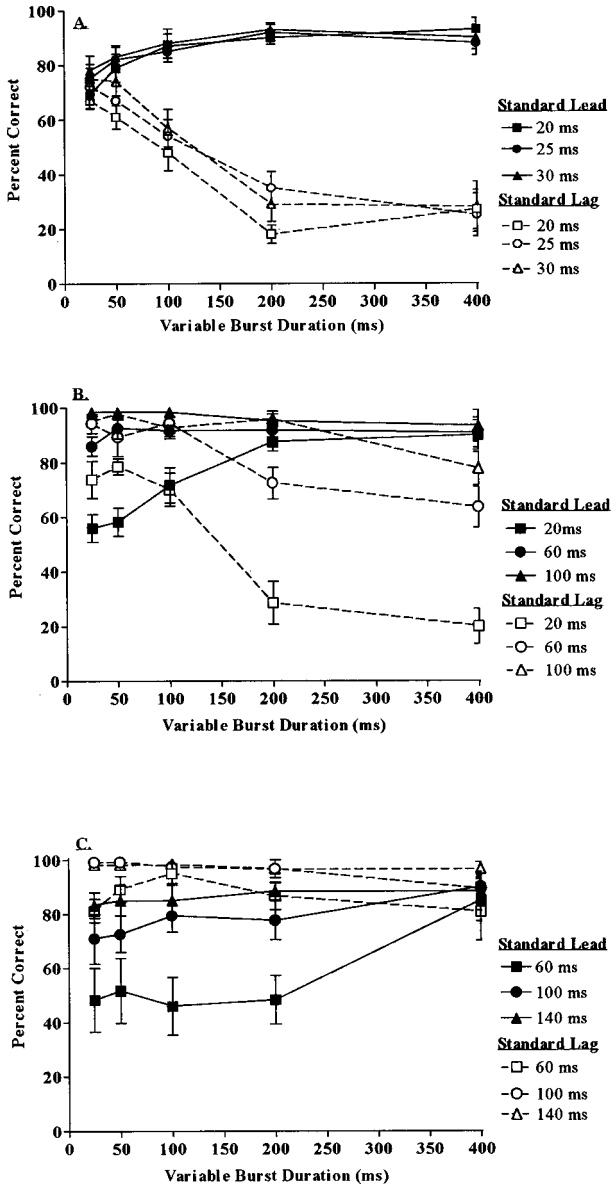


FIG. 3. Percent correct direction reports as a function of variable burst duration, order of presentation and SOA at each standard burst duration tested: (a) standard burst duration=25 ms, (b) standard burst duration=100 ms, and (c) standard burst duration=400 ms.

motion increased to 100 ms, and this SOA was effective in producing motion only when the standard was the lag stimulus. Given that Korte's second law predicts longer SOAs are required for longer burst durations, the interaction of SOA with variable burst duration and order suggests that the lead stimulus alone determines the SOAs that produce AAM.

The significant two-way interaction of variable burst duration by order at the 400-ms standard [$F(4,12)=8.26$, $p<0.01$] is also illustrated in Fig. 2. The order of stimuli mattered most at the variable duration of 200 ms. When both lead and lag stimuli were 400 ms, continuous motion was heard most often at SOAs of 60 and 100 ms, regardless of order. Very little motion was reported at SOA=140 ms.[3]

The direction judgments were also analyzed and are presented in Fig. 3. At a standard duration of 25 ms, a significant burst duration by order interaction was obtained [$F(4,12)=15.86$, $p<0.01$]. Direction discrimination improved with longer variable burst durations only when the standard was the lead stimulus. When the standard was the lag stimulus, performance worsened with increases in variable burst duration. At both 100- and 400-ms standards, significant three-way interactions between duration, order, and SOA were obtained [$F(8,24)=12.89$, $p<0.01$ and $F(8,24)=3.85$, $p<0.01$, respectively]. At 100-ms standard, performance was poorest at SOA=20 ms. At this SOA, discrimination performance improved with increases in variable burst duration when the standard led, but performance worsened when the standard lagged. At 400-ms standard duration, performance was at chance levels only at the shortest SOA (60 ms) when the standard was the lead stimulus. As with continuous motion reports, no significant effects of rise/decay proportion were observed on the percentage of correct direction judgments.

Figure 3 shows that performance on the direction task

was not greatly affected by whether motion was heard. Although performance was usually quite good at the variable burst duration-SOA-order combinations that produced reliable reports of continuous motion, performance was equally good (and sometimes superior) at combinations where little motion was reported. There are no peaks in Fig. 3 that correspond to the continuous motion peaks in Fig. 2. In fact, the differential effects of order and variable burst duration in some conditions suggest that participants may have used different cues to discriminate direction. One possible stationary cue to direction is the duration of the lag stimulus after lead-stimulus offset. For example, at the 25-ms standard–400-ms variable burst combinations, the lag stimulus remained on 395–405 ms after lead-stimulus offset when the standard led. Performance was near-perfect (88%–93%) in these conditions, as shown in Fig. 3(a). When the standard was the lag stimulus, the lag stimulus was offset 345–355 ms *before* lead-stimulus offset, and performance was very poor (25%–28%). These percentages suggest that participants were reliably reporting the direction as opposite the direction of the physical sequence.

III. DISCUSSION

Previous research indicated that a minimum burst duration of roughly 25 ms is required for AAM (Strybel *et al.*, 1992; Strybel and Neale, 1994). Also, as burst duration increases, the SOA must also increase for AAM to be heard (Briggs and Perrott, 1972; Strybel *et al.*, 1992; Strybel and Neale, 1994) The current study expanded on these timing requirements and found that AAM is heard most often when the duration of the lead and lag stimuli are roughly equal. At standards of 100 and 400 ms, the percentage of motion decreased with increasing disparity in burst durations. This rule did not strictly apply at 25 ms because AAM was produced most often with a 25-ms lead stimulus followed by a 50-ms lag stimulus. When the pulses are equivalent at 25 ms, the stimulus is near threshold for detection of AAM (Strybel *et al.*, 1992). Increasing the duration of the lag stimulus may have produced a stimulus sequence that exceeded this threshold value, thus increasing the likelihood that motion is heard. Nevertheless, when the lag burst duration is lengthened further (beyond 100 ms), motion reports declined dramatically.

When the difference in burst durations between the lead and lag stimuli was within the range that produced AAM, the lead stimulus duration determined the necessary SOA. This finding was shown in Fig. 2(a) and (b), where the optimal condition for motion depended on the order of stimuli, SOA, and variable burst duration. For example, when a 200-ms lead stimulus was followed by a 100-ms lag stimulus, motion was heard most often at SOA=100 ms. When the order was reversed, a SOA of 60 ms produced AAM. This three-way interaction was not found at the 400-ms standard, possibly because the SOAs tested were less than optimal for AAM. At 400 ms, motion was heard most often at SOAs=60 and 100 ms, and decreased substantially at SOA=140 ms. Since SOAs of 60 and 100 ms also produced the most AAM at burst duration=100 ms, the upper SOA for AAM may be constant for burst durations above 100 ms. This effect was also found in the binaural data of Strybel and Neale (1994). Note also that the overall percentage of motion reports at burst duration=400 ms (27%) was less than the value at burst duration=25 ms, suggesting an upper burst duration limit on AAM.

The results from the direction discrimination task show that participants can discriminate the direction of motion using cues that are not required for the perception of AAM. These data suggest that participants may have inferred direction from the location of the offset of the stimulus sequence. In fact, when the lag stimulus was offset before the lead stimulus, participants reliably reported the direction as opposite the order of presentation. As in most of our previous investigations of AAM (e.g., Perrott and Strybel, 1997), the effect of stimulus timing on AAM per se was different from the effect of stimulus timing on judgments of its direction.

[1] Strybel and Neale (1994), for example, tested burst durations between 25 and 300 ms. The lower SOA limit for AAM increased with burst duration, but the upper SOA limit, 100 ms, was constant for burst durations above 100 ms. In the present experiment, we chose SOAs above and below 100 ms to determine if the upper SOA limit remained constant.

[2] Order of presentation should not affect the percentage of continuous motion reported with equivalent burst durations. We believe that this effect of order at 100 ms resulted from a marginally significant four-way interaction between variable burst duration, SOA, order, and rise/decay proportion [$F(8,23)=2.13$, $p=0.07$]. At SOA=60 ms, more motion was reported when the rise/decay time of the variable burst duration stimulus was .15, and the standard was the lag stimulus. In other words, when a 100-ms burst with a 15-ms rise/decay period was followed by a 100-ms burst with a 0-ms rise/decay period, motion was reported 94% of the time. When the order was reversed, motion reports averaged only 70%.

[3] Based on a 67% criterion, all participants reported motion reliably at one or more SOAs for standard burst durations of 100 and 400 ms. At 25 ms, only two participants exceeded this value at one or more SOAs, consistent with the idea of a lower burst duration limit.

Briggs, R., and Perrott, D. R. (**1972**). ''Auditory apparent movement under dichotic listening conditions,'' J. Exp. Psychol. **92**, 83–91.

Kolers, P. A. (**1972**). *Aspects of Motion Perception* (Pergamon, Oxford).

Korte, A. (**1915**). ''Kinematoskipische untersuchungen,'' Zeitschrift fur Psychologie **72**, 193–206.

Perrott, D. R., and Strybel, T. Z. (**1997**). ''Some observations regarding motion without direction,'' in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Erlbaum, Hillsdale, NJ), pp. 275–294.

Strybel, T. Z., and Menges, M. L. (**1998**). ''Auditory apparent motion between sine waves differing in frequency,'' Perception **27**, 483–495.

Strybel, T. Z., and Neale, W. (**1994**). ''The effect of burst duration, interstimulus onset interval and loudspeaker arrangement on auditory apparent motion in the free field,'' J. Acoust. Soc. Am. **96**, 3463–3475.

Strybel, T. Z., Witty, A. M., and Perrott, D. R. (**1992**). ''Auditory apparent motion in the free field: The effects of stimulus duration and separation,'' Percept. Psychophys. **52**(2), 139–143.