Journal of Sound and Vibration (2000) **232**(1), 101–127
doi:10.1006/jsvi.1999.2687, available online at http://www.idealibrary.com on IDEAL®

FORMULAE DESCRIBING SUBJECTIVE ATTRIBUTES FOR SOUND FIELDS BASED ON A MODEL OF THE AUDITORY-BRAIN SYSTEM

Y. Ando, H. Sakai and S. Sato

Graduate School of Science and Technology, *Kobe University*, *Kobe* 657-8501, *Japan*

(*Accepted* 30 *June* 1999)

This article reviews the background of a workable model of the auditory–brain system, and formulae of calculating fundamental subjective attributes derived from the model. The model consists of the autocorrelation mechanisms, the interaural cross-correlation mechanism between the two auditory pathways, and the specialization of the human cerebral hemispheres for temporal and spatial factors of the sound field. Typical fundamental attributes, the apparent source width, the missing fundamental, and the speech intelligibility of sound fields, for example, in opera houses, are described in terms of the orthogonal spatial factors extracted from the interaural cross-correlation function, and the orthogonal temporal factors extracted from the autocorrelation function, respectively. Also, other important subjective attributes of the sound fields, the subjective diffuseness, and subjective preferences of both listeners and performers for single reflection are demonstrated here.

1. BACKGROUND OF THE MODEL

The auditory-brain model is based on the following facts found by extensive experimental studies. First of all, the sensitivity of the human ear to sound source in front of the listener is essentially formed by the physical system from the source point to the oval window of the cochlea [1]. By recording left and right auditory brainstem responses (*ABR*) it has been found [2] that (1) amplitudes of waves I and III correspond roughly to the sound pressure level as a function of the horizontal angle of incidence to the listener (ξ); (2) amplitudes of waves II and IV correspond roughly to the sound pressure level as a function of the contra horizontal angle ($-\xi$) implying the interchange of neural information flow between the left and right; and (3) results of analyses of *ABR*s indicate that possible neural activities at the inferior colliculus correspond well to the values of *IACC*. Moreover, it has been discovered by recording left and right slow vertex response (*SVR*) [3, 4] that (4) the left and right amplitudes of the early *SVR*, i.e., $A(P_1 - N_1)$, indicate that the left and right hemispheric dominances are due to the temporal factor (Δt_1) and spatial factors (*LL* and *IACC*) respectively and (5) both left and right latencies of N_2

correspond well to the scale values of subjective preference as a primitive response. It has also been reported [5–7] that (6) from the results of analyzing the autocorrelation function of α-wave in continuous brain wave (*CBW*), the cerebral hemispheric specialization of the temporal factors (Δt_1, and T_{sub}) indicate left hemisphere dominance and the *IACC* indicates right hemisphere dominance. Thus, a high degree of independence between the left and right hemispheric factors may be achieved. Based on such physiological responses, a model of the auditory–brain system is proposed (see Figure 1). This model consists of the autocorrelation mechanisms, the interaural cross-correlation mechanism between the two auditory pathways, and the specialization of human cerebral hemispheres for temporal and spatial factors of sound fields.

2. MODEL

As shown in Figure 1 [1, 8], the sound source $p(t)$ is located at a point r_0 in a three-dimensional space and a listener sitting at r is defined by the location of the center of the head, $h_{l,r}(r|r_0, t)$ being the impulse responses between r_0 and the left and right ear-canal entrances. The impulse responses of the external ear canal and the bone chain are respectively $e_{l,r}(t)$ and $c_{l,r}(t)$. The velocities of the basilar membrane are expressed by $V_{l,r}(x, \omega)$, x being the position along the membrane. The action potentials from the hair cells are conducted and transmitted to the cochlear nuclei, the superior olivary complex (including the medial superior olive, the lateral superior olive, and the trapezoid body), and the higher level of the two cerebral hemispheres.

The input power density spectrum of the cochlea $I(x')$ can be roughly mapped, according to the tuning of a single fiber [9], at a certain nerve position x'. This fact

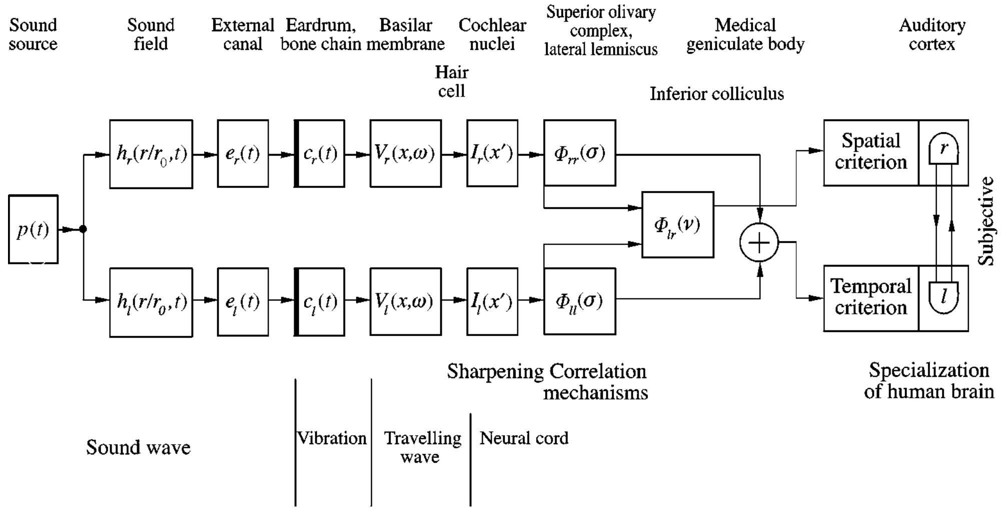

Figure 1. An auditory–brain model with the autocorrelation mechanisms, the interaural cross-correlation mechanism, and the specialization of human brain related to the spatial and temporal factors for subjective responses.

may be, at least, partially related supported by ABR waves (I–IV) which reflect the sound pressure levels as a function of the horizontal angle of incidence to a listener. Such neural activities, in turn, include sufficient information to attain the autocorrelation function (ACF) at a higher level, probably near the lateral lemniscus, as indicated by $\Phi_{ll}(\sigma)$ and $\Phi_{rr}(\sigma)$. For convenience, the interchange of neural signals is not included here. As discussed in reference [2], the neural activity (wave V) may correspond to the $IACC$. Thus, the interaural cross-correlation mechanism may exist at the inferior colliculus. It is concluded that the output signal to the interaural crosscorrelation mechanism including the $IACC$ and the loci of maxima may be dominantly connected to the right hemisphere. The sound pressure level may be expressed by a geometrical average of ACFs for the two ears at the time of origin ($\sigma = 0$) which, in fact appearing in the latency at the interior colliculus, may be processed in the right hemisphere. Effects of the initial time delay gap between the direct sound and the single reflection Δt_1 included in the autocorrelation function may activate the left hemisphere. Such specialization of the human cerebral hemisphere may be related to the highly independent contributions of spatial and temporal criteria on subjective attributes. It is notable that "cocktail party effects", for example, may be well explained by such specialization of the human brain because speech is processed in the left hemisphere and the spatial information is mainly processed in the right hemisphere.

3. FUNDAMENTAL SUBJECTIVE ATTRIBUTES IN RELATION TO THE INTERAURAL CROSS-CORRELATION MECHANISM

3.1. SUBJECTIVE DIFFUSENESS IN RELATION TO IACC

The interaural cross-correlation function is a significant factor in determining the perceived horizontal direction of the sound and the degree of subjective diffuseness of the sound field [10]. A well-defined direction is perceived when the normalized interaural cross-correlation function has one sharp maximum, a high value of the $IACC$, and a narrow values of the W_{IACC} as defined in Figure 2. On the other hand, subjective diffuseness or no spatial directional impression corresponds to a low value of $IACC$ ($< 0{\cdot}15$). The subjective diffuseness or spatial impression of the sound field in a room is one of the important attributes is describing good acoustics. If the sounds arriving at the two ears are dissimilar ($IACC = 0$), then different signals (but signals containing the same information) are conveyed through the two independent auditory channels to the brain. This condition, in turn, improves speech clarity [11].

The scale value of subjective diffuseness was obtained in paired-comparison tests with bandpass Gaussian noise by varying the horizontal angle of two symmetric reflections [12, 13]. Listeners judged which of two sound fields were perceived as the more diffuse. A notable finding is that the scale values of subjective diffuseness are inversely proportional to the $IACC$, as shown in Figure 3, and may be formulated in terms of the 3/2 power of the $IACC$ in a manner similar to that which

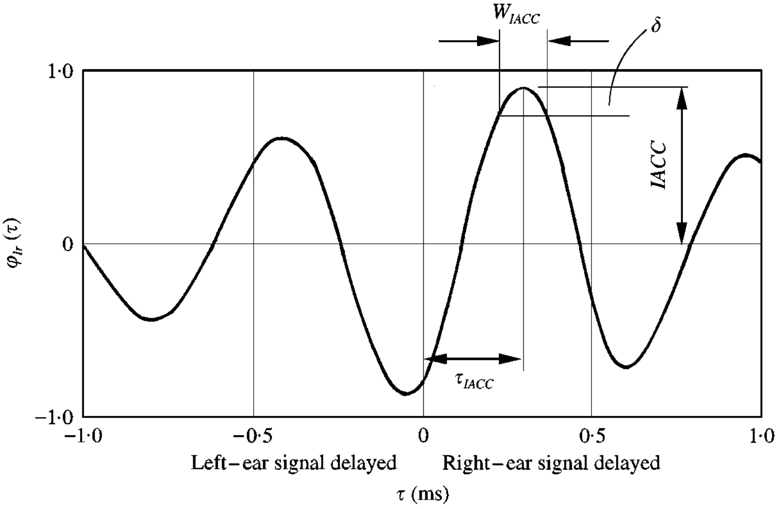

Figure 2. An interaural cross-correlation function, and definition of the $IACC$, W_{IACC} and τ_{IACC}.

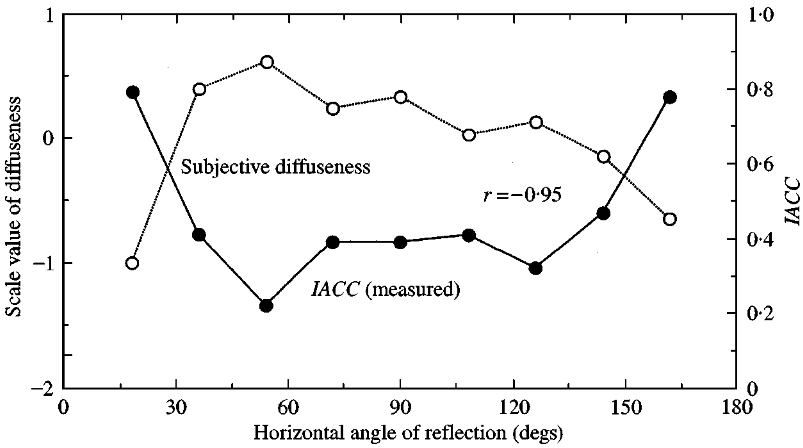

Figure 3. Scale values of subjective diffuseness and the $IACC$ as a function of the horizontal angle of incidence to a listener, with 1/3 octave bandpass noise with a center frequency of 1 kHz (7 subjects).

the subjective preference is formulated:

$$S \approx -\alpha(IACC)^{\beta}, \tag{1}$$

where $\alpha = 2{\cdot}9$ and $\beta = 3/2$.

The results of scale values by the paired-comparison test and values calculated by equation (1) are shown in Figure 4 as a function of the $IACC$. There is a great variation of data when $IACC < 0{\cdot}5$, but there are no essential differences between the results obtained at frequencies between 250 Hz and 4 kHz. The scale values of subjective diffuseness, which depend on the horizontal angle, are shown in Figure 3 for 1/3 octave-bandpass noise with a center frequency of 1 kHz. The most effective horizontal angles of reflections differ depending on the frequency range and are

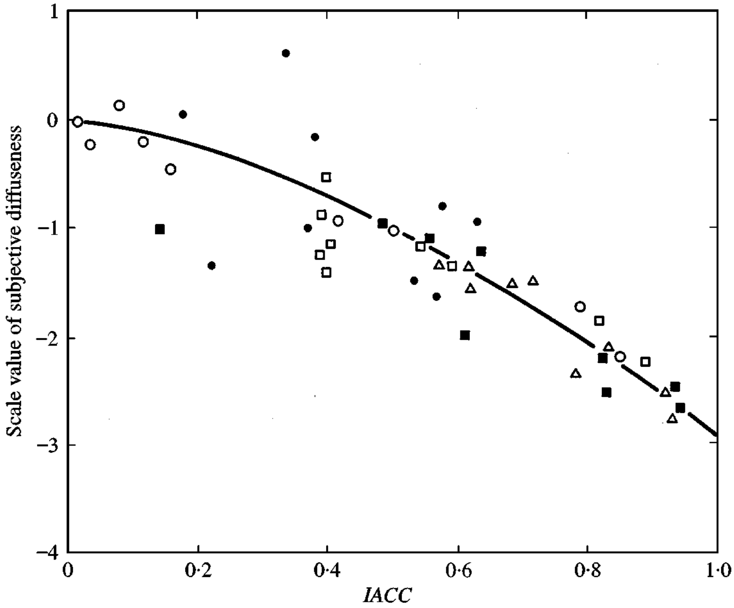

Figure 4. Scale values of subjective diffuseness as a function of the *IACC* calculated. Different symbols indicate different frequencies of the 1/3 octave bandpass noise: (△), 250 Hz; (○), 500 Hz; (□), 1 kHz; (●), 2 kHz; (■), 4 kHz. (———), Regression line by equation (1).

inversely related to the *IACC* values. As demonstrated in Figure 3, these are about ± 55° for the 1 kHz range.

3.2. APPARENT SOURCE WIDTH IN RELATION TO IACC AND W_{IACC}

For a sound field with a predominately low-frequency range, the interaural cross-correlation function has no sharp peak for the delay range of $|\tau| < 1$ ms, and W_{IACC} becomes wider (see Figure 5). The values of W_{IACC} for bandpass noise are calculated by using the equation [8]

$$W_{IACC}^{(\delta)} \approx \frac{4}{\Delta\omega_c} \cos^{-1}\left(1 - \frac{\delta}{IACC}\right) \quad (s), \tag{2}$$

where $\Delta\omega_c = 2\pi(f_1 + f_2)$ and f_1 and f_2 are the lower and upper frequencies of an ideal filter. For simplicity, δ is defined by 0·1(IACC) in obtaining W_{IACC} as shown in Figure 6. Note that a wider apparent source width (*ASW*) may be perceived within the low-frequency bands and by decreasing the *IACC*. More clearly, the *ASW* may be well described by both *IACC* and W_{IACC} [14]. The scale values of *ASW* were obtained by paired-comparison tests with 10 subjects. The values of W_{IACC} were varied by changing the center frequencies of 1/3 octave-bandpass noises. The center frequencies were 250 Hz, 500 Hz, 1 kHz, and 2 kHz. The values of *IACC* were adjusted by controlling the ratio of the sound pressure of the reflections to the sound pressure of the direct sound. Because the listening level affects *ASW*, the total sound pressure levels at the ear canal entrances were kept constant at a peak

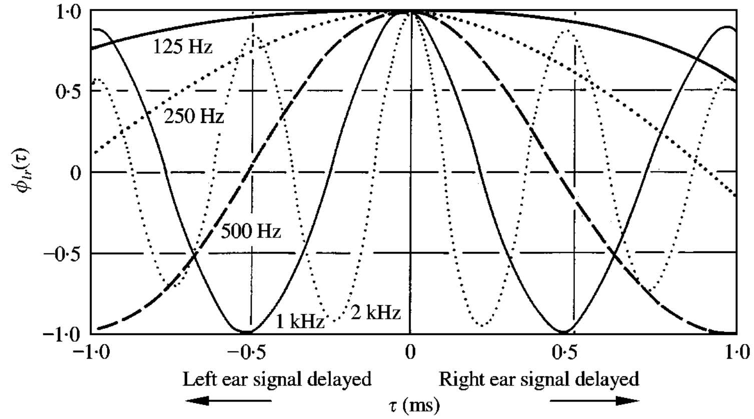

Figure 5. Measured interaural cross-correlation functions for the 1/3 octave-bandpass noises with center frequencies of 125 Hz–2 kHz.

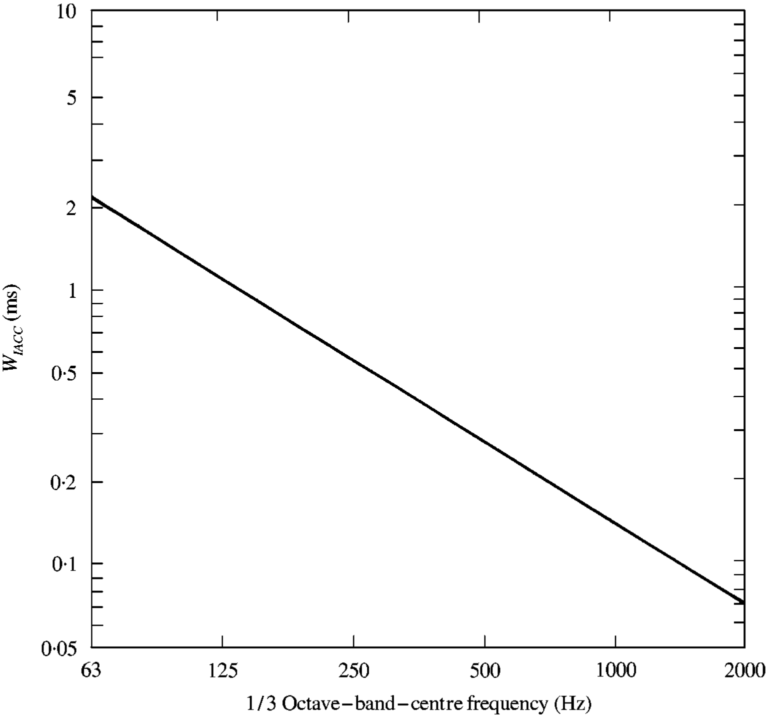

Figure 6. W_{IACC} as a function of the center frequency.

of 75 dB(A). Subjects judged which of two sound sources they perceived to be wider. The results of the analysis of variance for the scale values $s(ASW)$ indicate that both the factors $IACC$ and W_{IACC} are significant ($p < 0{\cdot}01$) and that contribute to the $s(ASW)$ independently such that

$$s(ASW) = f(IACC) + f(W_{IACC}) \approx a(IACC)^{3/2} + b(W_{IACC})^{1/2}, \tag{3}$$

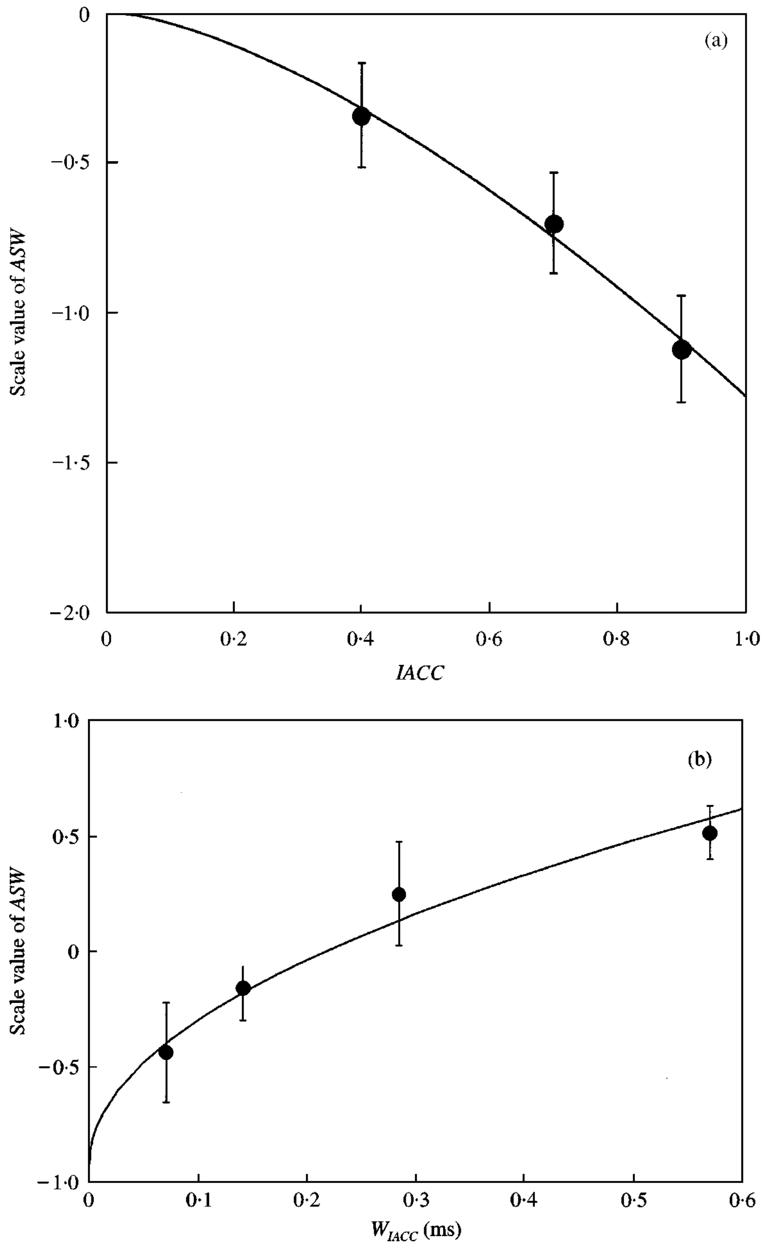

Figure 7. Average scale values of *ASW* as a function of the *IACC* (a), and the W_{IACC} (b).

where coefficients $a \approx -1{\cdot}64$ and $b \approx 2{\cdot}44$ are obtained by regressions of the scale values with 10 subjects as shown in Figure 7. As shown in Figure 8, scale values s(ASW) calculated by equation (3) and measured scale values are obviously in good agreement ($r = 0{\cdot}97$, $p < 0{\cdot}01$). The *ASW* for each subject can be obtained by using the same equation used in calculating the global *ASW* simply by changing the weighting coefficients a and b.

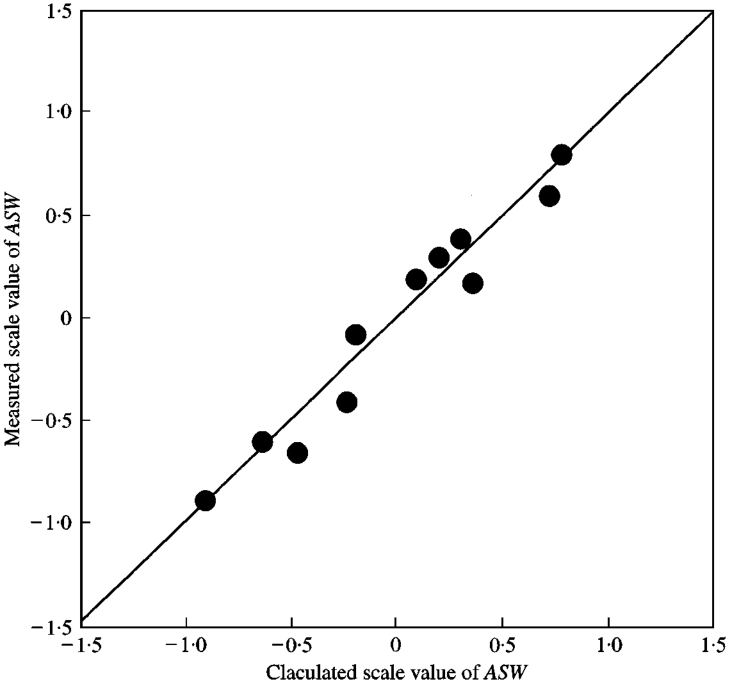

Figure 8. Relationship between the measured scale values of *ASW* and the scale values calculated by equation (3).

4. SUBJECTIVE ATTRIBUTES IN RELATION TO THE AUTOCORRELATION MECHANISM

4.1. PREFERRED DELAY TIME OF A SINGLE REFLECTION AND SUBSEQUENT REVERBERATION TIME FOR LISTENERS

Results of subjective preference as an overall psychological response for a sound field with a single reflection indicated that the most-preferred delay of the reflection may be found by the envelope curve of *ACF* (see Figure 9 for music Motifs A and B, $2T = 35$ s), $[\Delta t_1]_p \approx \tau_p$, such that

$$|\phi(\tau)|_{envelope} = kA^c \quad \text{at } \tau = \tau_p, \tag{4}$$

where A is the pressure amplitude of the single reflection, $k = 0{\cdot}1$, and $c = 1$. If the envelope of *ACF* is exponential, then the above equation is simply expressed by

$$[\Delta t_1]_p \approx (1 - \log_{10} A)\tau_e, \tag{5}$$

where τ_e is the effective duration defined by the 10 percentile delay of the ACF-envelope [15] as shown in Figure 10(a).

Results of subjective preference tests for sound fields with the direct sound and the single reflection are shown in Figure 11. The most-preferred delay times of the single reflection ($A = 1$) differ from 128 to 32 ms, but these delay times correspond well to the effective duration of *ACF* of Motif A ($\tau_e = 127$ ms) and Motif B ($\tau_e = 35$ ms) respectively. For more general cases with different amplitudes of the

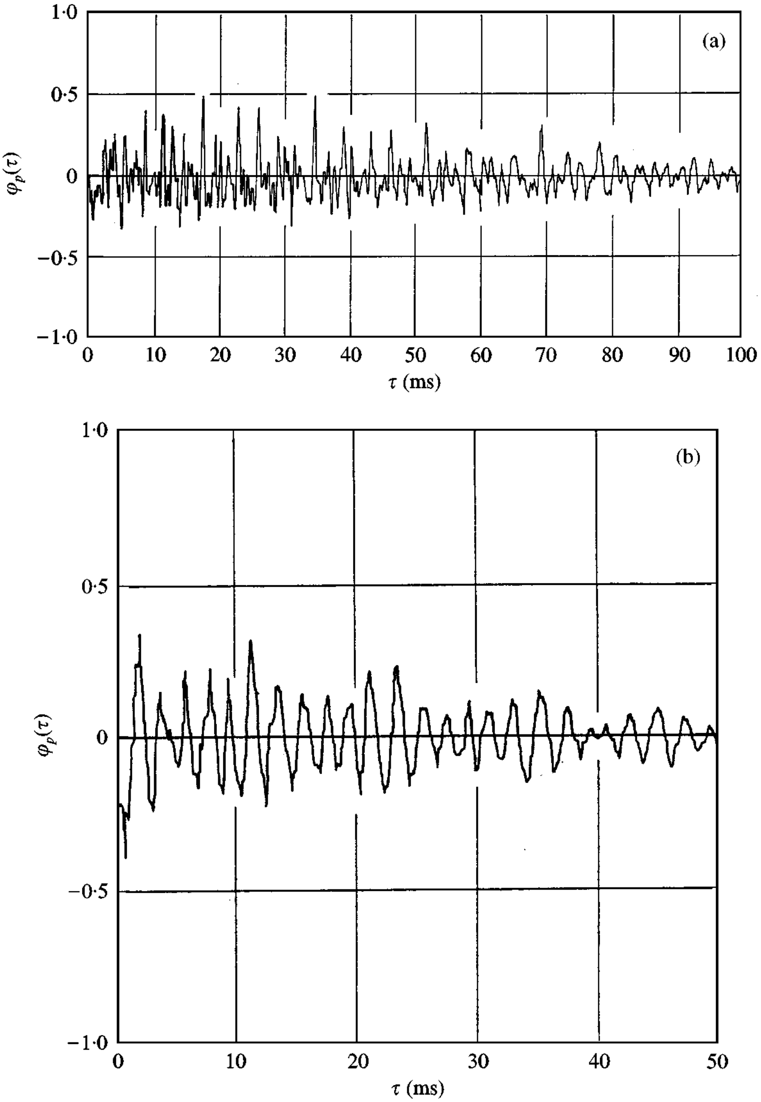

Figure 9. Measured autocorrelation functions for music Motiff A (Royal Pavane by Gibbons; $\tau_e \approx 127$ ms) (a) and Motif B (Sinfonietta by Arnold; $\tau_e \approx 35$ ms) (b).

reflections, the preferred delay times of the single reflection are shown in Figure 12, which also plots results with continuous speech. Similar results were also obtained with Korean subjects as well as German and Japanese ones [16].

Such a relationship also holds for other important subjective responses in relation to the temporal factor as discussed below. The constants k and c used in calculating subjective responses to a sound field, based on the ACF of source signals, and the ranges in the experiments are listed in Table 1 [15, 17–22]. Later, it is shown that the values of τ_e in equation (5) may be replaced by $(\tau_e)_{min}$ of the running $ACF(2T = 2\text{ s})$ with a running interval of 100 ms for music pieces used. This is because the fact that the corresponding music part of $(\tau_e)_{min}$ contains the most rapid movement including important information, and thus stimulates the

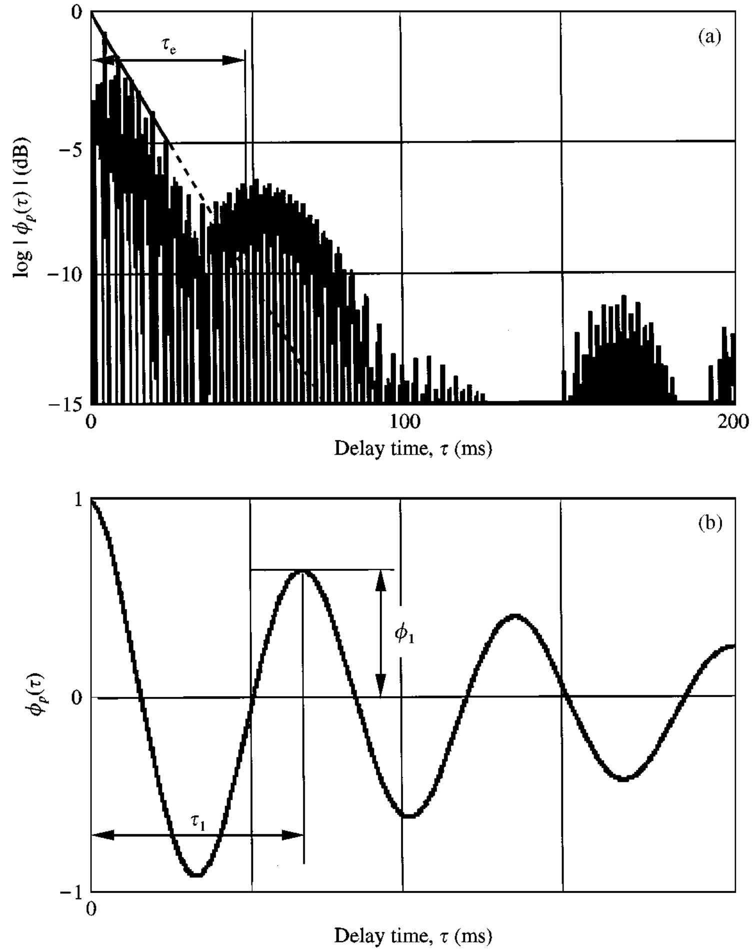

Figure 10. Definitions of orthogonal factors extracted from the *NACF*, τ_e defined by the 10-percentile delay (at -10 dB), obtained practically from the decay rate extrapolated in the range from 0 to -5 dB of *NACF* (a); and τ_1 and ϕ_1 in the fine structure of the *NACF* (b).

brain effectively [23]. If the sound field is a time-variant system due to, for example, an air conditioner, then $[\Delta t_1]_p$ in equations (4) and (5) is replaced by $[\Delta t_1 + \Delta/2]_p$, where Δ is the modulated interval in the delay time of the reflection caused by the airflow [24].

The most-preferred subsequent reverberation time is well described by

$$[T_{sub}]_p \approx 23\,\tau_e. \tag{6}$$

For vocal music, which is a typical source in an opera house, equation (6) also holds [25].

4.2. PREFERRED DELAY OF A SINGLE REFLECTION FOR PERFORMERS

From preference judgements with respect to ease of music performance by an altorecorder [21], the most-preferred delay time of the single reflection also may be

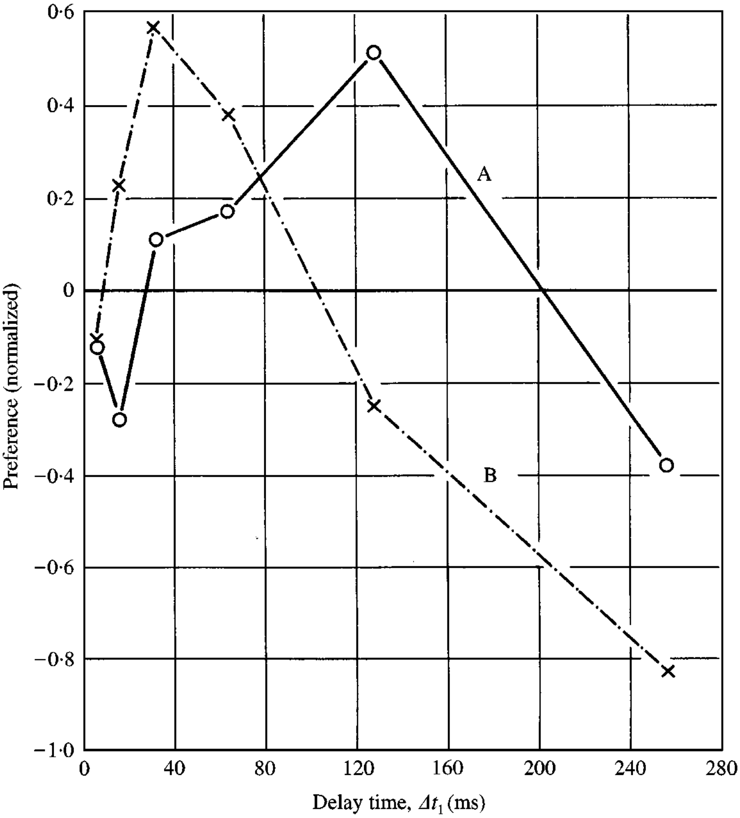

Figure 11. Preference scores as a function of the delay time of single reflection for music Motif A (Royal Pavane by Gibbons; $\tau_e \approx 127$ ms) and Motif B (Sinfonietta by Arnold; $\tau_e \approx 35$ ms).

described by equation (4). In this case, coefficients are $k = 2/3$ and $c = 1/4$. The coefficient k for performers differs by a factor of about 7 from that for listeners. This indicates that performers evaluate the amplitude of the reflection about as being 7 times greater than listeners do, which is known as the "missing reflection for performers". A similar tendency was found for cellists with $k = 1/2$ and $c = 1$ [22]. Musicians prefer weaker amplitudes than listeners do. Note that, for listerners, if $\Delta t_1/\tau_e = 1$, then $20 \log A_1 = 0$ dB. The constants k and c used in the formulae for calculating fundamental subjective attributes for sound fields, based on the ACF-envelope of source signals, and the ranges in the experiments are listed in Table 1.

4.3. PITCH IDENTIFICATION USING THE AUTOCORRELATION FUNCTION MODEL

In order to describe the phenomenon of the missing fundamental of a complex tone, perceived pitch is described in terms of τ_1, which is defined by the time delay at the first maximum peak of the normalized autocorrelation function (*NACF*), as shown in Figure 10(b).

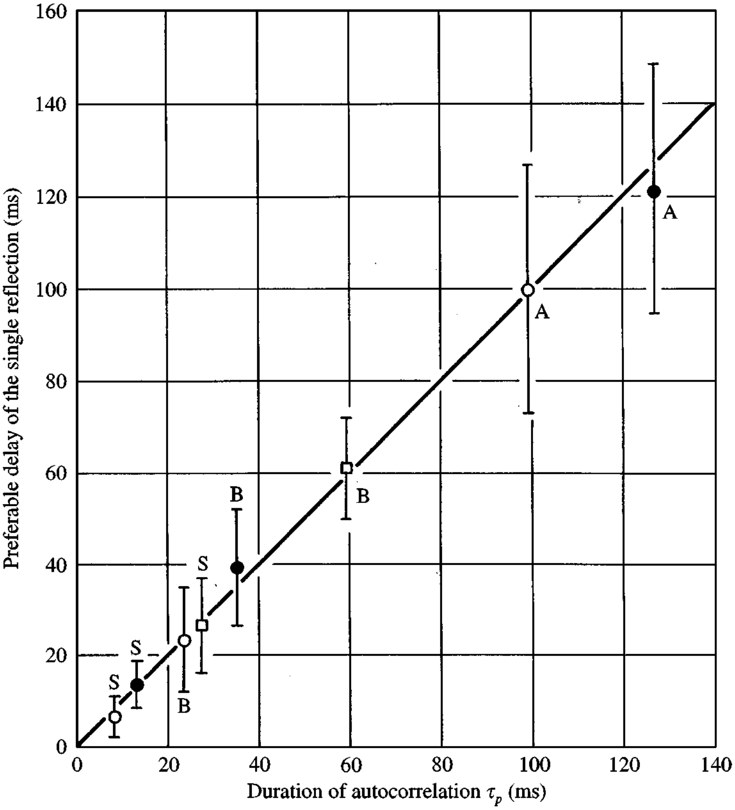

Figure 12. Relationship between the preferred delay time of the single reflection and the duration of *ACF* expressed by equation (4) with $k = 0{\cdot}1$ and $c = 1$ (the preferred delay time can be approximately calculated by equation (5)).

Pitch-matching tests were conducted in a semi-anechoic chamber with complex tones (pure tones of 600, 800, 1000, 1200, and 1400 Hz, with the same amplitudes as indicated in Figure 13(a)) with two different waveforms (in-phase and random phase) [26]. The starting phases of all components of in-phase stimuli were adjusted to zero. The phases of the components of random-phase stimuli were randomly set to avoid having any specific peak in its real waveform. The tests were performed by comparing pitches of a complex tone and a pure tone. The total *SPL* of each sound source was fixed at 74 dB. The waveforms of the in- and random-phase stimuli respectively, are shown in Figures 13(b) and 13(c). The *NACF*s of both stimuli were calculated at the integrated interval $2T = 0{\cdot}8$ s, as shown in Figure 13(d) for in-phase and random-phase stimuli. The waveforms differed greatly, as shown in Figures 14(b) and 14(c), but their *NACF*s were identical, τ_1 being at 5 ms. The corresponding frequency is not included in the spectrum nor in the real waveforms, but is included in the peak period in the *ACF*. The probability of matching data for each 1/12 octave band (chromatic scale) of the in- and random-phase stimuli for five musicians (two males and three females, 20–26 years old) are shown in Figures 14(a) and 14(b) respectively. In this experiment, none of the subjects identified the pitch above 600 Hz, which is the

Table 1

Constants related to the ACF envelope of source signals for calculating various subjective responses to sound fields with single reflection, in relation to the ACF envelope [15, 17–22]

Subjective atributes	In equation (4)		Delay time to be obtained	Range of amplitude examined (dB)	Source signals	Investigators
	k	c				
Preference of listeners	0·1	1	Preferred delay time	$7{\cdot}5 \geqslant A_1 \geqslant -7{\cdot}5$	Speech & music	Ando [15]
Threshold of perception of reflection	2	1	Critical delay time	$-10{\cdot}0 \geqslant A_1 \geqslant -50{\cdot}0$	Speech	Seraphim [17]
50% echo disturbance	0·01	4	Disturbed delay time	$0{\cdot}0 \geqslant A_1 \geqslant -6{\cdot}0$	Speech	Haas [18] Ando *et al.* [19]
Coloration	$10^{-5/2}$	-2	Critical delay time	$-7{\cdot}0 \geqslant A_1 \geqslant -27{\cdot}0$	Gaussian noise	Ando and Alrutz [20]
Preference of alto-recorder soloists	2/3	1/4	Preferred delay time	$-10{\cdot}0 \geqslant A_1 \geqslant -34{\cdot}0$	Music	Nakayama [21]
Preference of cello-soloists	1/2	1	Preferred delay times	$-15{\cdot}0 \geqslant A_1 \geqslant -21{\cdot}0$	Music	Sato *et al.* [22]

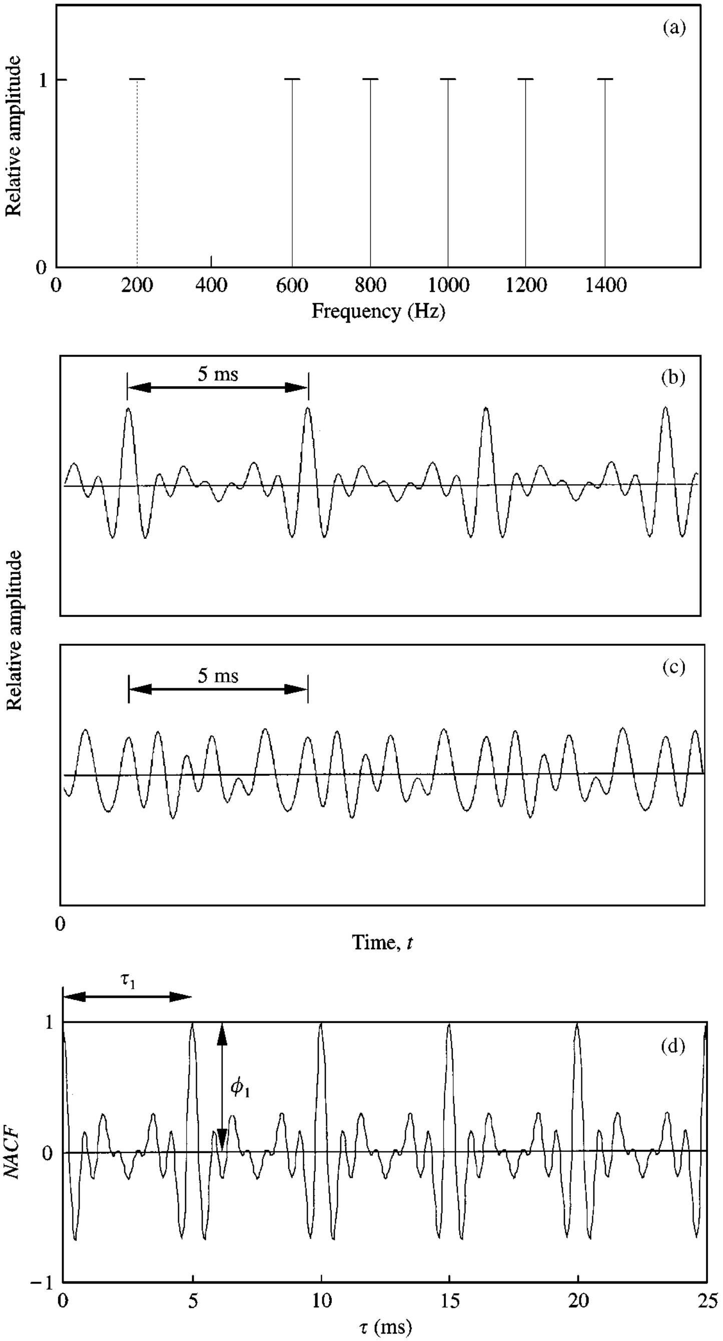

Figure 13. (a) Complex tones presented with pure-tone components of 600, 800, 1000, 1200, and 1400 Hz without the fundamental frequency of 200 Hz. (b) Real waveforms of complex tones in-phase components, (c) Real waveform of random-phase components. (d) Normalized *ACF* of the two complex tones with its period of 5 ms (200 Hz); Both τ_1 and ϕ_1 are defined in the fine structures of the NACF. The value of τ_1 is defined as the time duration at the first maximum peak of NACF except for $\tau = 0$. The ϕ_1 is the magnitude of *NACF* at τ_1.

lowest frequency of the components. For both in-phase and random phase, about 80% of the responses clustered around the first two octaves of the stimulus fundamental within a semitone. There were no fundamental differences in the distributions of pitch-matching data between the two conditions.

For more details, histograms between 183·3 and 218·0 Hz for all the subjects are shown in Figures 14(c) and 14(d) for in- and random-phase stimuli. Averaged values and standard deviations (SD) of the data obtained from each subject at frequencies near 200 Hz are listed in Table 2. The results obtained for the pitch under the two conditions are clearly similar. These results support the *ACF* model, but not the fine structure theory [27].

When the phase relations in complex tones consisting of the third to seventh harmonics of the fundamental frequency of 200 Hz are random, subjects indicated that the matched pitch was close to 200 Hz as when in-phase stimuli was used. The probability of pitch perception for random-phase stimuli was the same as that of the in-phased stimuli. This may be because the values of τ_1 and ϕ_1 (the *ACF*) of

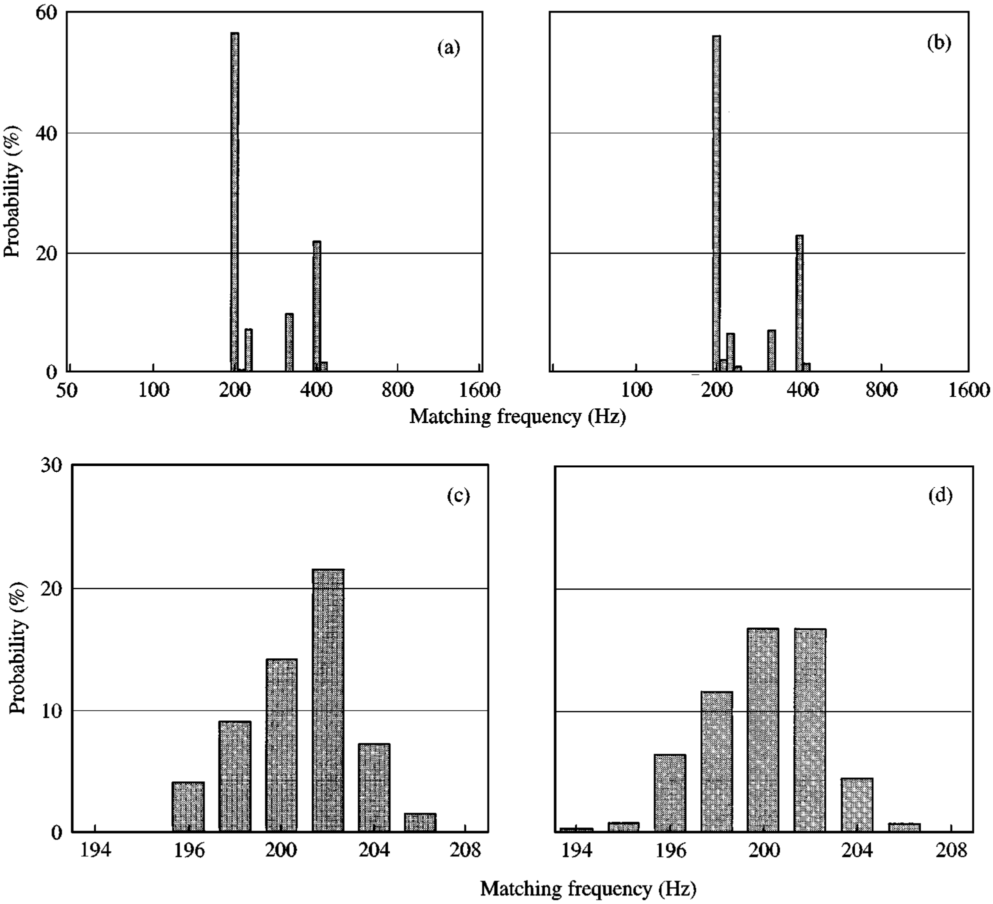

Figure 14. Results of the pitch-matching tests for global data of five subjects: in-phase stimuli (a), and random-phase stimuli (b). Results of the pitch-matching test around 200 Hz (between 183·3 and 218 Hz) of global data from five subjects: in-phase stimuli (c); and random-phase stimuli (d).

TABLE 2

Average values and standard deviations (SD) of pitch matching data

	Average (Hz)		SD	
Subject	In-phase	Random phase	In-phase	Random phase
A	202·6	201·0	1·89	2·44
B	199·1	198·3	1·70	1·42
C	202·5	202·1	1·18	1·76
D	203·7	201·7	2·29	1·65
E	202·2	202·2	1·87	2·07
Total	201·9	201·0	2·43	2·38

both stimuli are very similar. Therefore, the pitch can be calculated through the *ACF* analysis.

Another pitch-matching test was conducted by use of complex noises, which consist of bandpass noises changing their bandwidth instead of pure tones. This experiment was to determine whether pitch perception is changed by the magnitude of the maximum peak in the *NACF*. Factors of τ_1 and ϕ_1 (Figure 10(b)) extracted from the *NACF* are used to calculate the pitch.

The method used in this experiment was the same as that of the experiment mentioned above. The frequency components were changed by the width of the band-pass noise with a cut-off slope of 1080 dB/octave [28]. The center frequency of the five noise components was 600, 800, 1000, 1200, and 1400 Hz. The bandwidths (Δf) of the four components were 40, 80, 120 and 160 Hz (Figure 15(a)). Their waveforms are shown in Figures 15(b)–15(e). The *NACF*s for four conditions are shown in Figures 15(b′)– 15(e′). The integrated interval $2T$ was 2 s, and τ_1 for each stimulus was found to be about at 5 ms. The values of ϕ_1 were 0·804 (Δf: 40 Hz), 0·641 (Δf: 80 Hz), 0·471 (Δf: 120 Hz), and 0·416 (Δf: 160 Hz).

The probabilities of the matching data counted for each 1/12 octave band for five musicians (two males and three females, 20–25 years old) who were different from those in above experiment are shown in Figure 16. All historgrams show that there is a strong tendency to perceive a pitch of 200 Hz for any stimulus. This agrees with the prediction based on the value of τ_1. The results in Figure 16 indicate that a stimulus with a narrow bandwidth gives a stronger pitch corresponding to 200 Hz than does a stimulus with a wide bandwidth. The probabilities matched to 400 Hz (one octave higher than 200 Hz) keep increasing as the bandwidth narrows.

In the experiment, the standard deviation for the perceived pitches increased because the value of ϕ_1 decreased as Δf increased. The probabilities of the frequency ranges in Figure 16 are distributed more widely than the results of the experiment in the previous section.

The pitch strength corresponding to the value of ϕ_1 can be considered a probability of responses clustering into a certain range of a histogram. The relationship between the percentage of responses within 200 ± 16 Hz and the value

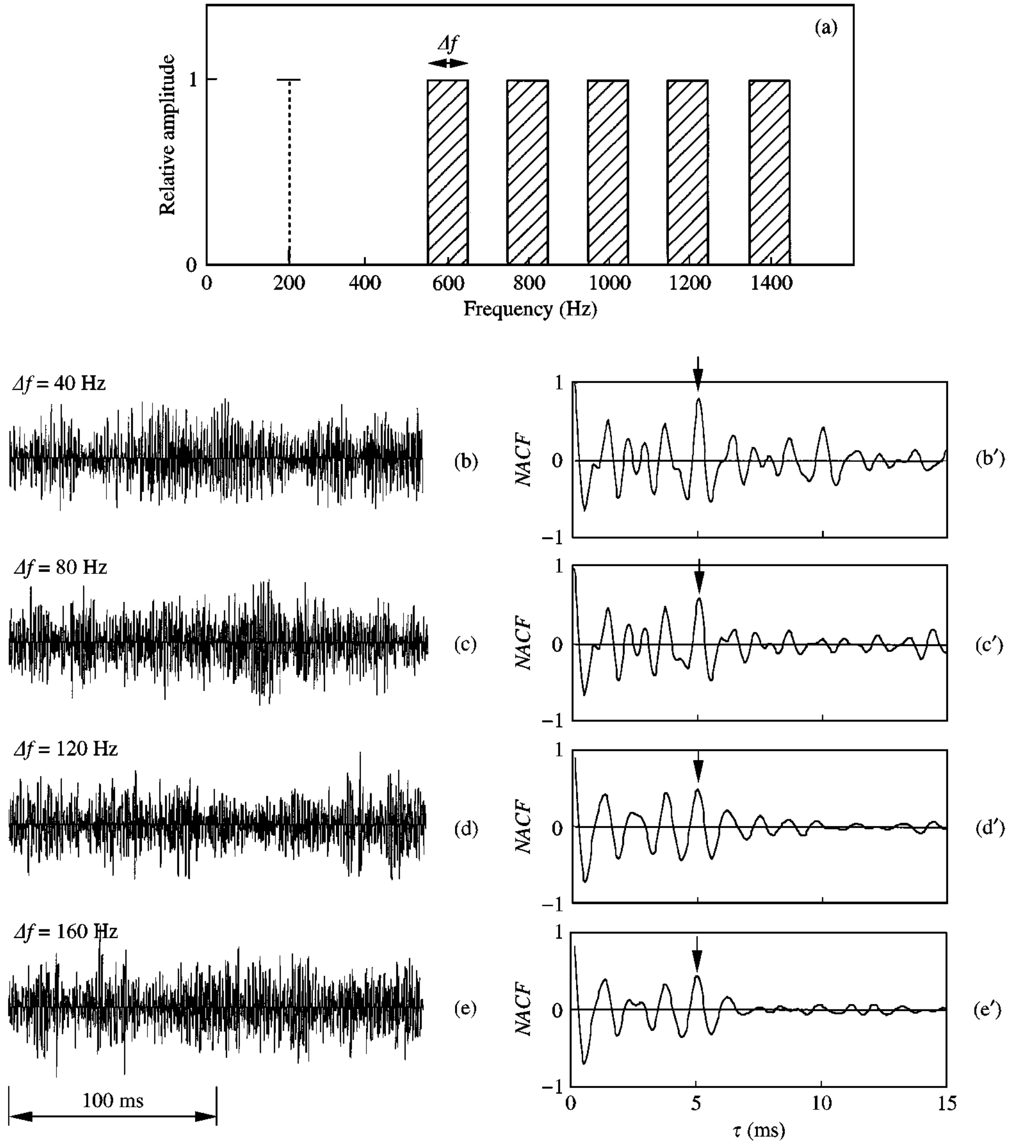

Figure 15. Complex noise containing the center frequencies: 600, 800, 1000, 1200, and 1400 Hz used (a). Its fundamental frequency is around 200 Hz. The Δf represents the bandwidth. Waveforms of the four complex noises applied: $\Delta f = 40$ Hz (b); $\Delta f = 80$ Hz (c); $\Delta f = 120$ Hz (d); and $\Delta f = 160$ Hz (e). The *NACF*s of the stimuli: $\Delta f = 40$ Hz (*b*′); $\Delta f = 80$ Hz (c′); $\Delta f = 120$ Hz (d′); and $\Delta f = 160$ Hz (e′).

of ϕ_1 for all subjects is shown in Figure 17 ($r = 0{\cdot}98$). The results of the above experiment at $\phi_1 = 1$ is also plotted as a filled square in this figure. The value of ϕ_1 appears to be an effective proxy for predicting the pitch strength. These results support Wightman's model [29], and indicate that the *ACF* is useful for estimating the pitch strength, as well as the pitch value.

When the bandwidth Δf of complex noises for third to seventh harmonics with the fundamental frequency 200 Hz are the smallest (40 Hz), the probability of the

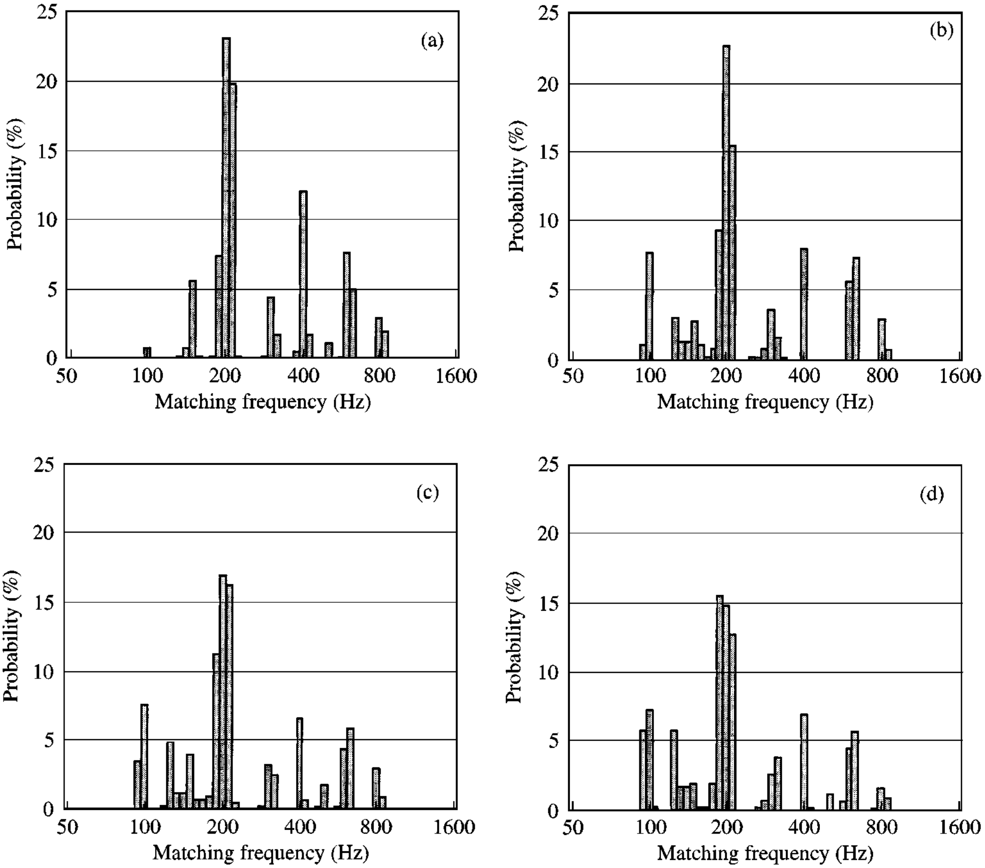

Figure 16. Results of the pitch-matching tests for the global data of five subjects: $\Delta f = 40$ Hz (a); $\Delta f = 80$ Hz (b); $\Delta f = 120$ Hz (c); and $\Delta f = 160$ Hz (d).

pitches perceived at 200 Hz is increased. This frequency can be predicted from the inverse of τ_1. Also, the pitch strength relies on the ϕ_1, and the pitch is strongly perceived for the largest ϕ_1. Therefore, the *ACF* model is acceptable for the prediction of the pitch of complex noise without a fundamental frequency, not only for complex tone but for ripple noise.

5. CALCULATIONS OF SPEECH INTELLIGIBILITY OF EACH SINGLE SYLLABLE IN RELATION TO THE FOUR ORTHOGONAL FACTORS EXTRACTED FROM THE AUTOCORRELATION FUNCTION

Concerning global speech intelligibility of sound fields, a speech transmission index (*STI*) has been proposed [30]; however, this index is calculated only by the effect of sound field. This section discusses the way that the speech intelligibility (*SI*) of each single syllable can be described in terms of the distance between the template source signal and a sound-field signal. In the calculation of this distance,

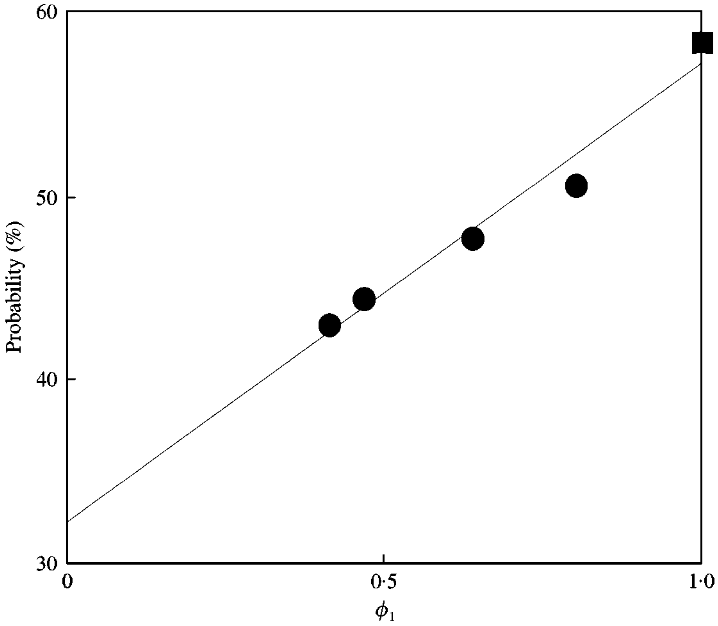

Figure 17. Relationship between ϕ_1 and probability of the pitch being within 200 ± 16 Hz.

only four independent factors extracted from the *ACF* of the direct sound as a template, and signals of sound fields are utilized.

Let S_K^T be the isolated template-syllable K, and let S_X^{SF} be another syllable X in the sound field, where T signifies the template and SF is the sound field. The distance between S_K^T and S_X^{SF} for one of the four factors is given by

$$d(x) = D_x\,(S_X^{SF}, S_K^T) = |\,C_X^{SF}, C_K^T\,|, \tag{7}$$

where $x = \tau_e$, τ_1, ϕ_1, or $\Phi(0)$, which are defined in Figure 10 and the factor $\Phi(0)$ is the energy which is represented at the origin of the delay of the *ACF*; C_K^T and C_X^{SF} are characteristics of an isolated template-syllable S_K^T and that of the another syllable of the sound-field S_X^{SF} in the brain respectively; and where

$$D_{\tau_e}(S_X^{SF}, S_K^T) = \left\{\sum_{i=1}^{n} |\log(\tau_e^i(S_X^{SF})) - \log(\tau_e^i(S_K^T))|\right\}\Big/ n,$$

$$D_{\tau_1}(S_X^{SF}, S_K^T) = \left\{\sum_{i=1}^{n} |\log(\tau_1^i\,(S_X^{SF})) - \log(\tau_1^i\,(S_K^T))|\right\}\Big/ n,$$

$$D_{\phi_1}(S_X^{SF}, S_K^T) = \left\{\sum_{i=1}^{n} |\log|(\phi_1^i\,(S_X^{SF})| - \log|\phi_1^i\,(S_K^T)\,|\right\}\Big/ n,$$

$$\begin{aligned} D_{\Phi_{(0)}}(S_X^{SF}, S_K^T) = \Bigg\{\sum_{i=1}^{n} & |\log(\Phi(0)^i\,(S_X^{SF})/\Phi(0)^{max}\,(S_X^{SF})) \\ & - \log(\Phi(0)^i(S_K^T)/\Phi\,(0)^{max}\,(S_K^T))|\Bigg\}\Big/ n, \end{aligned} \tag{8}$$

where n is the number of the frame of the running *ACF*. The procedure of calculation of the running *ACF* is described later (Figures 19 and 20).

Let the number of syllables of the same category with S_X^{SF} be N. Japanese syllables can be categorized by the type of consonant and the vowel as indicated in Table 3 because listeners do not confuse syllables over the category [31]. The intelligibility of the syllable k for one of the four factors may be obtained as

$$\Psi_k(x) = 100N \exp\left(-\frac{d_k(x)}{d_1(x)} \cdots \frac{d_k(x)}{d_{k-1}(x)} \, \frac{d_k(x)}{d_{k+1}(x)} \cdots \frac{d_k(x)}{d_N(x)} \right). \tag{9}$$

Let us now demonstrate an example of studying the intelligibility of the sound field which consists of the direct sound and the single reflection. The amplitude of reflection was the same as that of the direct sound, and the delay time of the reflection Δt_1 was varied in the range between 0 and 480 ms. The test signals consisted of 50 Japanese monosyllables with maskers (see Figure 18). The direct

TABLE 3

Categorization of each Japanese syllable

(a) *Unvoiced consonant*

	Category		Consonant				
	Vowel	Number	K	S	T	H	P
Not contracted (Category A)	A	1	KA	SA	TA	HA	PA
	I	2	KI	SI	TI	HI	PI
	U	3	KU	SU	TU	HU	PU
	E	4	KE	SE	TE	HE	PE
	O	5	KO	SO	TO	HO	PO
Contracted (Category B)	YA	6	KYA	SYA	TYA	HYA	PYA
	YU	7	KYU	SYU	TYU	HYU	PYU
	YO	8	KYO	SYO	TYO	HYO	PYO

(b) *Voiced consonant*

	Category		Consonant								
	Vowel	Number	N	M	Y	R	W	G	Z	D	B
Not contracted (Category C)	A	9	NA	MA	YA	RA	WA	GA	ZA	DA	BA
	I	10	NI	MI	—	RI	—	GI	ZI	—	BI
	U	11	NU	MU	YU	RU	—	GU	ZU	—	BU
	E	12	NE	ME	—	RE	—	GE	ZE	DE	BE
	O	13	NO	MO	YO	RO	—	GO	ZO	DO	BO
Contracted (CategoryD)	YA	14	NYA	MYA	—	RYA	—	GYA	ZYA	—	BYA
	YU	15	NYU	MYU	—	RYU	—	GYU	ZYU	—	BYU
	YO	16	NYO	MYO	—	RYO	—	GYO	ZYO	—	BYO

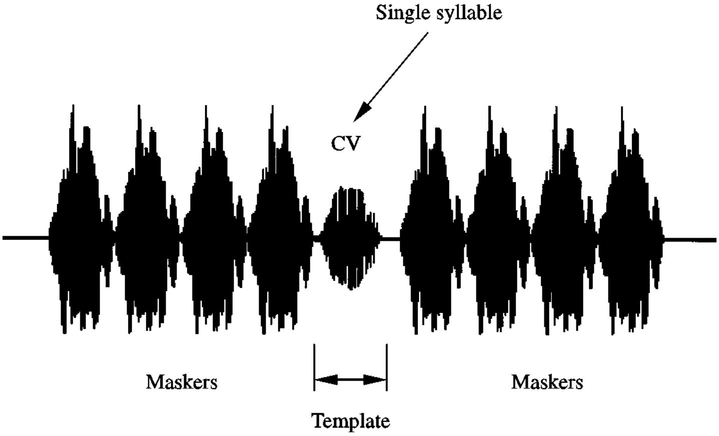

Figure 18. Example of a waveform of Japanese single syllables between artifical non-meaning forward and backward maskers. The direct sound without maskers used as a template.

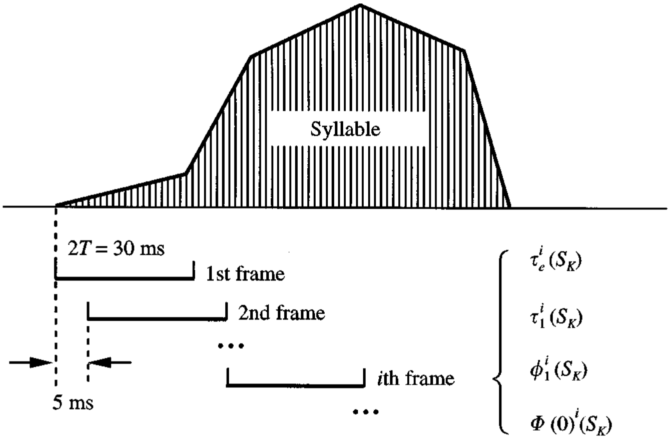

Figure 19. Running *ACF* of the syllable. Four orthogonal factors are obtained from each step of the running *ACF*, integration interval $2T = 30$ ms, with a running step of 5 ms for both the template-syllable and the syllable in the sound field.

sound without maskers was used as a template. The running *ACF*s, integration interval $2T = 30$ ms, with the running step of 5 ms were calculated (see Figure 19). Four orthogonal factors were obtained from each step of running *ACF*. The important former half-parts of the *ACF*s $\phi(0) < 0{\cdot}5$ (see Figure 20) of both a template and a test syllable were analyzed. The values of $\phi(0)$ are defined as the $\Phi(0)$ normalized by the maximum value ($\Phi(0)^{max}$) among all frames. Actually, two types of distances are calculated with the four orthogonal factors. One is the distance between the template and the direct sound with maskers, and the other is the distance between the template and the reflected sound with maskers. The shorter of those distances is selected in the calculation of syllable intelligibility. By using the distances given by equation (8), the intelligibility of all syllables can be calculated by equation (9).

The intelligibility for four factors can be combined linearly due to the expression given by

$$\Psi_k = a\Psi_k(\tau_e) + b\Psi_k(\tau_1) + c\Psi_k(\phi_1) + d\Psi_k(\Phi(0)), \tag{10}$$

where a, b, c, and d are the coefficients to be evaluated. Results of both calculated and tested intelligibility for single syllables belonging to each category are demonstrated in Figure 21. Twenty-one subjects were used in the experiments, so that about a 5% error results for a single subject's judgement in these figures. Averaged results for each category are shown in Figure 22 and mean values of

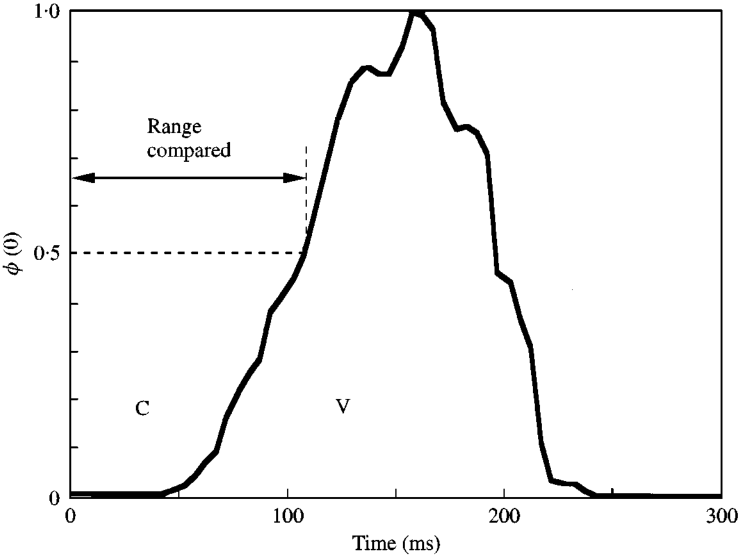

Figure 20. The power function of a syllable. The important former half-parts of running *ACF*s $\phi(0) < 0{\cdot}5$ of both of template and a test syllable were analyzed for four orthogonal factors of the *ACF*.

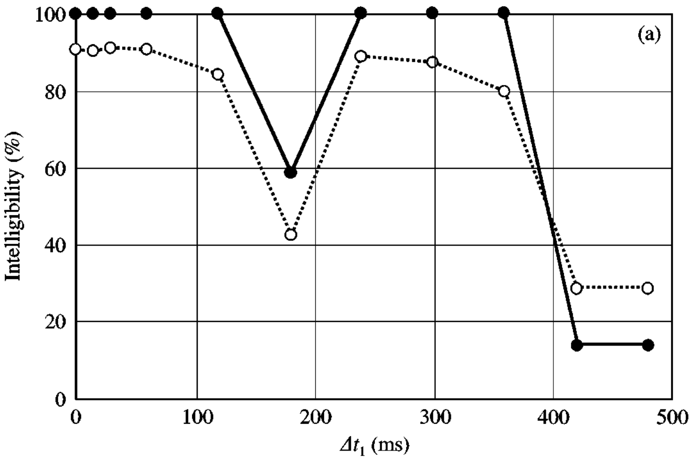

Figure 21. Results of calculated and testing intelligibilities for single syllables belonging to each category: /ha/ (a); /be/ (b); /hya/ (c); /nyu/ (d). Twenty-one subjects were used in the experiments: (——); measured; and (······) calculated

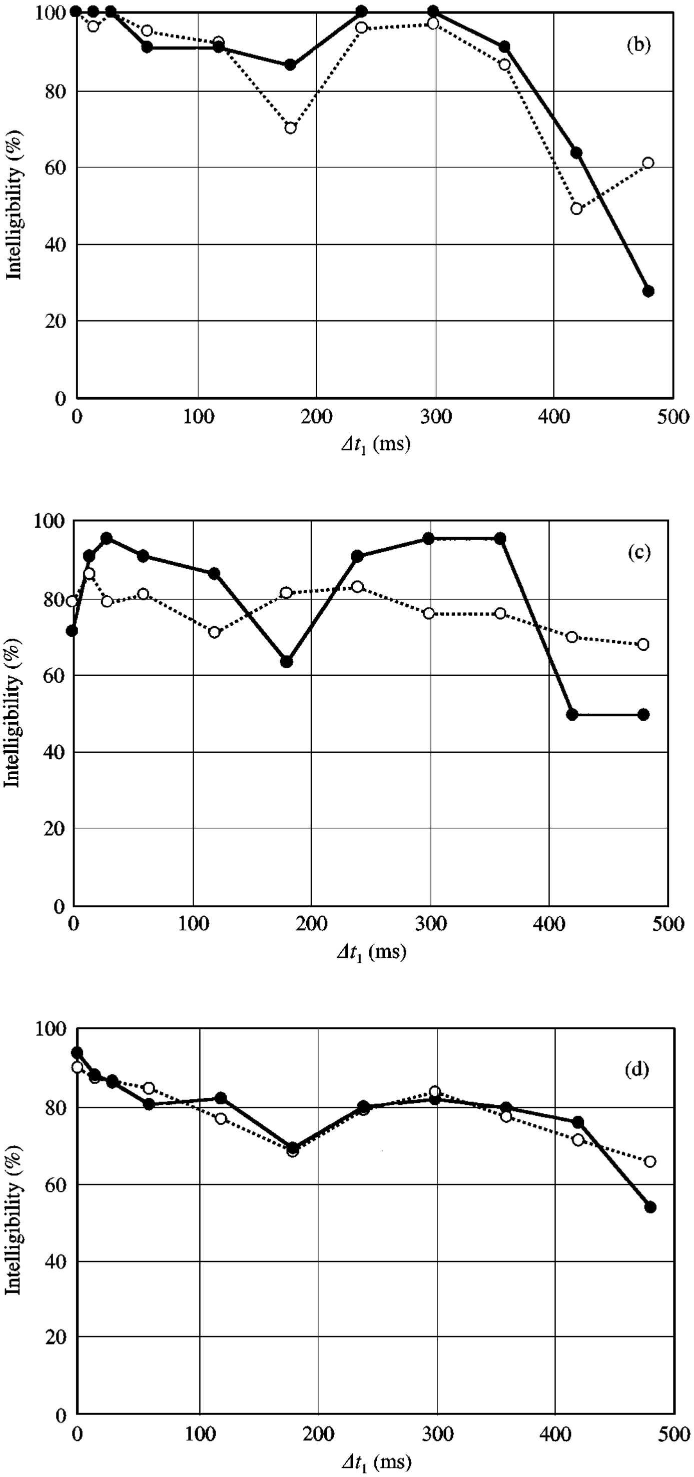

Figure 21. Continued.

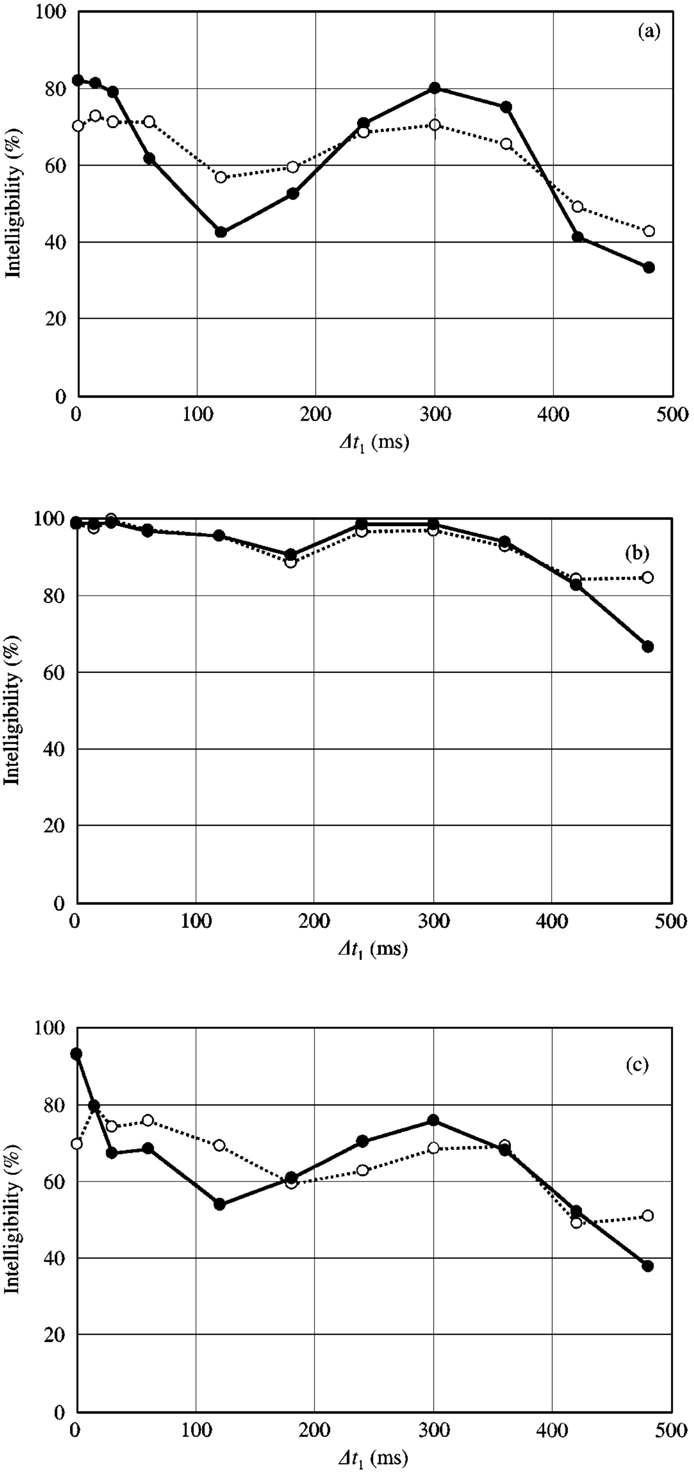

Figure 22. Averaged results of calculated and tested intelligibilities for each category: category A (a); category B (b); category C (c); category D (d). (———): Measured; and (······): calculated.

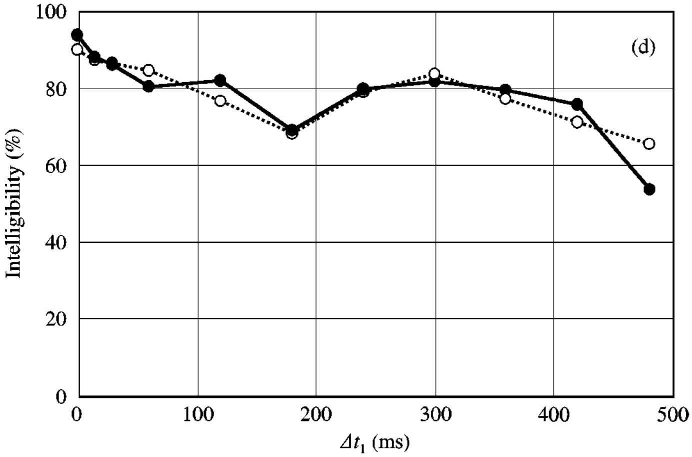

Figure 22. Continued.

TABLE 4

Contribution of each factor extracted from ACF to speech intelligibility. Values were normalized by the maximum of four coefficients obtained from the multiple regression [32]. *The most of maximum values may be found for the factor of* τ_e

Consonant	Vowel	τ_e	τ_1	ϕ_1	$\Phi(0)$
	A	0·50	0·60	0·36	1·00
	U	0·07	0·08	0·25	1·00
	E	1·00	0·66	0·23	0·46
Unvoiced	O	0·62	1·00	0·13	0·25
	YA	1·00	0·17	0·55	0·56
	YU	1·00	0·19	0·30	0·47
	YO	1·00	0·30	0·41	0·88
	A	0·74	0·92	0·89	1·00
	I	0·88	0·38	0·01	1·00
	U	1·00	0·80	0·30	0·29
	E	0·19	0·42	1·00	0·01
Voiced	O	1·00	0·25	0·09	0·80
	YA	1·00	0·17	0·55	0·56
	YU	1·00	0·19	0·30	0·47
	YO	0·25	0·21	1·00	0·51

calculated and tested intelligibility for all syllables are shown in Figure 23. It is quite remarkable that the calculated values are in good agreement with the tested values.

Therefore, it is concluded that the four orthogonal factors extracted from the *ACF* of the source signals and the sound-field signals may be significant to recognize speech. As indicated in Table 4, the most significant of the four factors is the effective duration of the *ACF*. Such evaluation of speech intelligibility considering human brain activity has not been considered in speech intelligibility studies.

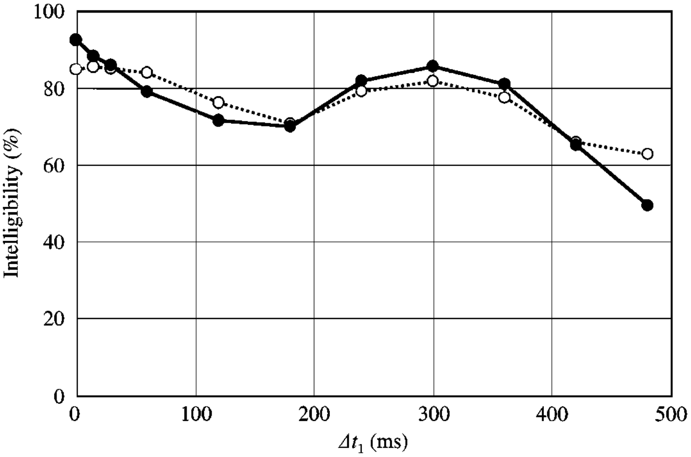

Figure 23. Mean values of calculated and tested intelligibilities for all syllables.

REFERENCES

1. Y. Ando 1985 *Concert Hall Acoustics*. Heidelberg: Springer-Verlag.
2. Y. Ando, K. Yamamoto, H. Nagamatsu and S. H. Kang 1991 *Acoustic Letters* **15**, 57–64. Auditory brainstern response (ABR) in relation to the horizontal angle of sound incidence.
3. Y. Ando, S. H. Kang and H. Nagamatsu 1987 *The Journal of the Acoustical Society of Japan* E **8**, 183–190. On the auditory-evoked potential in relation to the IACC of sound field.
4. Y. Ando, S. H. Kang and K. Morita 1987 *The Journal of the Acoustical Society of Japan* E **8**, 197–204. On the relationship between auditory-evoked potential and subjective preference for sound field.
5. Y. Ando and C. Chen 1996 *Journal of Architecture, Planning and Environmental Engineering* (*Transaction of AIJ*) **488**, 67–73. On the analysis of autocorrelation function of α-waves on the left and right cerebral hemispheres in relation to the delay time of single sound reflection.
6. C. Chen and Y. Ando 1996 *Journal of Architecture, Planning and Environmental Engineering* (*Transaction of AIJ*) **489**, 73–80. On the relationship between the autocorrelation function of the α-waves on the left and right cerebral hemisphere and subjective preference of the reverberation time of music sound field.
7. K. Nishio and Y. Ando 1996 *The Journal of the Acoustical Society of America* **100**, 2787. On the relationship between the autocorrelation function of the continuous brain waves and the subjective preference of sound field in change of the IACC.
8. Y. Ando 1998 *Architectural Acoustics—Blending Sound Sources, Sound Fields and Listeners*. New York: AIP Press/Springer-Verlag.
9. Y. Katsuki, T. Sumi, H. Uchiyama and T. Watahabe 1958 *The Journal of Neurophysiology* **21**, 569–588. Electric responses of auditory neurons in cat to sound stimulation.
10. P. Damaske and Y. Ando 1972 *Acustica* **27**, 232–238. Interaural crosscorrelation for multichannel loudspeaker reproduction.
11. T. Nakajima and Y. Ando 1991 *The Journal of Acoustical Society of America* **90**, 3173–3179. Effects of a single reflection with varied horizontal angle and time delay on speech intelligibility.
12. Y. Ando and Y. Kurihara 1986 *The Journal of Acoustical Society of America* **80**, 833–836. Nonlinear response in evaluating the subjective diffuseness of sound field.

13. P. K. Singh, Y. Ando and Y. Kurihara 1994 *Acustica* **80**, 471–477. Individual subjective diffuseness responses of filtered noise sound fields.
14. S. Sato and Y. Ando 1996 *The Journal of Acoustical Society of America* **100**, 2592. Effects of interaural crosscorrelation function on subjective attributes.
15. Y. Ando 1977 *The Journal of Acoustical Society of America* **62**, 1436–1441. Subjective preference in relation to objective parameters of music sound fields with a single echo.
16. S. H. Kang and Y. Ando 1985 *Memoirs of Graduate School of Science and Technology, Kobe University* **3**-A, 71–76. Comparison between subjective preference judgements for sound fields by different nations.
17. H. P. Seraphim 1961 *Acustica* **11**, 80–91. Ueber die Wahrnehmbarkeit mehrerer Rueckwuerfe von Sprachshall.
18. H. Haas 1951 *Acustica* **1**, 49–58. Ueber den Einfluss eines Einfachechos auf die Hoersamkeit von Sprache.
19. Y. Ando, S. Shidara and Z. Maekawa 1974 *Proceedings of the* 8*th International Congress of Acoustics* (London), 611. Simulation of sound propagation with boundary and subjective test.
20. Y. Ando and H. Alruz 1982 *Journal of Acoustic Society of America.* **71**, 616–618. Perception of coloration in sound fields in relation to the autocorrelation function.
21. I. Nakayama 1984 *Acustica* **54**, 217–221. Preferred time delay of a single reflection for performers.
22. S. Sato, Y. Ando and S. Ota 1999 *Journal of Sound and Vibration* (Special Issue on Opera House Acoustics). Subjective preference of cellists for the delay time of a single reflection in a performance.
23. K. Mouri, K. Akiyama and Y. Ando 1999 *Journal of Sound and Vibration* (Special Issue on Opera House Acoustics). Relationship between subjective preference and the α-brain wave in relation to the initial time delay gap with vocal music.
24. J. Atagi, Y. Ando and Y. Ueda 1999 *Journal of Sound and Vibration* (Special Issue on Opera House Acoustics). On the effects of time-variant sound fields on subjective preference.
25. H. Sakai, Y. Ando and H. Setoguchi 1999 *Journal of Sound and Vibration* (Special Issue on Opera House Acoustics). Individual subjective preference of listeners to vocal music sources in relation to the subsequent reverberation time of sound fields.
26. Y. Ando, S. Sato and H. Sakai 1999 *Computational Acoustics in Architecture* (J. J. Sendra Editor), Southampton: WIT Press, chaper 4. Fundamental Subjective Attributes of Sound Fields Based on the Model of Auditory-Brain System.
27. J. F. Schouten 1970 *Frequency Analysis and Periodicity Detection in Hearing, Leiden, Sijithoff, The Residue Revisited* (R. Plomp, G. F. Smoorenburg, editors), 41–58.
28. I. Gde. N. Merthayasa and Y. Ando *Unpublished.* Variation in the autocorrelation function of narrow-band noises: their effect on loudness judgement.
29. F. L. Wightman 1973 *Journal of the Acoustical Society of America* **54**, 407–416. The pattern-transformation model of pitch.
30. T. Houtgast, H. J. M. Steeneken and R. Plomp 1980 *Acustica* **46**, 60–72. Predicting speech intelligibility in rooms from the modulation transfer function. I. General room acoustics.
31. Y. Korenaga and Y. Ando 1996 *The Journal of the Acoustical Society of Japan* **52**, 940–947. A method of calculating intelligibility of sound field in relation to temporal structure of reflections—on the trend of syllable confusion under sound fields composed of a direct sound and up to two reflections (in Japanese).
32. T. Shoda and Y. Ando 1998 *Proceedings of the* 16*th International Congress on Acoustics*, Seattle, 2163–2164. Calculation of speech intelligibility using four orthogonal factors extracted from the autocorrelation function of source and sound field signals, which is cited in Table 4.