Journal of Sound and Vibration (2001) **241**(1), 105–116
doi:10.1006/jsvi.2000.3282, available online at http://www.idealibrary.com on IDEAL®

THE FREQUENCY RANGE APPLICABLE TO PITCH IDENTIFICATION BASED UPON THE AUTO-CORRELATION FUNCTION MODEL

M. Inoue and Y. Ando

Graduate School of Science and Technology, Kobe University, Rokkodai, Nada, Kobe 657-8501, *Japan. E-mail*: 995t007n@kobe-u.ac.jp

and

T. Taguti

Faculty of Science, Konan University, Kobe, Okamoto, Higashinada 658-8501, *Japan*

(*Accepted* 7 *August* 2000)

Pitch identification experiments were conducted to determine the upper boundary of the frequency range within which the autocorrelation function (ACF) model described by Ando (see Reference [16]) is applicable to the phenomenon of the missing fundamental of a complex tone. Pitch-matching tests were conducted using: (1) complex tones consisting of $2f$, $3f$, and $4f$ (where f denotes the fundamental frequencies of 0·5, 1·0, 1·2, 1·6, 2·0, or 3·0 kHz) and (2) complex tones consisting of $2f$ and $3f$. These tests showed that, for both kinds of complex tones, the ACF model can account for the perception of pitch when the fundamental frequency is below 1200 Hz.

1. INTRODUCTION

The primary sensations of sound are timbre, loudness, and pitch. Pitch is related to the frequency of a pure tone, and to the fundamental frequency of a periodic complex tone. One finds periodic complex tones, for example, in the environmental noise, voices, and instrumental music.

The phenomenon of the missing fundamental is something that needs to be accounted for by a model of pitch perception. That is, it must account for the fact that the pitch of harmonic components without the fundamental frequency is perceived as being the same as the pitch of a pure tone of the fundamental frequency. The perceived frequency of a tone consisting of harmonic components without the fundamental frequency is called the residue pitch. Although many models of pitch perception for stimuli in the auditory organ, including place theory and temporal theory, have been proposed, none of them provide a precise mechanism by which the residue pitch can be determined. The autocorrelation function (ACF) model of pitch perception was originally developed from a "duplex" model [1] based on a network of delay lines and coincidence counters arranged along the axes of frequency and delay. An ACF was used in the network. However, this model has not received much attention, because it was very abstract and lacked a precise rule that could be used to estimate a pitch, and is too conceptual. The model was improved, however, by adding new rules [2, 3].

0022-460X/01/110105 + 12 $35.00/0

According to the fine-structure theory [4], a pitch is derived from the time interval between the peaks of the waveform of a stimulus. This theory is related to the pattern of nerve firing, which involves the phase-locking effect. Such a theory, based on the temporal information concerning nerve firing, can be applied only in a range of less than about 5 kHz. The interference between higher harmonics without dividing each other is necessary in this theory. The range of the interference differs from that in the dominance region [5]. Waveforms change with changing phase relationships, but the perceived pitch is invariable [6], and this cannot be accounted for by the fine-structure theory. Three models called "pattern recognition" models were proposed to improve the fine-structure theory and these were generally accepted for many years [6–8]. In order to overcome the difficulty in the previous models—that the phase relation influences the pitch perception—a pattern-transformation model was proposed [6, 9]. This model is based on the concept that pitch is generally insensitive to phase changes. In the model, a stimulus is roughly divided into separate frequencies by the auditory system. Then these outputs are translated into other patterns, one in each frequency band. A pitch transformer based on an ACF detects the location of peaks in the output waveform. The residue pitch can be estimated from the distance between these peaks, and the pitch strength can be estimated from the height of the peak of the stimuli. In an experiment with complex tones and changing the phase relation [10], octave changes in the pitch were also observed. Virtual pitch theory based on a learning matrix was also proposed [8]. According to that theory, the residue pitch can be estimated by extracting the virtual pitch and spectral pitch from a matrix. In the optimum processor theory [7], in which an optimum central processor predicts the pitch of complex tones, information on the frequency of each component—but not on amplitude or phase —is conveyed to the processor. These models all assume the existence of information-processing mechanisms that can account for the rough spectral analysis in the spectral system and for the pattern recognition in the central system. They are also based on the assumption that the spectral analysis of each partial tone of the phase relation is not related to the pitch perception.

A similar phenomenon has been observed even in non-periodic complex tones as well as in some types of noises. The effectiveness of the pattern-transformation model was tested by using ripple noises as the peripheral weighting model [11–13]. It has also been suggested that the pitch strength can be predicted from the range of the weighting dominance region [5] by the use of filtering as well as with the pattern-transformation model.

In this study, the fine ACF structures shown are compared with the pitch matching data. A cancellation model involving an array of delay lines and inhibitory gating neurons has also been proposed as an extension of the ACF model [14]. It seems reasonable, however, to suppose that both spatial and temporal cues are related to the perception of residue [15].

When one uses bandwidth noises, it is thought that the pitch can be calculated by ACF analysis. This is because the pitch can be predicted from the value of τ_1 (see section 2). Also the pitch strength relies on the ϕ_1, and the pitch is strongly perceived for the largest ϕ_1. The ACF model is applicable to the prediction of the pitch not only of complex tone and ripple noise but also of complex noise without a fundamental frequency [16, 17]. The purpose of the present experiment is to find the frequency range applicable to pitch identification by the ACF model.

2. AUTOCORRELATION FUNCTION (ACF) MODEL

The ACF model is, equivalent to the power density spectrum, insensitive to the relative phases [16].

The ACF $\Phi(\tau)$ is expressed by

$$\Phi(\tau) = \frac{1}{2T}\int_{-T}^{T} p(t)p(t+\tau)\,\mathrm{d}t, \tag{1}$$

where $p(t)$ is a sound signal at the entrances of the ears of a listener, τ is the time delay, and $2T$ is the integral interval. The normalized ACF (NACF), $\phi(\tau)$, is defined as

$$\phi(\tau) = \frac{\Phi(\tau)}{\Phi(0)}, \tag{2}$$

where $\Phi(0)$ is an average level of the signal strength for the signal interval of $2T$.

The factors of τ_1 and ϕ_1 extracted from the NACF are used to calculate the pitch. In this model, it is thought that, τ_1, the time delay at the first maximum peak of the NACF, has a close relation with the pitch one perceives.

3. EXPERIMENTAL METHOD

3.1. SOURCE STIMULI

For fundamental frequencies of 500, 1000, 1200, 1600, 2000, and 3000 Hz, stimuli consisting of two or three pure tone components were produced in a computer. The two-component stimuli consisted of the second and third harmonics of the fundamental frequency, and the three-component stimuli consisted of the second, third, and fourth harmonics (see Table 1). The starting phase of all components was adjusted at zero. The total sound pressure level at the center of the listener's head was fixed at 74 dB.

3.2. ACF ANALYSIS

The NACFs of all stimuli were calculated and some examples are shown in Figure 1. Each NACF has peaks τ_1 related to the fundamental frequency. So if the subjects perceive the pitches of complex tones in accordance with the ACF model, they should perceive the pitches of all stimuli to be around the fundamental frequencies.

TABLE 1

Stimuli used for the experiments; stimuli consist of two or three pure tones with fundamental frequencies, 500, 1000, 1200, 1600, 2000 *and* 3000 Hz

Fundamental frequency (Hz)	Stimuli with two components (Hz)	Stimuli with three components (Hz)
500	1000, 1500	1000, 1500, 2000
1000	2000, 3000	2000, 3000, 4000
1200	2400, 3600	2400, 3600, 4800
1600	3200, 4800	3200, 4800, 6400
2000	4000, 6000	4000, 6000, 8000
3000	6000, 9000	6000, 9000, 12000

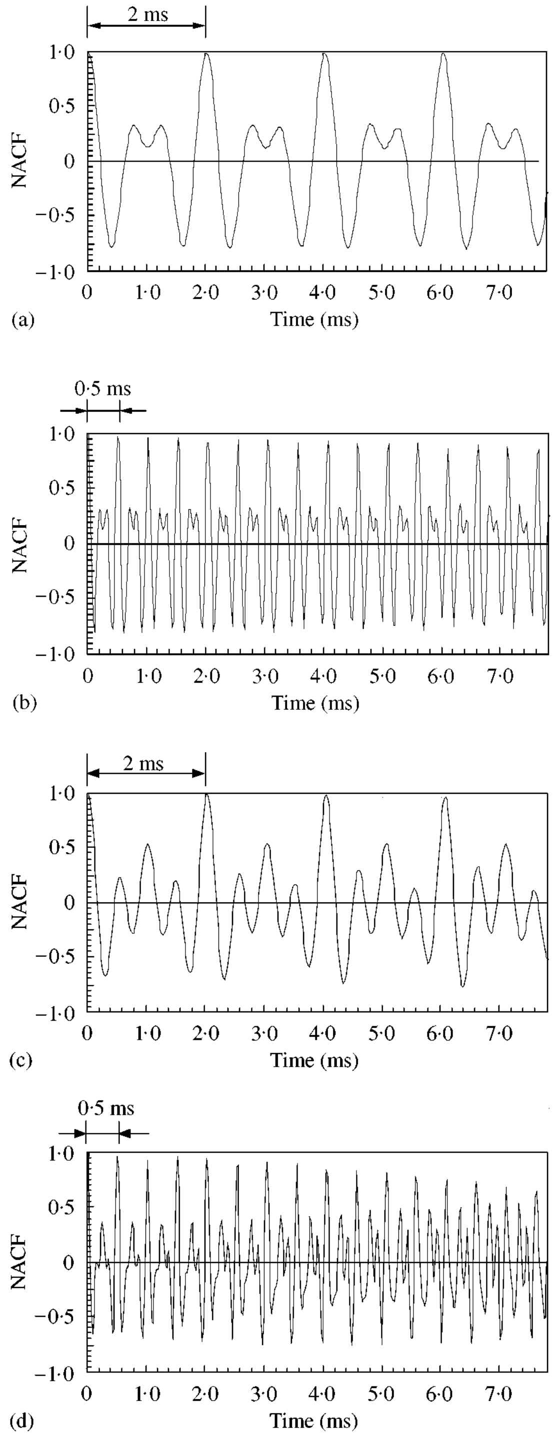

Figure 1. Examples of the normalized autocorrelation function (NACF) calculated for various stimuli: (a) The fundamental frequency is 500 Hz and there are two components; (b) the fundamental frequency is 2000 Hz and there are two components; (c) the fundamental frequency is 500 Hz and there are three components; and (d) the fundamental frequency is 2000 Hz and there are three components.

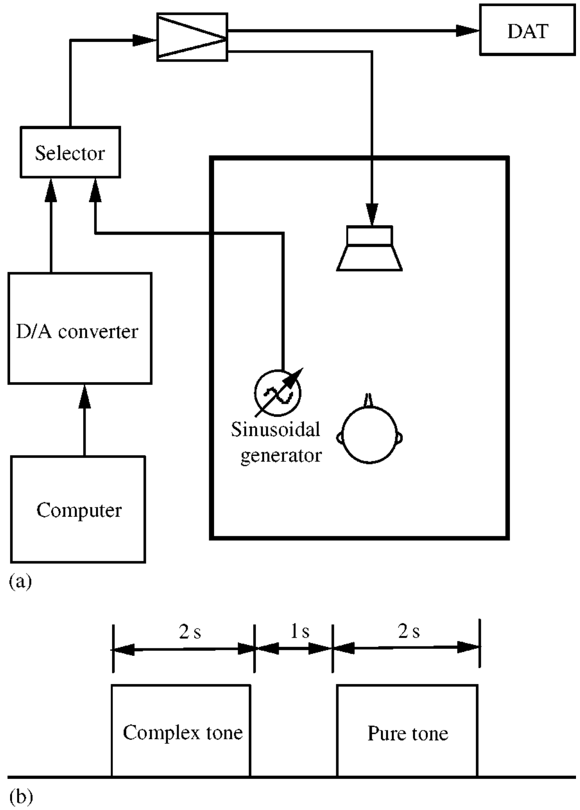

Figure 2. (a) Experimental set-up for the subjective pitch matching test. (b) Stimulus presentation order: complex tone for test stimuli, pure tone for reference.

3.3. APPARATUS

Figure 2(a) illustrates the apparatus used. A loudspeaker was placed in front of a subject in an anechoic chamber. The distance between the center of the subject's head and the loudspeaker was 0·8 m.

3.4. SUBJECTS

The three experimental subjects had musical experience (2 male and 1 female) between 21 and 27 years old (Subjects A–C).

3.5. EXPERIMENT

Pitch-matching tests were conducted using complex tones as test stimuli and a pure tone generated by a sinusoidal generator as a reference. Test stimuli were generated by a computer and reference tones were generated by a sinusoidal generator. Both were delivered by a loudspeaker.

As shown in Figure 2(b), 1 s after a complex tone was presented for 2 s, a pure tone was presented for 2 s. The initial frequency of the pure tone was lower than the fundamental frequency of the test stimulus and, while the pure tone was being presented, the subjects adjusted its frequency until its pitch was perceived as identical to that of the complex tone.

4. RESULTS

4.1. RESULTS FOR EACH SUBJECT

The experimental results for Subject A are shown in Figure 3. The results for Subjects B and C are, respectively, shown in Figures 4 and 5. When the fundamental frequencies of

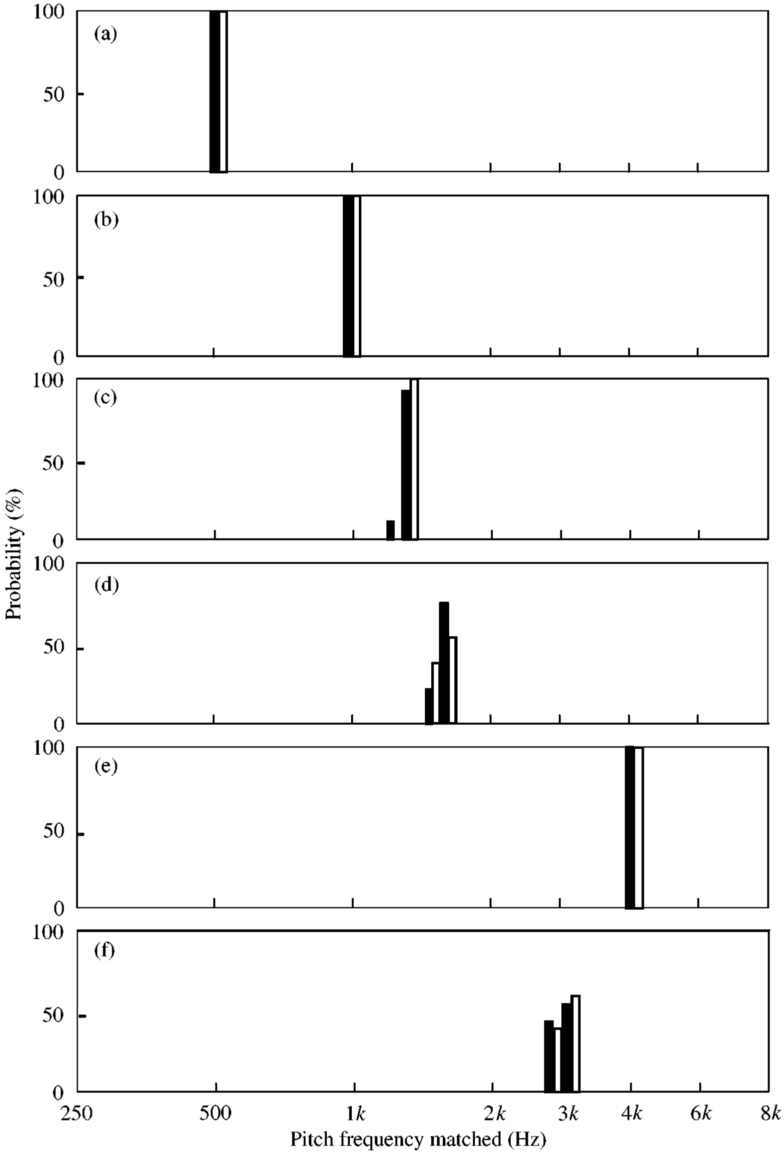

Figure 3. Pitch-matching test results for Subject A. Fundamental frequency: (a) 500 Hz, (b) 1000 Hz, (c) 1200 Hz, (d) 1600 Hz, (e) 2000 Hz and (f) 3000 Hz. Different colors indicate results obtained under different conditions: black for two components, and white for three components.

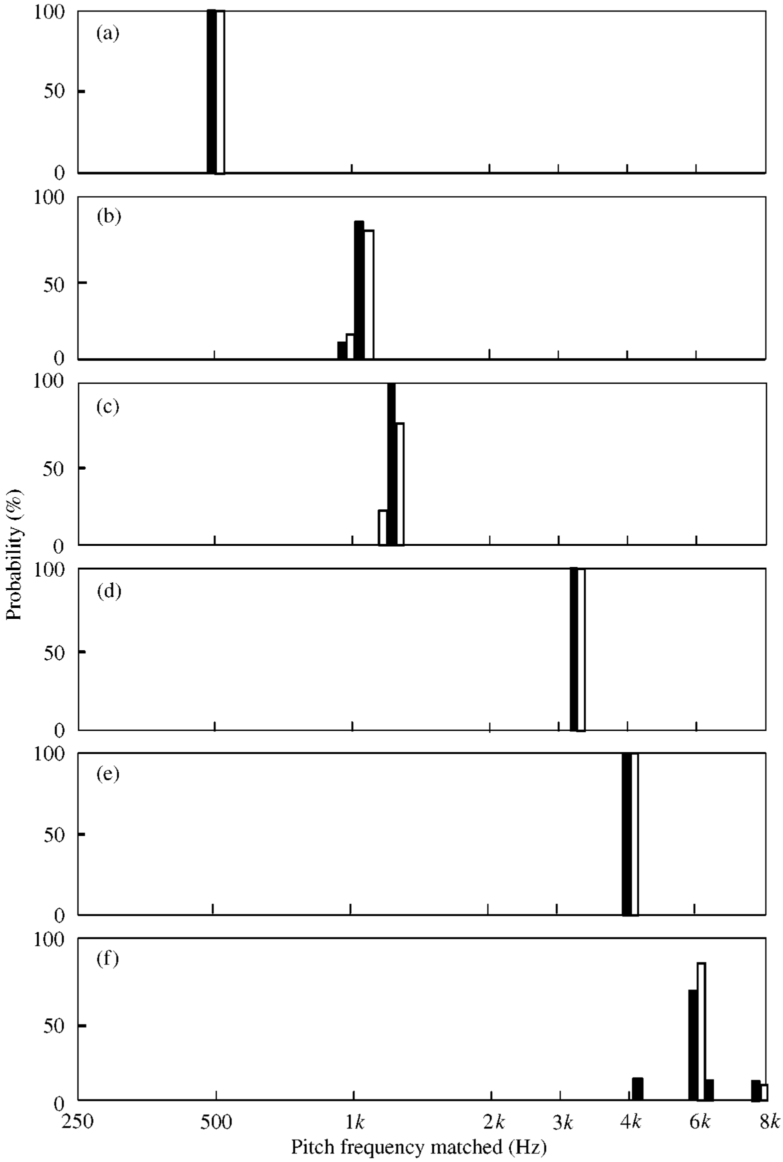

Figure 4. Pitch-matching test results for Subject B. Fundamental frequency: (a) 500 Hz, (b) 1000 Hz, (c) 1200 Hz, (d) 1600 Hz, (e) 2000 Hz and (f) 3000 Hz. Different colors indicate results obtained under different conditions: black for two components, and white for three components.

the stimuli were 500, 1000 or 1200 Hz, more than 90% of the responses obtained under both conditions (two and three components) clustered around each fundamental frequency for all subjects. When the fundamental frequency was 1600 Hz, the responses obtained under both conditions clustered around 1600 Hz for Subject A, although the responses for Subjects B and C clustered around 3200 Hz. When the fundamental frequency was 2000 Hz, the responses obtained under both conditions clustered around 4000 Hz for Subjects A and B. For Subject C, however, 60% of the responses obtained when the stimulus had two components clustered around 2000 Hz and 40% of the responses clustered around 4000 Hz.

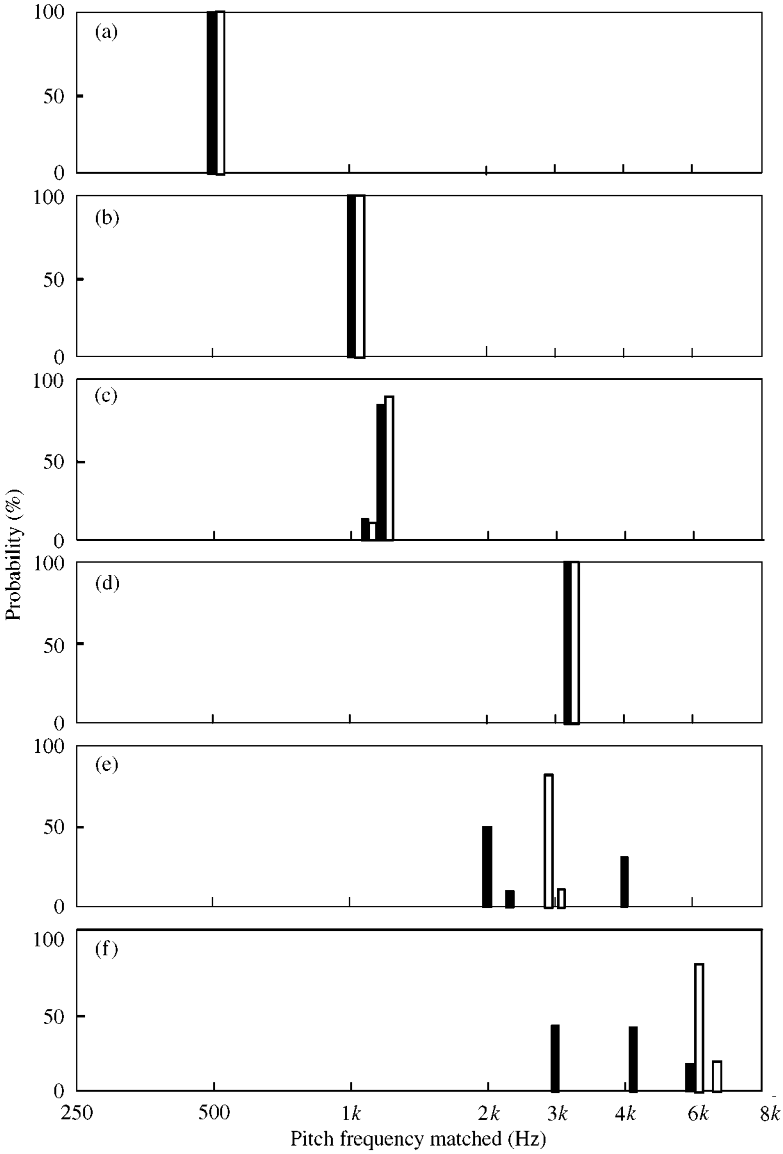

Figure 5. Pitch-matching test results for Subject C. Fundamental frequency: (a) 500 Hz, (b) 1000 Hz, (c) 1200 Hz, (d) 1600 Hz, (e) 2000 Hz and (f) 3000 Hz. Different colors indicate results obtained under different conditions: black for two components, and white for three components.

For the condition with three components, 80% of the responses clustered around 2000 Hz, and when the fundamental frequency was 3000 Hz, the responses obtained under two conditions clustered around 3000 Hz for Subject A. For Subject B, about 70% of the responses obtained when the stimulus had two components clustered around 6000 Hz, 15% clustered around 4005 Hz, and the rest clustered around 8009 Hz. For the condition with three components, 90% of the responses were clustered around 6000 Hz and the rest clustered around 8009 Hz. For Subject C, about 40% of the responses obtained when the stimulus had two components clustered around 3000 Hz, 40% clustered around 5046 Hz, and 20% clustered around 6000 Hz. For the condition with three components, the responses clustered around 6000 Hz.

4.2. TOTAL RESULTS

The results for all subjects are shown in Figure 6. Whenever the fundamental frequency of the stimulus was 500, 1000 or 1200 Hz, more than 90% of the responses obtained from all

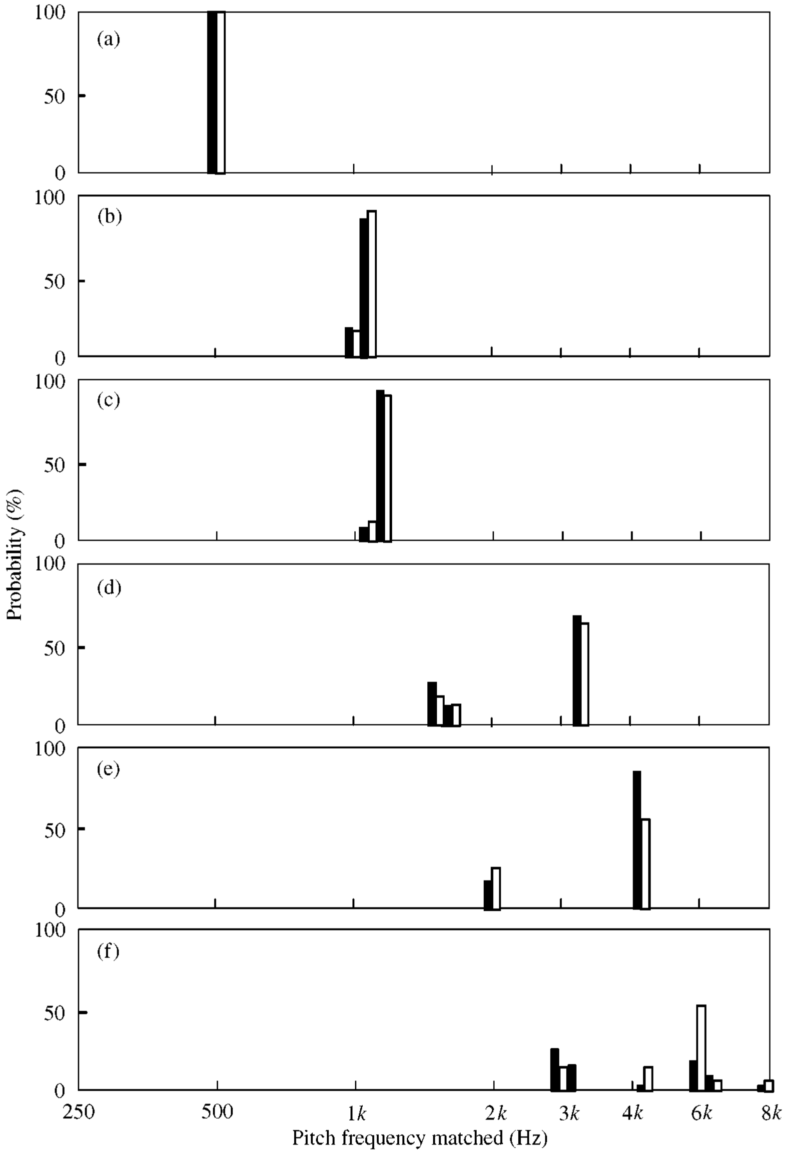

Figure 6. Pitch-matching test results for all three subjects. Fundamental frequency: (a) 500 Hz, (b) 1000 Hz, (c) 1200 Hz, (d) 1600 Hz, (e) 2000 Hz and (f) 3000 Hz. Different colors indicate results obtained under different conditions: black for two components, and white for three components.

subjects under both conditions clustered around the fundamental frequency. When the fundamental frequency of the stimuli were 1600, 2000, or 3000 Hz, however, the responses varied from subject to subject.

5. DISCUSSION

The results of these experiments indicate that when the fundamental frequency is 500, 1000 or 1200 Hz, regardless of the number of stimulus components the perceived matching pitch is almost the same as the fundamental frequency calculated from the period of the NACF. When the fundamental frequencies were 1600, 2000 and 3000 Hz, however, the probability that the subjects adjusted the frequency of the pure tone to the frequencies close to the calculated fundamental frequencies is extremely decreased (see Figure 7). Some responses, however, clustered around the fundamental frequencies (see Figures 3–5), even when the fundamental frequency was 1600, 2000 or 3000 Hz. This is because there are individual differences in the way the subjects perceived pitch. And when the fundamental frequencies are 1600 and 3000 Hz, Subject A judged slightly higher than the fundamental frequency. This is because the pitch of the pure tone is perceived to be a little bit lower as the SPL increased [18], although the pitch of the complex tone was not influence by SPL [19]. As described in section 3.2, the NACF for any stimulus has a delay time of the first peak τ_1 related to the fundamental frequency of that stimulus. So the pitches of all stimuli should be perceived as their fundamental frequencies [17]. The present results, however, showed that most of the subjects perceived the pitch of the stimulus to be the same as the fundamental frequency calculated from the interval between the peaks of NACF of the stimulus only when the fundamental frequencies were below 1200 Hz. These results indicate that the ACF model is applicable when the fundamental frequencies of stimuli are below 1200 Hz. It is considered that this is because nerve cells have an absolute refractory period of about 1 ms limiting the maximum firing rate.

Even when the fundamental frequencies are below 1200 Hz, the fundamental frequency is not included in the spectrum but in the NACFs. Thus, the method of measuring not by the use of band-pass filters but by the ACF model can obtain factors determining this type of pitch.

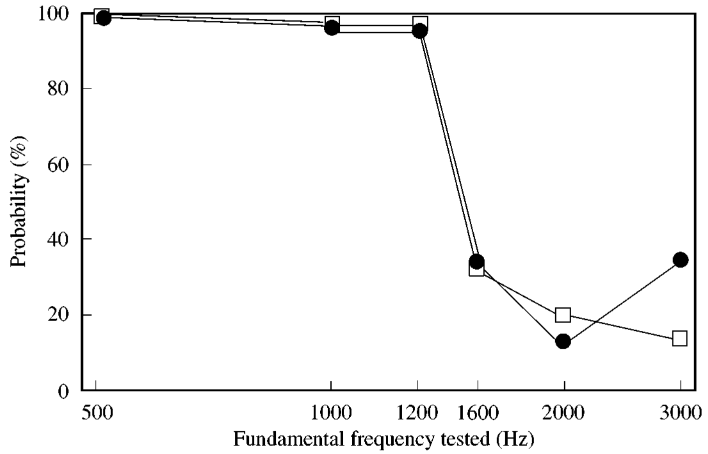

Figure 7. Probability that all three subjects adjusted pure tone to a frequency of the complex tone near the fundamental frequency. Different symbols indicate results obtained under different conditions: black circle for two components, and white square for three components.

When the fundamental frequency of the stimulus was 3000 Hz, some responses of Subject B clustered around 4005 and 8009 Hz and some responses of Subject C clustered around 5046 Hz. These frequencies were neither part of the test stimulus nor integer multiples of the fundamental frequency, and this would be hard to predict from the peaks of the NACF.

Because the ACF model cannot account for the perception of pitch when the fundamental frequency of the stimulus is greater than 1600 Hz, it seems reasonable to suppose that temporal cues are related to the perception of the residue pitch [15]. From 1000 to 2000 Hz, the precision of phase-locking decreases. To compare the results of this experiment and that, it is considered that the applicable frequency range of ACF model is very likely to correspond to the range from 1000 to 1200 Hz, where the phase-locking occurs with high precision.

Although there are individual differences in perceived pitch, the ACF model is applicable to complex tones that have a fundamental frequency below 1200 Hz.

6. CONCLUSION

We have come to the two following conclusions.

(1) The ACF model is applicable when the fundamental frequency is below 1200 Hz.
(2) Within this range, even if the stimulus has the fewest number (two) of components, pitch can be predicted by the ACF model.

REFERENCES

1. J. C. R. Licklider 1951 *Experimentia* **7**, 128–134. A duplex theory of pitch perception.
2. R. Meddis and M. J. Hewitt 1991a *The Journal of the Acoustical Society of America* **89**, 2866–2882. Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: phase identification.
3. R. Meddis and M. J. Hewitt 1991b *The Journal of the Acoustical Society of America* **89**, 2883–2894. Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II: phase sensitivity.
4. J. F. Schouten 1970 *Leiden* 41–58. Sijithoff. *The residue revisited*, in (R. Promp and G. F. Smoorendurg editors). Frequency analysis and periodicity detection in hearing.
5. R. J. Ritsuma 1967 *The Journal of the Acoustical Society of America* **42**, 191–198. Frequencies dominant in the perception of pitch complex sounds.
6. F. L. Wightman 1973a *The Journal of the Acoustical Society of America* **54**, 397–406. Pitch and stimulus fine structure.
7. J. L. Goldstein 1973 *The Journal of the Acoustical Society of America* **54**, 1496–1516. An optimum processor theory for the central formation of the pitch of complex tones.
8. E. Terhardt 1974 *The Journal of the Acoustical Society of America* **55**, 1061–1069. Pitch, consonance, and harmony.
9. F. L. Wightman 1973b *The Journal of the Acoustical Society of America* **54**, 407–416. The pattern-transformation model.
10. C. Lundeen and A. M. Small Jr. 1984 *The Journal of the Acoustical Society of America* **75**, 1578–1587. The influence of temporal cue on the strength of periodicity pitches.
11. W. A. Yost, R. Hill and T. Perez-Falcon 1978 *The Journal of the Acoustical Society of America* **63**, 1166–1173. Pitch and pitch discrimination of broadband signals with rippled power spectra.
12. W. A. Yost 1996a *The Journal of the Acoustical Society of America* **99**, 1066–1078. A time domain description for the pitch strength of iterated rippled noise.
13. W. A. Yost 1996b *The Journal of the Acoustical Society of America* **100**, 511–518. Pitch of iterated noise.
14. A. Cheveigne 1998 *The Journal of the Acoustical Society of America* **103**, 1261–1271. Cancellation model of theory for the central formation of the pitch of complex tones.

15. K. OHOGUSI 1978 *The Journal of the Acoustical Society of America* **64**, 764–771. On the role of spatial and temporal cues in the perception of the pitch of complex tones.
16. Y. ANDO 1998 *Architectural Acoustics–Blending Sound Sources, Sound Fields, and Listeners*. New York: AIP/Springer-Verlag.
17. T. SUMIOKA and Y. ANDO 1997 *Memoirs of the Graduate School of Science and Technology, Kobe University*. Pitch identification using the autocorrelation function model.
18. S. S. STEVENS 1935 *The Journal of the Acoustical Society of America* **6**, 150–154. The relation of pitch to intensity.
19. E. ZWICKER and H. FASTL 1990 *Psychoacoustics*. Berlin: Springer-Verlag; chapter 5.