Theory of Auditory Temporal and Spatial Primary Sensations[1]

Yoichi Ando
1-4-132-105 Hiyodoridai, Kita, Kobe 651-1123 Japan

(Received 23 January 2008; accepted 12 December 2008)

Getting results of a number of previous papers into together, this paper proposes a theory of auditory temporal and spatial primary sensations. The model consists of the autocorrelation function (ACF) mechanism and the interaural crosscorrelation function (IACF) mechanism for signals arriving at two ear entrances, and the specialization of human cerebral hemispheres. Evidences of this auditory model have been found by recording the auditory brainstem response (ABR), the slow vertex response (SVR), the electroencephalogram (EEG) and the magnetroencephalogram (MEG). It may describe four temporal primary sensations including pitch or missing fundamental, loudness, timbre, and in addition duration sensation. These four temporal sensations may be formulated by the temporal factor, which may be extracted from the ACF associated with left hemisphere. Three spatial sensations consisting of localization in the horizontal plane, apparent source width (ASW) and subjective diffuseness are described by the spatial factor extracted from the IACF associated with the right hemisphere.

Keywords: temporal primary sensations and spatial primary sensations, temporal and spatial factors extracted from ACF and IACF, specialization of left and right cerebral hemispheres.

1. INTRODUCTION

The purpose of this paper is to propose a theory of auditory temporal and spatial primary sensations by getting results of a number of previous papers into together. The neurally grounded theory of subjective preference for sound fields in concert halls that is based on the model of human auditory system has been previously described (Ando, 1985, 1998, 2007a). Most generally, subjective preference itself is regarded as a primitive response of the living creature that entail judgments that steer an organism in the direction of maintaining life, so as to enhance prospects for survival. Thus, brain activities corresponding to subjective preference have been well observed (Ando, 2003). The auditory preference model consists of two kinds of internal representations of sound that are based on the correlation structure of the sound as it presents itself to the two ears. The autocorrelation function (ACF) describes the internal correlation structure of the monaural signal at each of the two ears, while the interaural cross correlation function (IACF) describes the correlations between the signals arriving at two ear entrances. The activities corresponding to the magnitude of IACF (IACC) has been found by records of the auditory brainstem responses (Ando, Yamamoto, Nagamatsu, and Kang, 1991). The ACF mechanism has a firm neural basis in the temporal patterning of spike activity in early stages of auditory processing (Cariani, 1996a; 1996b; 2001).

Since Helmholz it has been well appreciated that the cochlea carries out a rough spectral analysis of sound signals that manifests itself in coarse tonotopically organized spatial patterns of neuronal response from auditory nerve to cortex. The resulting auditory frequency maps, however, are relatively coarse; this "place principle" is capable of frequency distinctions across octaves, but very much poorer at distinguishing tones within a given octave. Time domain representations complement those based on frequency, and it has also been appreciated for almost a century that neural patterns of discharge faithfully follow the time structure of the acoustic stimulus. Temporal patterns of spike activity provide time-domain representations of the stimulus that can be analyzed by the central auditory system. The auditory representations from cochlea to cortex that have been found to be related to subjective preference of the sound field in a deep

[1]The main part of this paper was presented at the 9th Western Pacific Acoustics Conference (WESPAC IX 2006), Seoul, Korea, 26-28 June 2006.

way involve these temporal response patterns, which have a character much different from those related to power spectrum analysis.

The ACF and IACF mechanisms existing in auditory processing subserve tonal quality and auditory space perception. Many aspects of tonal quality, including pitch, loudness, timbre and duration may be described in terms of temporal factors extracted from the autocorrelation representation. On the other hand, spatial sensations such as localization in the horizontal plane, apparent source width (ASW) and subjective diffuseness may be described in terms of spatial factors extracted from the crosscorrelation representation. Consequently, so called the "primary sensation" may be classified into two categories: temporal primary sensations and spatial primary sensations. Any subjective responses of the sound signal and the sound field, therefore, may be well described based on both the temporal and spatial factors. These factors appear to be associated with processing in different cerebral hemispheres. The temporal factors extracted from the ACF appear to be associated with the left cerebral hemisphere, while the spatial factors extracted from the IACF appear to be associated with the right hemisphere as summarized in this journal (Ando, 2003). Such a hemispheric specialization, and thus the independence of the temporal and spatial processing, that they engender may play important roles in shaping the structure of any subjective responses.

2. MODEL OF HUMAN AUDITORY SYSTEM

2.1. Summary of Neural Evidences

The auditory-brain model is based on the following evidences.

A. Physical system

First of all, one is interested in the fact that the human ear sensitivity to the sound source in front of the listener is essentially formed by the physical system from the source point to the oval window of cochlea (Ando, 1985, 1998). The transfer function of such a cascade system includes the human head and pinnae, the external canal and the eardrum and borne chain and excludes a feedback system. For the sake of practical convenience, the A-weighting network may be utilized representing the ear sensitivity.

B. Auditory brainstem response (ABR) from the left and right auditory pathway

The records of left and right auditory brainstem responses (ABR) implies that (APPENDIX: Ando, et al, 1991):

1) Amplitudes of wave $I_{l,r}$ & $III_{l,r}$ correspond roughly to the sound pressure level at two ear entrances as a function of the horizontal angle of incidence to the listener (ξ).
2) Amplitudes of waves $II_{l,r}$ & $IV_{l,r}$ correspond roughly to the sound pressure levels as a function of the contra horizontal angle ($-\xi$).
3) Analysis of activities of waves $IV_{l,r}$, and wave V averaged at the inferior colliculus, correspond well with the value of IACC.

C. Slow vertex response (SVR) on the left and right hemisphere

Recording left and right SVR has discovered following facts (Ando, 1992):

4) The left and right amplitudes of the early SVR, the amplitude difference between amplitudes of first positive and negative $A(P_1-N_1)$ indicate that the left and right hemispheric dominance, respectively, are due to the temporal factor: the initial time delay gap between the direct sound and the first reflection, Δt_1, and spatial factors: the sensation level, SL and the magnitude of the IACF defined by Eq. (5), IACC. At first, it was considered that the SL or the listening level (LL) was classified as a temporal-monaural factor from a physical viewpoint. However, results of SVR indicate that SL is the right hemisphere dominant. Reasons are that binaural LL is measured by the geometric average of sound energies arriving at the two ears, and the left hemisphere is not interested in such a factor without any information and/or meaning in its temporal sequence.
5) Both left and right latencies of the second N_2 correspond well to the IACC, and thus these are well related to the scale value of subjective preference as an overall-primitive response.

D. Electroencephalogram (EEG) on left and right hemispheres

Analyses of EEG recorded on left and right cerebral hemispheres reconfirm that (Ando, 1998):

6) Δt_1 and the subsequent reverberation time, T_{sub}, indicate left hemisphere dominance, and the IACC response indicates right hemisphere dominance. Thus, a high degree of independence between the left and right hemispheric factors may be resulted.
7) The scale value of subjective preference corresponds well to the value of τ_e, which is extracted from ACF of the α-wave in EEG, when one of the temporal and spatial factors of the sound field is varied. At the preferred environmental condition, brain would like to enjoy for a long time corresponding to τ_e of the α-wave.

E. Magnetroencephalogram (MEG) on the left and right hemisphere

Analyses of MEG recorded on left and right cerebral hemispheres reconfirm that (Soeta, et al., 2002):

8) Amplitudes of MEG recorded when Δt_1 is changed show the left hemisphere specialization.
9) The scale value of individual subjective preference is directly related to the value of τ_e of extracted from ACF of the α-wave of MEG in the left hemisphere.

So far, it has been found that LL and IACC are associated with the right cerebral hemisphere, and the temporal factors, Δt_1 and T_{sub}, the sound field in a room are associated with the left (Table 1).

Table 1. Hemispheric specialization of the temporal and the spatial factors determined by analyses of AEP (SVR), EEG and MEG.

Factors changed	AEP(SVR) $A(P_1\text{-}N_1)$	EEG, ratio of ACF τ_e values of α-wave	MEG, ACF τ_e value of α-wave
Temporal			
Δt_1	L > R (speech)[1]	L > R (music)	L > R (speech)
T_{sub}	---	L > R (music)	---
Spatial			
LL	R > L (speech)	---	---
IACC	R > L (vowel /a/)	R > L (music)[2]	R > L[3]
	R > L (band noise)		
τ_{IACC}			R > L[3]

[1] Sound source used in experiments is indicated in the bracket.

[2] Flow of EEG α-wave from the right hemisphere to the light hemisphere for music stimuli in change of the IACC was determined by the CCF $|\phi(\tau)|_{max}$ between α-waves recorded at different electrodes.

[3] Soeta and Nakagawa (2006).

2.2. Model

Based on above mentioned physical system and evidence of physiological responses to pitch and temporal structure (Cariani, 1996a; 1996b; 2001), a model of the central auditory signal processing system may be formed for the comprehensive temporal and spatial factors. The model consists of the autocorrelation mechanisms, the interaural crosscorrelation mechanism between the two auditory pathways, and the specialization of human cerebral hemispheres for temporal and spatial factors of the sound field. In addition, according to the relationship of subjective preference and physiological phenomena in changes with variation to the temporal and spatial factors (Ando, 2003), a model may be proposed as shown in Fig. 1. The specialization of the human cerebral hemisphere may relate to the highly independent contribution between the spatial and temporal factors on any subjective attributes. For example, "cocktail party effects" might well be explained by such specialization of the human brain, because speech is processed in the left hemisphere with all of temporal factors extracted from the ACF, and independently the spatial factors extracted from the IACF are mainly processed in the right hemisphere.

3. TEMPORAL AND SPATIAL FACTORS

3.1. Temporal Factors

The running ACF of the sound signals $f_l(t)$ and $f_r(t)$ at the left and the right ear entrances, respectively, is defined by,

$$\Phi_{ll}(\tau) = \frac{1}{2T}\int_{-T}^{+T} f_l'(t) f_l'(t+\tau)dt$$
$$\Phi_{rr}(\tau) = \frac{1}{2T}\int_{-T}^{+T} f_r'(t) f_r'(t+\tau)dt \qquad (1)$$

where $f_{l,r}'(t) = f_{l,r}(t)*s(t)$, s(t) being the ear sensitivity, which is essentially formed by the transfer function of physical system of the ear (Ando, 1985, 1998). Rigorously, individual sensitivity represented by s(t) is different, for the purpose of practical use, however, s(t) may be chosen as the impulse response of an A-weighted network. For frontal incidence of the sound signal, it yields $\Phi_{ll}(\tau) \approx \Phi_{rr}(\tau) \equiv \Phi(\tau)$, and $\phi(\tau) = \Phi(\tau)/\Phi(0)$, being $\phi(0) = 1$.

From the ACF analyzed, temporal factors are extracted, which may play important role for temporal primary sensations:

1. Energies at two ear entrances represented are given by $\Phi_{ll}(0)$ and $\Phi_{rr}(0)$.
2. As shown in Fig. 2a, the width of amplitude $\phi(0)$ around the origin of the delay time defined at a value of 0.5, is $W_{\phi(0)}$, according to the fact that $\phi(\tau)$ is en even function.
   In the fine structures including peaks and delays, for instance, τ_1 and ϕ_1 are the delay time and the amplitude of the first maximum of the ACF. Then, there are τ_n and ϕ_n being the delay time and the amplitude of the n-th local peak (n >1). Usually there are certain correlations between τ_n and τ_{n+1}, and between ϕ_n and ϕ_{n+1}, thus, the most significant factor can obtain at the first minimum, so that:
3. Factor τ_1: the delay time of ACF at the maximum; and
4 factor ϕ_1: the amplitude, at which τ_1 is observed.
5. The effective duration of the envelope of the normalized

ACF, τ_e, is defined by the tenth-percentile delay, and represents a repetitive feature containing the direct sound signal itself. This factor has determined the most preferred temporal factors of the sound field in a concert hall (Ando, 1985, 1998, 2007a).

The normalized ACF is defined by

$$\phi_p(\tau) = \Phi_p(\tau)/\Phi_p(0) \tag{2}$$

As in the manner shown in Fig. 2b, τ_e is obtained by fitting a straight line for extrapolation of delay time at –10 dB, if the initial envelope of the ACF decays exponentially.

It is worth noticing that the auditory temporal window recommended is given by (Mouri, Akiyama and Ando, 2001),

$$[2T]_{\text{recommend}} \approx 30(\tau_e)_{\min} \tag{3}$$

where $(\tau_e)_{\min}$ is the minimum value of τ_e extracted from the running ACF of the source signal.

3.2. Spatial Factors

The running IACF for the sound pressures, $f_{l,r}(t)$, arriving at the two-ear entrances is expressed by,

$$\Phi_{lr}(\tau) = \frac{1}{2T}\int_{-T}^{+T} f_l'(t)\, f_r'(t+\tau)dt$$

The normalized IACF in the possible range of maximum interaural delay times is expressed by,

$$\phi_{lr}(\tau) = \Phi_{lr}(\tau)/[\Phi_{ll}(0)\Phi_{rr}(0)]^{1/2} \tag{4}$$

where $\Phi_{ll}(0)$ and $\Phi_{rr}(0)$ are the ACFs at $\tau = 0$, or sound energies arriving at the left- and right-ear entrances, respectively.

From the IACF analysis, four spatial factors are determined (Fig. 3):

1. The magnitude of the interaural cross-correlation is defined by,

$$\text{IACC} = |\phi_{lr}(\tau)|_{\max} \tag{5}$$

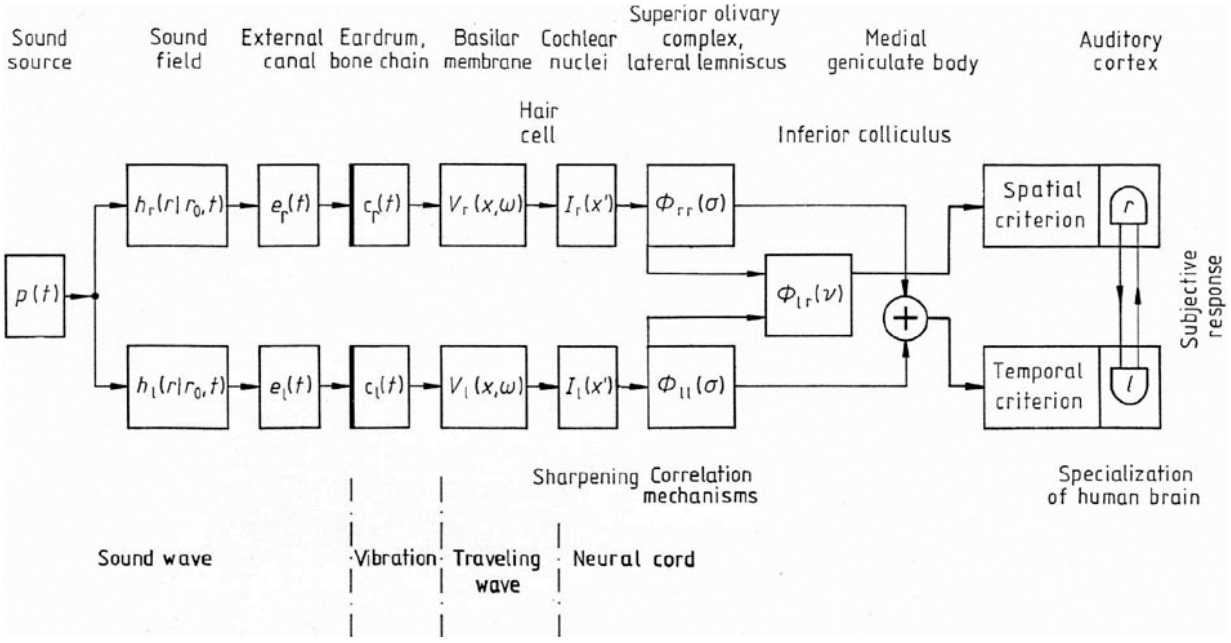

Fig. 1. Model of processors in the auditory-brain system for subjective responses. p(t): source sound signal in the time domain; $h_{l,r}(r/r_0,t)$: head-related impulse responses from a source position to the left and right ear-entrances; $e_{l,r}(t)$: impulse responses of left and right external canals, from the left and right ear-entrances to the left and right eardrums; $c_{l,r}(t)$: impulse responses for vibration of left and right bone chains, from the eardrums to oval windows, including transformation factors into vibration motion at the eardrums: $V_{l,r}(x,\omega)$: travelling wave forms on the basilar membranes, where x the position along the left and right basilar membrane measured from the oval window; $I_{l,r}(x')$: sharpening in the cochlear nuclei corresponding to roughly the power spectra of input sound, i.e., responses of a single pure tone ω tend to approach a limited region of nuclei. These neural activities may be enough to convert into activities similar to the ACF; $\Phi_{ll}(\sigma)$ and $\Phi_{rr}(\sigma)$: ACF mechanisms in the left and right auditory-pathways, respectively. Symbol ⊕ signifies that signals are combined; $\Phi_{lr}(\nu)$: IACF mechanism (Ando, 1985); r and l: specialization for temporal and spatial factors of the left and right human cerebral hemispheres, respectively. The left and right hemisphere may make temporal sensation and spatial sensations, respectively. Both hemispheres may make overall subject response such as subjective preference of the sound field (Ando, 2002).

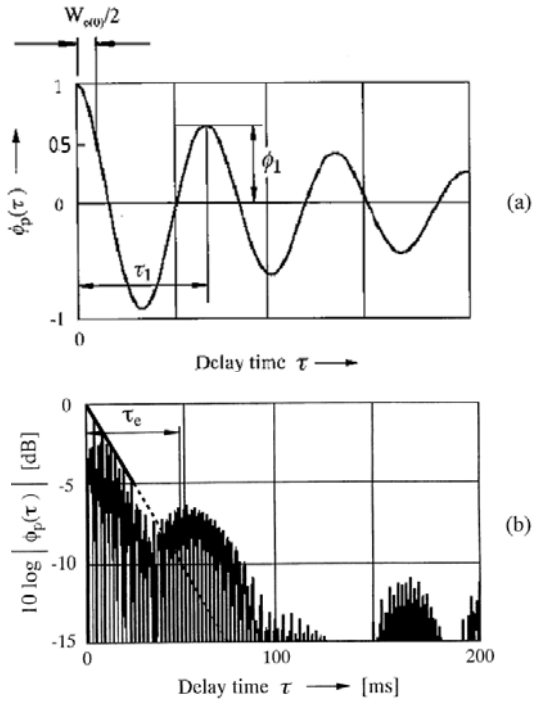

Fig. 2. Definition of two temporal factors, $W_{\phi(0)}$, τ_1 and ϕ_1, extracted from the ACF (a). Determination of the effective duration of running ACF, τ_e (b).

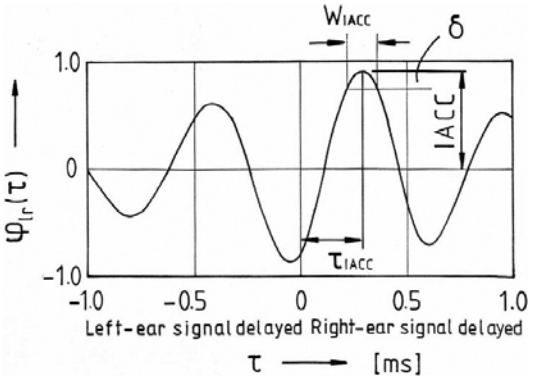

Fig. 3. Definition of the three spatial factors IACC, τ_{IACC} and W_{IACC}, extracted from the IACF.

for the possible maximum interaural time delay, –1 ms < τ < +1 ms.

2. The interaural delay time, at which IACC is defined, is IACC.
3. The width of the IACF defined by the interval of delay time at a value of below the IACC that may correspond to the just noticeable difference (JND) of the IACC is given by W_{IACC}. This value mainly depends on the frequency component of the source signals, and correlates to above-mentioned $W_{\phi(0)}$.
4. The denominator of Eq. (4) defined by the binaural LL is the geometrical mean of the sound energies arriving at the two ears, $\Phi_{ll}(0)$ and $\Phi_{rr}(0)$.

3. THEORY OF TEMPORAL AND SPATIAL PRIMARY SENSATIONS

According to the auditory model, we consider differences in response patterns between the cerebral hemispheres of human listeners. Here temporal factors are more prominent in the left hemisphere, while spatial factors are more prominent in the right hemisphere (Table 1). The model in which an internal variable associated with each hemisphere is modeled as the linear combination of hemisphere-specific factors can explain these differences. Temporal primary sensations S_L and spatial primary sensations S_R can be modeled in terms of the contributions of different factors that dominate in the neural response:

$$S_L = [f(c_{1l})+f(c_{2l})+...+f(c_{ml})]_{\text{left hemisphere}}$$
$$S_R = [f(c_{1r})+f(c_{2r})+...+f(c_{nr})]_{\text{right hemisphere}} \quad (6)$$

where c_{ml} and c_{nr} are the temporal factors extracted from the ACF and the spatial factors extracted from the IACF, respectively as defined by previous section.

Over all subjective responses including preference and annoyance, can be expressed by both temporal and spatial factors, such that

$$S = S_L + S_R \quad (7)$$

4. TEMPORAL AND SPATIAL PRIMARY SENSATIONS

4.1. Pitch of Complex Tones

What is interesting is that harmonic complexes having no energy at the fundamental frequency in their power spectra (i.e., they have only "upper" partials) still can produce strong "low" pitches at the fundamental itself. It is thus the cases for complex tones with a "missing fundamental" that strong pitches are heard that correspond to no individual frequency component, and this raises deep questions about whether patterns of pitch perception are consistent with frequency-domain representations.

A pitch-matching test compared pitches of pure and complex tones was performed to reconfirm previous results (Sumioka and Ando, 1996). The test signals were all complex tones consisting of harmonics 3-7 of a 200 Hz fundamental. All tone components had the same amplitudes, as shown in Fig. 4. As test signals, the two waveforms of complex tones, (a) in-phases and (b) random-phases, were applied as shown in Fig. 5. Starting phases of all components

of the in-phase stimuli were set at zero. The phases of the components of random-phase stimuli were randomly set to avoid any periodic peaks in the real waveforms. As shown in Fig. 6, the normalized ACF of these stimuli were calculated at the integrated interval 2T = 0.8 s. Though the waveforms differ greatly from each other, as shown in Fig. 6, their ACF are identical. The time delay at the first maximum peak of the ACF, τ_1 equals to 5 ms (200 Hz) corresponding to the fundamental frequency. The subjects were five musicians (two male and three female, 20-26 years of age). Test signals were produced from the loudspeaker in front of each subject in a semi-anechoic chamber. The sound pressure level (SPL) of each complex tone at the center position of the listener's head was fixed at 74 dB by analysis of the ACF $\Phi(0)$. The distance between a subject and the loudspeaker was 0.8 m±1 cm.

Probability of matching frequencies counted for each 1/12-octave band (chromatic scale) of the in-phase stimuli and random-phase stimuli are shown in Fig. 7. The dominant pitch of 200 Hz is included neither in the spectrum nor in the real waveform of random phases. But, it is obviously included in the period in the ACF. For both in-phase and random-phase conditions, about 60% of the responses clustered within a semitone of the fundamental. Results obtained for pitch under the two conditions are definitely similar. In fact, the pitch strength remains the same under both conditions. Thus, pitch of complex tones can be predicted from the time delay at the first maximum peak of the ACF, 1.This result agrees well with those obtained by Yost (1996) who demonstrated that pitch perception of iterated rippled noise is greatly affected by the first major ACF peak of the stimulus signal.

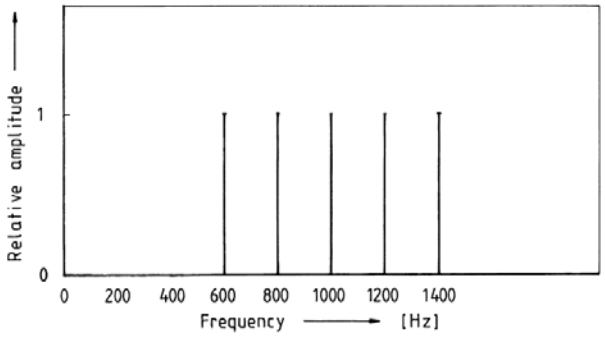

Fig. 4. Complex tones presented with pure-tone components of 600 Hz, 800 Hz, 1000 Hz, 1200 Hz, and 1400 Hz without the fundamental frequency of 200 Hz.

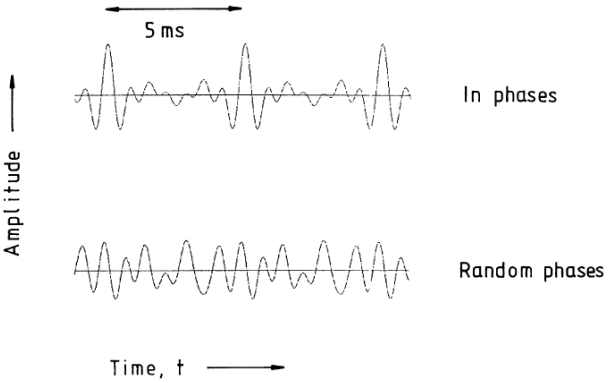

Fig. 5. Real waveforms of the complex tone in phase components (above) and in random phase components (below).

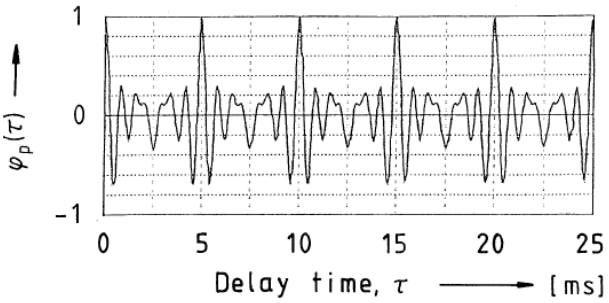

Fig. 6. Normalized autocorrelation function (NACF) of the two complex tones with its period of 5 ms (200 Hz).

4.2. Limitations of the ACF Model

For fundamental frequencies of 500, 1000, 1200, 1600, 2000, and 3000 Hz, stimuli consisting of two or three pure tone components were produced (Inoue, Ando and Taguti, 2001). 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 forth harmonics. The starting phase of all components was adjusted to zero (in phase). The total sound pressure level at the center of the listener's head was fixed at 74 dB. The ACF of all stimuli was calculated obtaining the peak τ_1 related to the fundamental frequency. The 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. Three subjects with musical experience (two male and one female, ages between 21 and 27 years old) participated. Pitch-matching tests were conducted using complex tones as test stimuli and a pure tone generated by a sinusoidal generator as a reference.

Results for all subjects are shown in Fig. 8. Whenever the fundamental frequency of the stimulus was 500, 1000, or 1200 Hz, more than 90% of the responses obtained from all subjects under both conditions clustered around the fundamental frequency. When the fundamental frequencies of the stimuli were 1600, 2000, or 3000 Hz, however, the probabil-

ity, that the subjects adjusted the frequency of the pure tone to the calculated fundamental frequency, is extremely decreased. These results imply that the ACF model is applicable when the fundamental frequency of stimuli is below 1200 Hz.

4.3. Loudness

The bandwidth (Δf) of a sharp filter was changed by using the cut-off slope of 2068 dB/octave, which was produced by repeating of passing through the cascade-two filters (Sato, Kitamura and Ando, 2002). Factors of τ_1, τ_e and ϕ_1 extracted from the ACF. It is worth noticing to obtain a possible longest e value that the filter bandwidth of "0 Hz" was used also, in which a sharp filtered noise consisting only its slope component. All source signals were the same sound pressure level at 74 dBA, which was accurately adjusted by measurement of the ACF at the origin of the delay time, $\Phi(0)$.

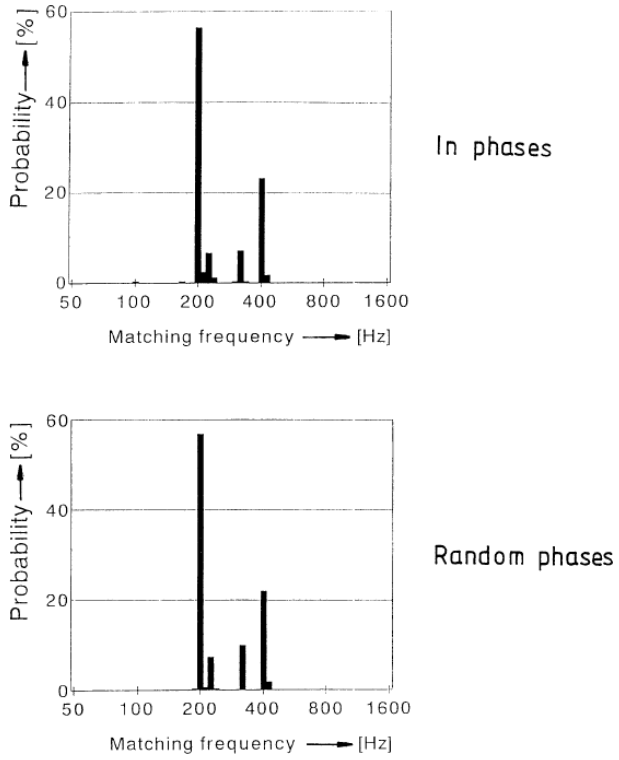

Fig. 7. Results of the pitch-matching tests for five subjects.

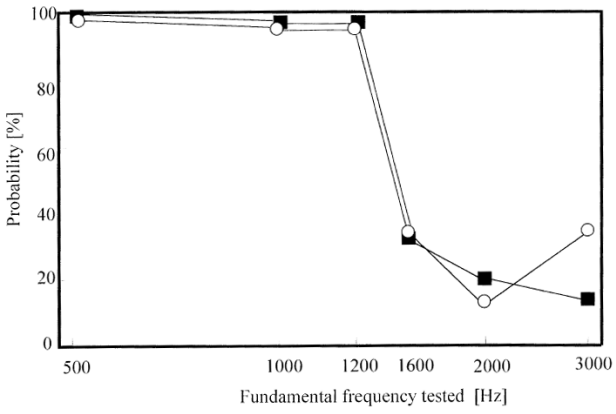

Fig. 8. Probability that three subjects adjusted a pure tone near to the fundamental frequency of the complex tone. Different symbols indicate results obtained under different conditions: empty circles are for two omponents, and full squares are for three components.

The loudness judgment was performed by the paired-comparison test (PCT) for which the ACF of the band-pass noise was changed. A headphone delivered the same sound signal to the two ears. Thus, the IACC was kept constant at nearly unity. Sound signals were digitized at a sampling frequency of 48 kHz. The paired comparison test (PCT) was conducted with five subjects with normal hearing, who were seated in an anechoic chamber and asked to judge which of two paired sound signals was perceived to be louder. The stimulus duration was 1.0 s, rise and fall times were 50 ms, and a silent interval between the stimuli was 0.5 s. The silent interval of 3.0 s separated each pair of stimuli, and the pairs were presented in random order.

Fifty responses (5 subjects 10 sessions) to each stimulus were obtained. Consistency tests indicated that all subjects had a significant ($p < 0.05$) ability to discriminate loudness. The test of agreement also indicated that there was significant agreement among all subjects ($p < 0.05$). A scale value of loudness was obtained by applying the law of comparative judgment (Thurstone's case V). The scale value was confirmed by goodness of fit throughout this paper.

The relationship between the scale value of loudness and the filter bandwidth is shown in Fig. 9. The scale value difference of 1.0 corresponds about 1 dB due to the preliminary experiment. For all three-center frequencies (250, 500, 1000 Hz) the scale value of loudness is maximal for the pure tone with the infinite value of τ_e and wide bandwidths, with minima at narrower bandwidths (40, 80, 160 Hz respectively). From the dependence of τ_e on filter bandwidth, we found that loudness increases with increasing τ_e within the "critical bandwidth". Results of analysis of variance for the scale values of loudness are indicated that for all center frequencies tested, the scale value of loudness of pure tone was significantly larger than that of other band pass noises within the critical band ($p < 0.01$). Consequently, loudness of the band-pass noise with identical sound pressure level was not constant within the critical band. Also, loudness of the pure tone was significantly larger than that of sharply filtered noises, and loudness increased with increasing τ_e within the critical band.

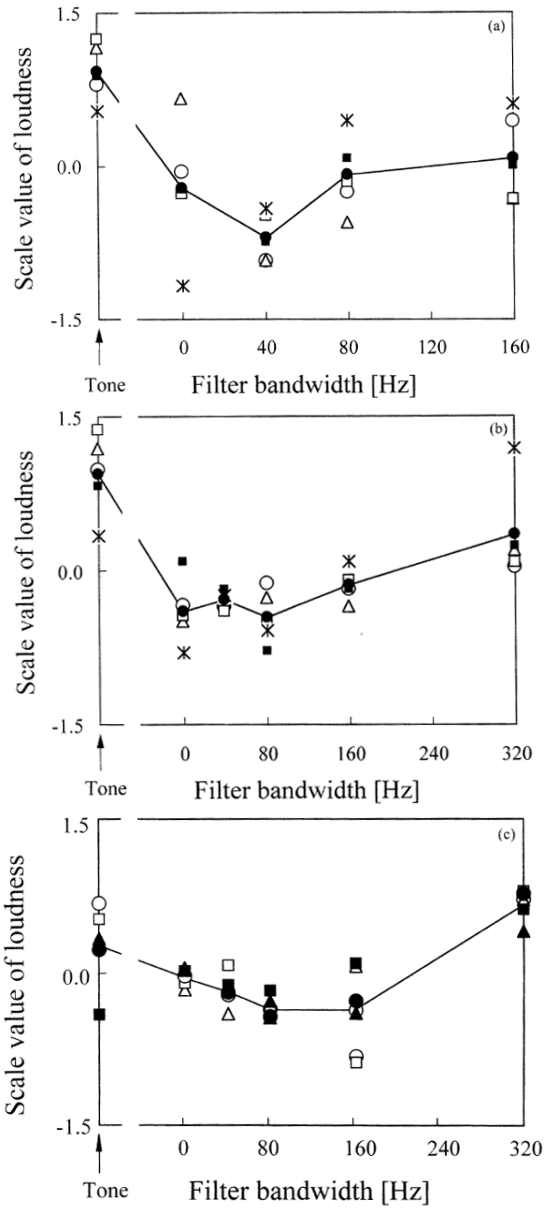

Fig. 9. Scale values of loudness as a function of the bandwidth. Different symbols indicate the scale values obtained with different subjects. (a) f_c = 250 Hz (a). f_c = 500 Hz (b). f_c = 1000 Hz (c).

4.4. Timbre

Timbre is defined as an aspect of sound quality independent of loudness, pitch and duration, i.e. that the quality of sound texture that distinguishes two notes of equal pitch, loudness and duration that are played by different musical instruments. An attempt is made here to investigate the relationship between the temporal factor extracted from the ACF of an electric guitar sound and dissimilarity representing the difference of timbre with a difference of distortion. As shown in Fig. 2(a), a factor $W_{\phi(0)}$ is defined by the delay time at the first 0.5 crossing of the normalized ACF, $\phi(t)$. It is worth noticing that this value is equivalent to the factor W_{IACC} extracted from the IACF.

An electric guitar with the "distortion" is a main instrument in pops and rock music. Previously, Marui and Martens (2005) investigated timbre variations by use of three types of nonlinear distortion processors with three differing level of Zwicker Sharpness (Zwicker and Fastl, 1999).
In this study, we examined whether or not timbre is described by the temporal factor extracted from the running ACF of the source signal that distinguishes notes of equal pitch, loudness and duration, which are played by different distortion levels (Hanada, Kawai and Ando 2007).

4.4.1. Experiment 1

The purpose of this experiment is to found the factor extracted from the running ACF contributing to the dissimilarity of simulated sounds changing the strength of distortion by the use of computer program. The distortion of music signal p(t) was processed, such that: when $|p(t)| \le C$

$$p(t) = p(t) \tag{8a}$$

and when $|p(t)| \ge C$

$$p(t) = +C,\ p(t) \ge C;\ p(t) = -C,\ p(t) \le -C \tag{8b}$$

where C is the cut-off pressure amplitude, and its level is defined by

$$CL = 20\log_{10}(C/|p(t)|_{max}) \tag{9}$$

and $|p(t)|_{max}$ is the maximum amplitude of the signal.

The value of CL was varied from 0 to −49 dB (7 dB step), so that eight stimuli were applied for test signals. As indicated in Table 2, pitch, listening level, and signal duration were fixed. Subjects participated were 19 students (male and female of 20 years of age). Subjects listened to three stimuli, and judged degree of dissimilarity. The number of combinations of this experiment was ${}_8C_3 = 56$ triads. The dissimilarity matrix was made according to the judgments giving the number that, 2 for the most different pair, 1 for the neutral pair, and 0 for the most similar pair. After the analysis of multi-dimensional scaling, we obtained the scale value (SV) of dissimilarity. It is worth noticing that this value is different from the scale value obtained by the law of comparative judgment by the PCT.

Table 2. Conditions of two dissimilarity experiments.

Condition	Experiment 1	Experiment 2
1) Conditions fixed		
Note (Pitch)	A4 (220 Hz) By use of 3rd string and 2nd fret	A4 (220 Hz) By use of 3rd string and 2nd fret
Listening level in L_{AE} [dB]	80	70
Signal duration [s]	4.0	1.5
2) Conditions varied		
CL [dB] by Eq. (9)	Eight signals tested changing the cut-off level for 0-49 dB (7 dB step)	
Distortion type	-	Three different types: VINT, CRUNCH and HARD
Drive level	-	Three levels due to the strength of distortion: 50, 70, 90 due to effectors Type ME-30 (BOSE)

We analyzed contributions of factors to the SV, for example, the mean value of $W_{\phi(0)}$, the decay rate of SPL [dBA/s] and the mean value of 1 (pitch strength). It was found that the most significant factor contributing to the SV was the mean value of $W_{\phi(0)}$. There were certain amount of correlations between the mean value of $W_{\phi(0)}$ and other factors were found,

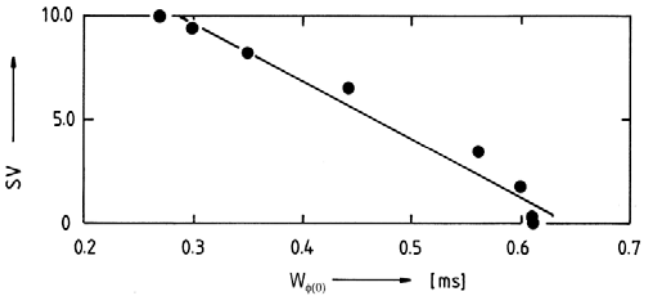

Fig. 10. Results of regression analysis for SV and the mean value of $W_{\phi(0)}$ (Experiment 1).

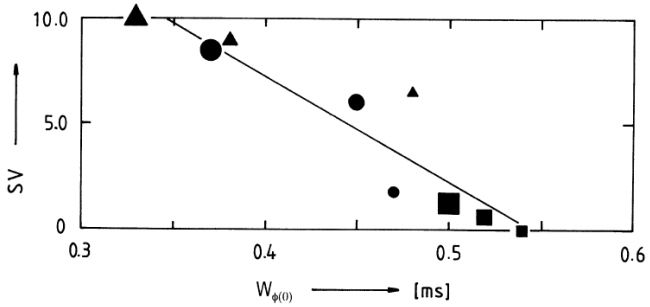

Fig. 11. Results of regression analysis for SV and the mean value of $W_{\phi(0)}$ (Experiment 2).

however, the mean value of $W_{\phi(0)}$ is considered as a representative. The SV as a function of the mean value of $W_{\phi(0)}$ is shown in Fig. 10. The correlation between the SV and the value of $W_{\phi(0)}$ is 0.98 ($p < 0.01$).

4.4.2. Experiment 2

The purpose of this experiment is to found the factor extracted from the running ACF contributing to the dissimilarity of sounds changing the strength of distortion by the use of available commercial effectors. As indicated in Table 2, nine stimuli were produced with three kind of effect types (VINT, CHUTCH and HARD) and three drive levels due to strength of distortion, 50, 70 and 90 by the effectors Type ME-30 (BOSE). Subjects participated were 20 students (male and female of 20 years of age). The method of experiment was same as above. Thus, the number of combinations of this experiment was ${}_9C_3 = 84$ triads. As shown in Fig. 10, results achieved were similar to Experiment 1. The correlation between the SV and the value of $W_{\phi(0)}$ is 0.92 ($p < 0.01$).

So far, it has been found by two experiments that the most effective factor extracted from the running ACF of the source signal is $W_{\phi(0)}$ determining timbre. This factor is deeply related to the frequency component of the source signal. It is worth noticing that the ACF of sinusoidal wave is the cosine function, so that if the frequency is low then $W_{\phi(0)}$ is large. The identical fundamental frequency (pitch) F_0 may be produced by, for example, complex tones with the frequency components of ($2F_0$ and $3F_0$) and the components of ($3F_0$ and $4F_0$) and so on. Even if the pitch is the same, therefore, timbre is different, due to difference of $W_{\phi(0)}$. Ohgushi (1980) was reported that the lowest and highest components of complex tones affected on timbre. In fact, it could be deeply related to $W_{\phi(0)}$.

4.5. Duration Sensation

As a fourth of the temporal sensation, we introduce the duration sensation. Experiments for pure and complex tones were performed by PCT (Saifuddin, Matsushima and Ando, 2002). Throughout this investigation, the listening level was fixed at 80 dBA. Perceived durations of two-component complex tones (3000 and 3500 Hz) having a fundamental at 500 Hz were compared with those by pure tone stimuli at 500 and 3000 Hz. Pairs consisting of two stimuli were presented randomly to obtain scale values for duration sensation (DS). Three signal durations, including rise/fall segments, were used for each of the stimuli: D = 140, 150, and 160 ms. Waveform amplitudes during stimulus onsets and offsets were

ramped with rise/fall times of 1 ms for all stimuli tested, the time required to reach a 3 dB below the steady level. There were thus 9 stimuli, and 36 pairs in total. The source stimuli were presented in a darkened soundproof chamber from a single loudspeaker at the horizontal distance of 74 cm from the center of the seated listener's head. Ten students participated in both experiments as subjects of normal hearing levels (22 and 36 years old). Each pair of stimuli was presented five times randomly within every session for each subject.

Observed scale values for the perceived durations of the 9 stimuli are shown in Fig. 12. While signal duration and stimulus periodicity had major effects on perceived duration, the number of frequency components (1 vs. 2) did not. Perceived durations of tones with the same periodicity (f = 500 Hz and F_0 = 500 Hz) were almost identical, while durations for pure tones of different frequencies (f = 500 Hz and f = 3000 Hz) differed significantly, by approximately 10 ms (judging from equivalent scale values, the 500 Hz pure tone appeared ~10 ms longer than the 3000 Hz tone). Thus, the duration (DS) of the higher frequency pure tone (3000 Hz: τ_1 = 0.33 ms) was found to be significantly shorter ($p < 0.01$) than that of either the pure tone (frequency: 500 Hz; τ_1 = 2 ms) or the complex tone (F_0 = 500 Hz: τ_1 = 2 ms). Also, the scale values of DS between the two pure tones: τ_1 = 2 (500 Hz) and 0.33 ms (3000 Hz) are almost parallel, so that the effects of periodicity (τ_1) and signal duration (D) on the apparent duration (DS) are independent and additive. Therefore, for these experimental conditions, we may obtain

$$S_L = f(\tau_1, D) = f(\tau_1) + f(D) \quad (10)$$

where τ_1 is extracted from the ACF of the source signal.

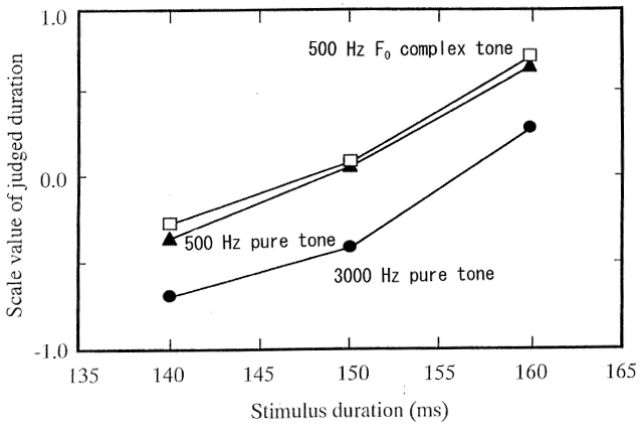

Fig. 12. Scale values of DS obtained by paired-comparison tests. □: complex tone, f_f = 500 Hz with 3000 Hz f_f = 500 Hz with 3000 Hz and 3500 Hz pure-tone components; ▲: 500-Hz pure tone; ●: 3000-Hz pure tone.

The four temporal primary sensations with related factors extracted from the ACF of the source signal are summarized in Table 3.

Table 3. Auditory temporal sensations in relation to the temporal factors extracted from the ACF.

Temporal Factors		**Temporal Sensations**			
		Loudness	Pitch	Timble[1]	Duration Sensation
ACF	$\Phi(0)$	**X**[2]	---[3]	---	---
	$W_{\phi(0)}$			**X**	
	τ_1	**X**	**X** ($F = 1/\tau_1$)	---	**X**
	ϕ_1	X[4]	**X** (Strength of pitch)	---	x
	τ_e	**X**	x	---	**X**
	D	---	---	---	**X**

[1]Timbre for the sound field has been discussed in relation to the temporal and spatial factors (Hotehama, Sato and Ando, 2002).
[2]**X**: Major factors describing each sensation.
[3]—-: Factors to be examined whether or not it influences on the respective sensation.
[4]x: An attention should be paid due to factors of ϕ_1 and τ_e are mutually correlated.

5. SPATIAL SENSATIONS OF THE BINAURAL SIGNALS

5.1. Localization in the Horizontal Plane

The perceived direction of a sound source in the horizontal plane is assumed to be expressed in terms of the spatial factors extracted from the IACF associated with the right hemisphere, such that

$$L_{holizontal} = f[\Phi_{ll}(0), \Phi_{rr}(0), IACC, \tau_{IACC}, W_{IACC}]_{right\ hemisphere} \quad (11)$$

where $\Phi_{ll}(0)$ and $\Phi_{rr}(0)$ signify sound energies of the signals arriving at the left and right ear-entrances. It is well known that the most significant factor for the horizontal localization in the five spatial factors in Eq. (4) is the inter-aural delay time, IACC, as well as sound energies at the two ears including the inter-aural level difference. A well-defined direction is perceived when the normalized IACF has one sharp maximum with a large value of IACC, and with a narrow value of W_{IACC} due to the high frequency components above 2 kHz.

5.2. Apparent Source Width (ASW)

The scale value of ASW was obtained by the PCT with ten subjects (Sato and Ando, 1996). In order to control the value of WIACC, the center frequency of 1/3-octave band pass noises was changed as 250 Hz, 500 Hz, 1 kHz, and 2 kHz. The values of IACC were adjusted by controlling the sound pressure ratio between reflections ($\xi = \pm 54°$) and the direct sound ($\xi = 0°$). In order to avoid effects of the listening level on ASW (Keet, 1968), the total sound pressure level at the ear canal entrances of all sound fields was kept constant at a peak of 75 dBA. Listeners judged which of two sound sources they perceived to be wider.

Results of the analysis of variance for the scale values S(ASW) indicate that both of factors IACC and W_{IACC} are significant ($p < 0.01$) and contribute to S(ASW) independently, so that

$$S_R = f(IACC) + f(W_{IACC}) \approx \alpha(IACC)^{3/2} + \beta(W_{IACC})^{1/2} \quad (12)$$

where coefficients $\alpha \approx -1.64$ and $\beta \approx 2.44$ are obtained by regressions of the scale values with ten subjects as shown in Fig. 13. This holds under the conditions of a constant LL and $\tau_{IACC} = 0$. Obviously, as shown in Fig. 14, calculated scale values by Eq. (5) and measured scale values are in good agreement ($r = 0.97$, $p < 0.01$). It is worth noticing that the wider ASW is perceived for the sound source with a predominately low frequency is reflected in the W_{IACC}.

Nest, we shall show that ASW depends on the W_{IACC} as a representative of the spectral component of the source signal, IACC and the LL as well (Sato and Ando, 2002). This study examines the ASW of the complex noise, which consists of the band-pass noises whose center frequencies are the harmonics of the fundamental frequency. An additional question is that whether or not pitch (the fundamental frequency) contributes to the ASW. The complex noises with the fundamental frequencies centered on 200, 400, and 800 Hz in the applicable range of ACF (Section 4.2) were used as the source signal. The complex noise consisted of three band-pass noise components, and the center frequencies of the lowest band noise components were fixed at 1600 Hz. When the fundamental frequency $F_0 = 200$ Hz, the three band pass noise components were 1600, 1800, and 2000 Hz. Similarly, $F_0 = 400$ Hz (1600, 2000, 2400 Hz) and $F_0 = 800$ (1600, 2400, 3200 Hz). The amplitudes of all band noise components were adjusted exactly to the same by measuring $\Phi(0)$. The component band-noise width of the complex noise was 80 Hz with a cut-off slope of 2068 dB/octave.

By use of a single frontal loudspeaker for direct sound ($\xi = 0°$) and two symmetrical loudspeakers ($\xi = \pm 54°$) for reflections added were simulated in an anechoic chamber. To reconfirm the effects of LL on ASW, the SPL at the listener's head position was changed from 70 to 75 dB. The values of the IACC of all sound fields were adjusted to 0.90 ± 0.01 by controlling the sound pressure ratio of the reflections relative to the level of the direct sound. The PCT of twelve sound fields (6×2) was performed on five subjects with normal hearing ability. Twenty-five responses (five subjects $\times$ five repeats) to each stimulus were obtained. The scale values of ASW obtained are shown in Fig. 15 as a function of W_{IACC} of the source signal and a parameter of LL. The results of the analysis of variance for scale values of ASW revealed that the explanatory factors W_{IACC} and LL are significant ($p < 0.01$). The interaction between W_{IACC} and LL is insignificant, so that

$$S_R = f(W_{IACC}) + f(LL) \approx a(W_{IACC})^{1/2} + b(LL)^{3/2} \quad (13)$$

where a and b are coefficients. The curves in Fig. 15 confirm the calculated scale values of Eq. (13) with coefficients $a \approx 2.40$ and $b \approx 0.005$. These coefficients were obtained by

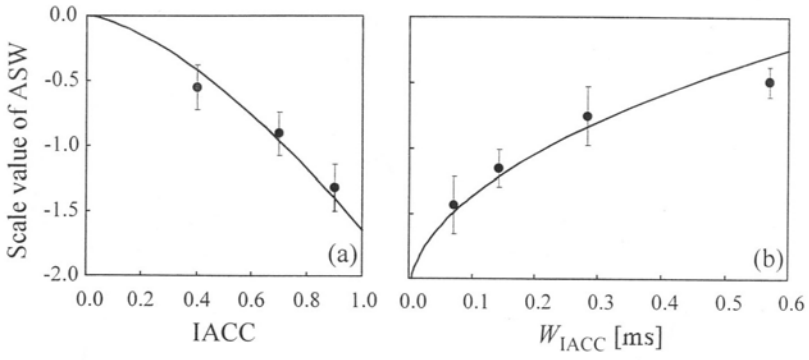

Fig. 13. Average scale values of ASW for the 1/3 octave-band pass noises with 95% reliability as a function of (a) the IACC and (b) the W_{IACC}. The regression curves are calculated with $\alpha = 1.64$ and $\beta = 2.44$.

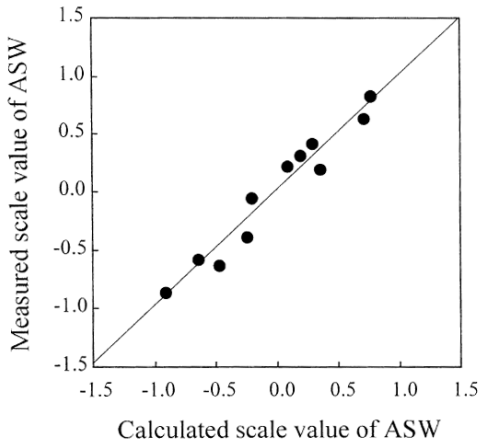

Fig. 14. Relationship between the measured scale values of ASW and scale values of ASW calculated by Eq. (12) with $\alpha = -1.64$ and $\beta = 2.44$. Correlation coefficient r = 0.97 ($p < 0.01$).

multiple regression analyses. It is noteworthy that the scale value of ASW for the1/3-octave band pass noise is also expressed in terms of the 1/2 power of W_{IACC} as expressed by Eq. (12) and that the coefficient for W_{IACC} ($\beta \approx 2.44$) is close to that of this study.

Here, the factor W_{IACC} may be determined by the frequency component of the source signal only, thus the pitch or the fundamental frequency itself did not affect on ASW. Results of ANOVA for scale values of the ASW indicated that the explanatory factor LL is also significant ($p < 0.01$). The scale values of the ASW increase with an increase in the LL as similar to those by Keet (1968). Fig. 16 shows the relationship between the measured scale value of the ASW and the scale value of the ASW calculated by Eq. (13) with the coefficients a = 2.40, and b = 0.003. The correlation coefficient between the measured and calculated scale values is 0.97 ($p < 0.01$).

The facts that the weighting coefficients of $(W_{IACC})^{1/2}$ are most similar according to Eqs. (12) and (13) might read to a common formula, such that

$$S_R = f(IACC)+f(W_{IACC})+f(LL) \approx \alpha(IACC)^{3/2}+\beta(W_{IACC})^{1/2}+ (LL)^{3/2} \quad (14)$$

where $\alpha \approx -1.64$, $\beta \approx 2.42$, $\gamma \approx 0.005$. It is worth noticing that the unit of scale values of subjective preference was almost constant even applying different sound sources (Ando, 1983).

5.3. Subjective Diffuseness

In order to obtain the scale value of subjective diffuseness, the PCT with band-pass Gaussian noise, varying the horizontal angle of two symmetric reflections, has been conducted (Ando and Kurihara, 1986; Singh, Ando and Kurihara, 1994). Listeners judged which of two sound fields were perceived as more diffuse. Subjective diffuseness are inversely proportional to the IACC, and may be formulated in terms of the 3/2 power of the IACC in a manner similar to the subjective preference values, i.e.,

$$S_R = -\alpha(IACC)^{\beta} \quad (15)$$

where $\alpha = 2.9$, $\beta = 3/2$.

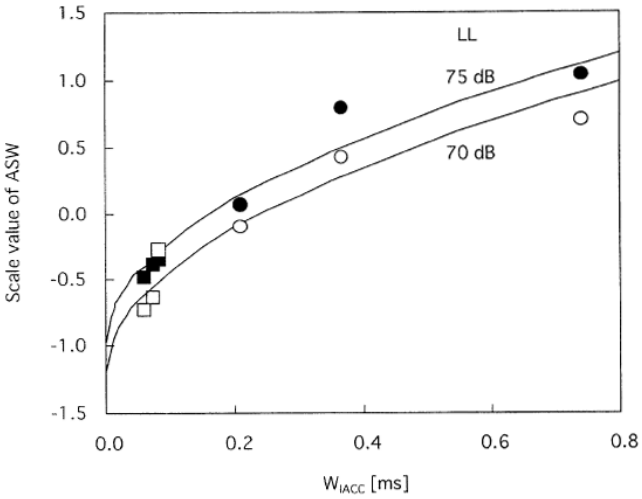

Fig. 15. Average scale values of ASW as a function of W_{IACC} and as a parameter of LL. ●: band pass noise; LL = 75 dB; ○: band pass noise; LL = 70 dB; ■: complex noise; LL = 75 dB; □: complex noise; LL = 70 dB. The regression curve is expressed by Eq. (13) with a = 2.40 and b = 0.005.

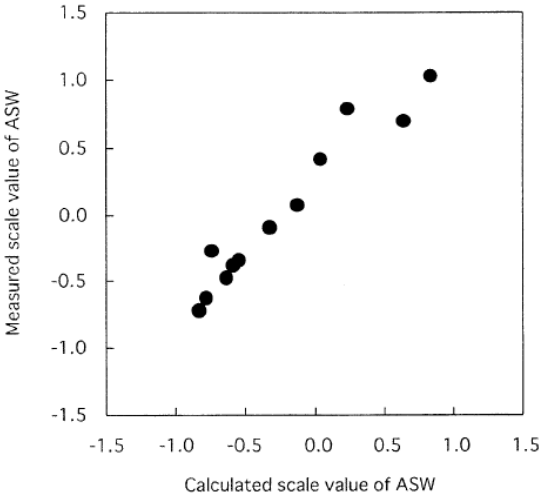

Fig. 16. Relationship between the measured scale values of ASW and the scale values of ASW calculated by Eq. (13) with a = 2.40 and b = 0.005. Correlation coefficient r = 0.97 ($p < 0.01$).

Relationship between the scale value of subjective diffuseness obtained by PCT and calculated value by Eq. (15), as a function of the IACC, are shown in Fig. 17. There is a great variation of data in the range of IACC < 0.5, however, no essential difference may be found in the results with frequencies between 250 Hz - 4 kHz. The scale values of subjective diffuseness for 1/3 octave-band pass noise with the center frequencies of 250 Hz, 500 Hz, 1 kHz, 2 kHz and 4 kHz, which depend on the horizontal angle, are shown in Fig. 18, Clearly, the most effective horizontal angles of reflections depend on the frequency range. These are about ±90° for the 500 Hz range and the frequency range below 500 Hz, around ±55° for the 1 kHz range (that is the most important angle for the music), and smaller than 20° for the 2 kHz and 4 kHz ranges. It is considered, when the listening level is increased, the scale value of subjective diffuseness of the sound field is increased also.

The three spatial primary sensations with related factors extracted from the IACF of the sound field are summarized in Table 4.

6. REMARKS AND FUTURE WORKS

The auditory-brain model has been well grounded by physiological evidences. Accordingly, we may well describe temporal sensations and spatial sensations in terms of the temporal factor extracted from the ACF and the spatial factor extracted from the IACF, respectively, as shown in Table 3

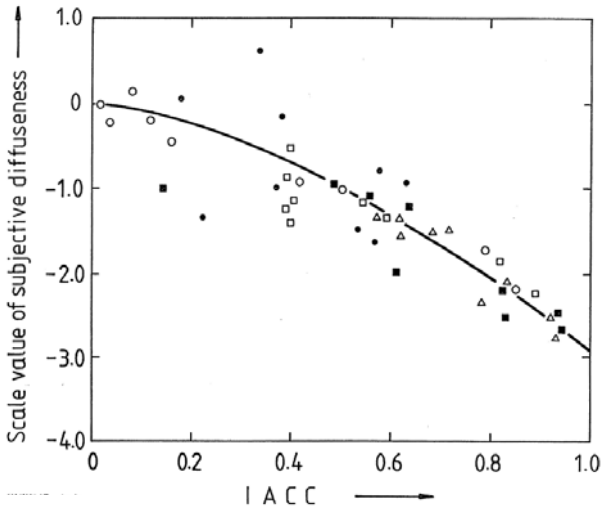

Fig. 17. Scale values of subjective diffuseness as a function of the IACC (calculated). Different symbols indicate different frequencies of the 1/3 octave band pass noise: △: 250 Hz, ○: 500 Hz, □: l kHz, ●: 2 kHz, ■: 4 kHz. ____: Regression line.

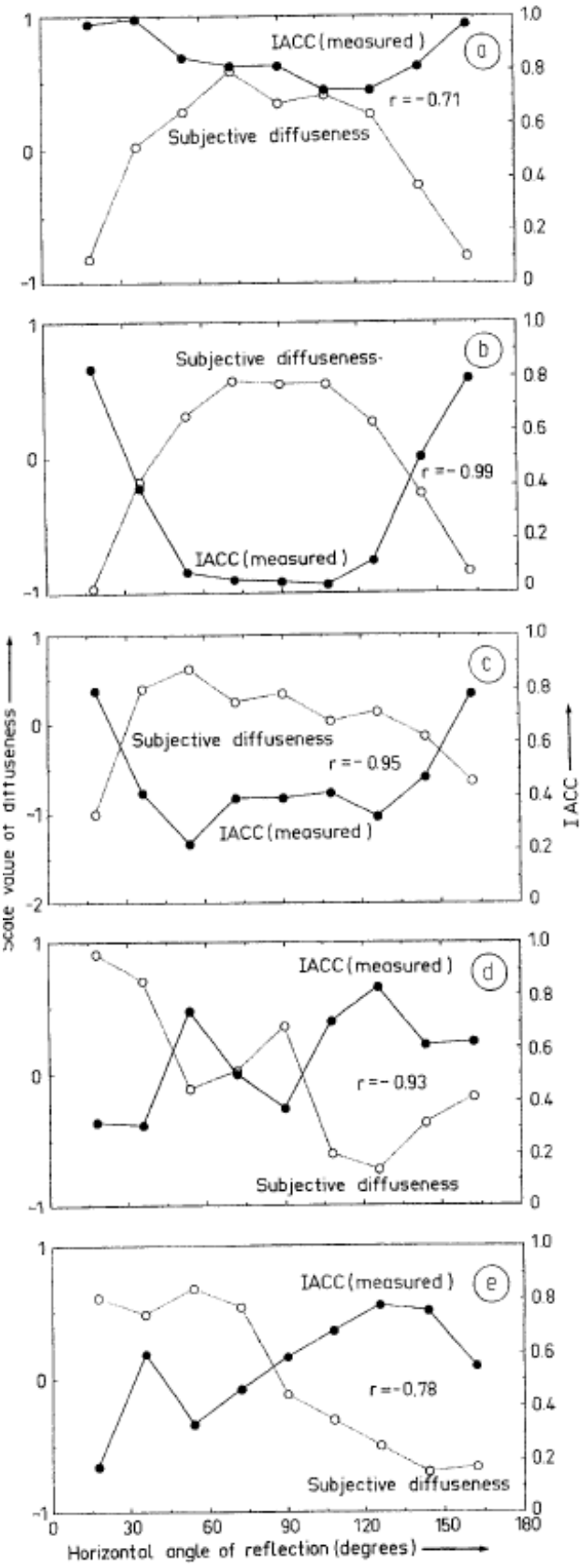

Fig. 18. Scale values of subjective diffuseness and the IACC as a function of the horizontal angle of incidence to a listener, with 1/3 octave band noise of center frequencies. (a) 250 Hz. (b) 500 Hz. (c) 1 kHz. (d) 2 kHz. (e) 4 kHz.

Table 4. Auditory temporal sensations in relation to the temporal factors extracted from the ACF.

Spatial Factors		Spatial Sensations		
		Localization in the Horizontal Plane	ASW	Subjective Diffuseness
IACF	LL[1]	---	---[2]	---
	τ_{IACC}	**X**[3]	---	---
	W_{IACC}	**X**	$(W_{IACC})^{1/2}$	---
	IACC	**X**	$(IACC)^{3/2}$	$(IACC)^{3/2}$

[1] LL = 10 log $[\Phi(0)/\Phi(0)_{reference}]$, where $\Phi(0) = [\Phi_{ll}(0)\ \Phi_{rr}(0)]^{1/2}$.

[2] —-: Factors to be examined whether or not they influence on the respective sensation.

[3] **X**: Major factors describing localization in the horizontal plane. Note that for localization in the vertical plane, factors are extracted from the ACF (Sato, Ando and Mellert, 2001).

and Table 4. Thus, any subjective attributes including, for example, speech identification in the sound field could be described in term of the ACF and IACF processes in the auditory pathway and the specialization of two cerebral hemispheres (Ando, Sato and Sakai, 1999; Ando, 2002). Also, in accordance with the model, a method of noise measurement and identification were proposed (Ando, 2001; Ando and Pompoli, 2002), and substantial measurements were conducted (Fujii, Soeta and Ando, 2001; Kitamura, Shimokura, Sato and Ando, 2002). For example, annoyances of noise in relation to the temporal factor (Soeta, Maruo and Ando, 2004), and the spatial factor (Sato, Kitamura and Ando, 2004) have been well described.

Since the purpose of this paper is to propose the comprehensive theory of the temporal and spatial sensations, some of research works are left, which should be investigated in the future. Further research works as indicated by the symbol (---) in Table 3 and Table 4 should be conducted.

As an application for speech recognition, for instance, this theory might be a successful approach because of a limited number of temporal and spatial factors, which may be attained by simple algorithms. It is worth noticing that many auditory theorists have postulated that hearing is based on dual frequency- and time-domain auditory representations. Maps based on cochlear "place" have been thought to cover the frequency range of pure tone hearing and cochlear resonances, while the temporal representation has been thought to cover the range of periodicities available in neuronal firing patterns (roughly up to 4-5 kHz). Even when after getting such dual frequency- and time-domain representations, no cues describing the primary sensations may be extracted. Moreover, the auditory system might not be operative for further complicated mathematical algorithms.

This theory has a potential as a "seamless theory" for other modalities as well. For example, the visual temporal and spatial sensations based on the temporal and spatial factors, which may be extracted from the ACF of temporal and spatial signals. Also, it has been found that the temporal factors of visual field related to subjective preference are associated with the left hemisphere (Ando, 2007b). As is well known that the right hemisphere is associated with spatial abilities and nonverbal identification including face and pattern identification (Sperry, 1974).

APPENDIX (Ando, 1998)
AUDITORY BRAINSTEM RESPONSES (ABR) CORRESPONDING TO IACC

The spatial factors extracted from the IACF may provide a representation for judging subjective preference and spatial sensations. The left and right auditory-brainstem responses (ABR) of human subjects were recorded here, and analyzed in order to identify such a mechanism.

A.1. The Neuronal Response Correlates of Sound Direction in the Horizontal Plane

To probe the neural correlates of horizontal sound direction (azimuth), the source signals p(t) of trains of clicks (50 μs pulses), presented every 100 ms for 200 s (2,000 times). Signals were supplied to loudspeakers positioned at various horizontal angles (0-180°) with respect to the front of the subject, all on the subject's right hand side. The distance between each loudspeaker and the center of the head was kept at 68±1 cm. The speakers had a frequency response of ±3 dB for 100 Hz to 10 kHz.

The left and right ABRs were recorded through electrodes placed on the vertex, and the left and right mastoids (Ando, Yamamoto, Nagamatsu and Kang, 1991). Typical examples of recorded ABR waveforms as a function of the horizontal angle of sound incidence are shown in Fig. A1. It can be readily appreciated that waves I-VI differ in peak amplitude and latency as the sound location changes its angle of incidence relative to the subject's head. Similar ABR waveforms were obtained from each of four participating subjects (males, 23±2 years of age). Their ABRs were averaged together and the mean amplitude of the ABR waveform peaks (waves I-VI) was computed as a function of the horizontal angle (Fig. A2a-A2f). Of particular interest is that the average peak

I amplitudes from the right electrode are greater than those from the left, r > l for angles ξ = 30°-120° ($p < 0.01$), which may reflect interaural differences in sound pressure (head shadowing) produced by the source location on the right hand side. This tendency is reversed for wave II for two angles ξ = 60°-90° (l > r, $p < 0.05$, Fig. A2b). The behavior of wave III (Fig. A2c) is similar to that of wave I (r > l, $p < 0.01$). This tendency again reverses for wave IV (Fig. A2d, l > r, $p < 0.05$), and is maintained further in wave VI (Fig. A2f, l > r , $p < 0.05$) even though absolute values are amplified.

From these patterns, it could be inferred that the flow of the left and right neural signals is interchanged three times at the cochlear nucleus, the superior olivary complex and the lateral lemniscus as shown in Fig. A3. The interaction at the inferior colliculus in particular may be operative for the interaural signal processing as discussed below. In wave V as shown in Fig. A2e, such a reversal cannot be seen, and the relative behavior of amplitudes of the left and the right are parallel and similar. Thus, these two amplitudes were averaged and plotted in Fig. A6 (V symbols). For comparison, the amplitudes of wave IV (left-l and right-r) normalized to their respective ABR amplitudes at the frontal sound incidence, which may correspond to the normalized sound pres-

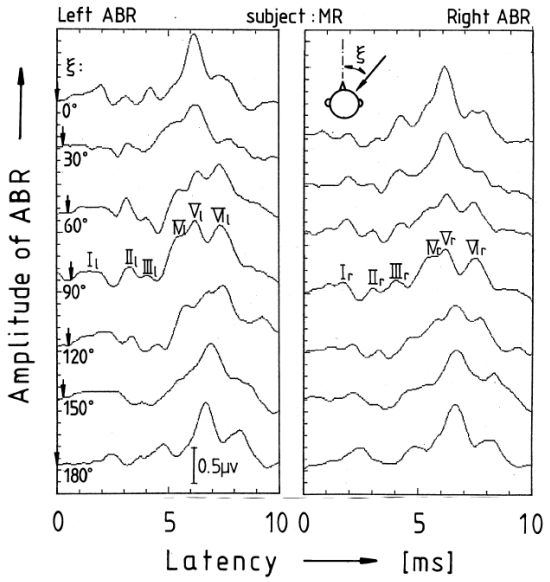

Fig. A1. Examples of auditory brainstem response (ABR) obtained between the vertex and left- and right-mastoids, as a parameter of the horizontal angle of sound incidence. The abscissa is the time relative to the time when the single pulse arrives at the right ear entrance. Arrows indicate the time delay, depending upon the sound source location of the right hand side of the subject, and the null amplitude of ABR.

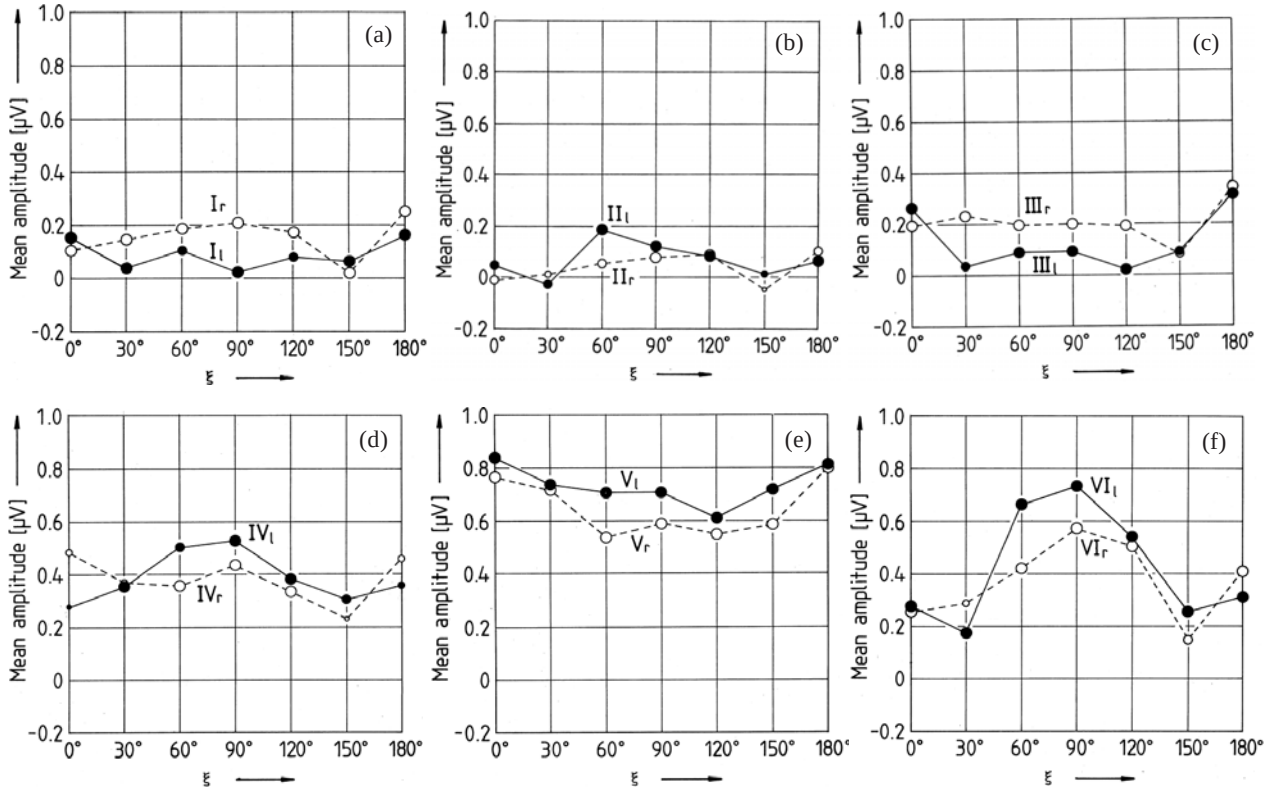

Fig. A2. Averaged amplitudes of ABR for each waves I - VI. The size of circles indicated the number of available data from four subjects. Filled circles: Left ABRs; Empty circles: Right ABRs. a) Wave I. b) Wave II. c) Wave III. d) Wave IV. e) Wave V. f) Wave VI.

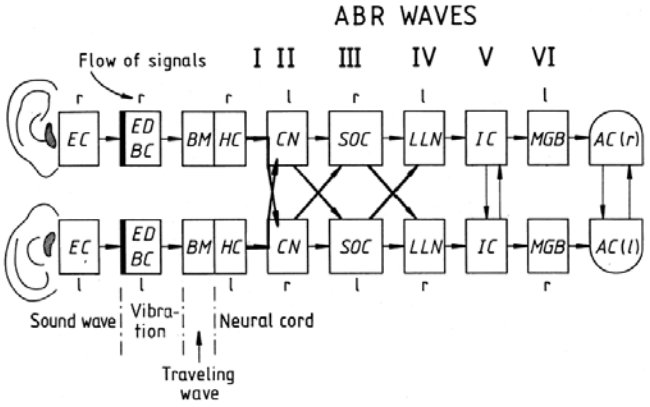

Fig. A3. Diagram for flow of neural signals in the auditory pathways. EC: External canal; ED and BC: Eardrum and bone chain; BM and HC: Basilar membrane and hair cell; CN: Cochlear nucleus; SOC: Superior olivary complex; LLN: Lateral lemniscus nucleus; IC: Inferior colliculus; MGB: Medial denticulate body; AC: Auditory cortex of the right and left hemispheres. The source location of each wave (ABR) was previously investigated for both animal and human subjects (Jewett, 1970; Lev and Sohmer, 1972; Buchwald and Huang, 1975).

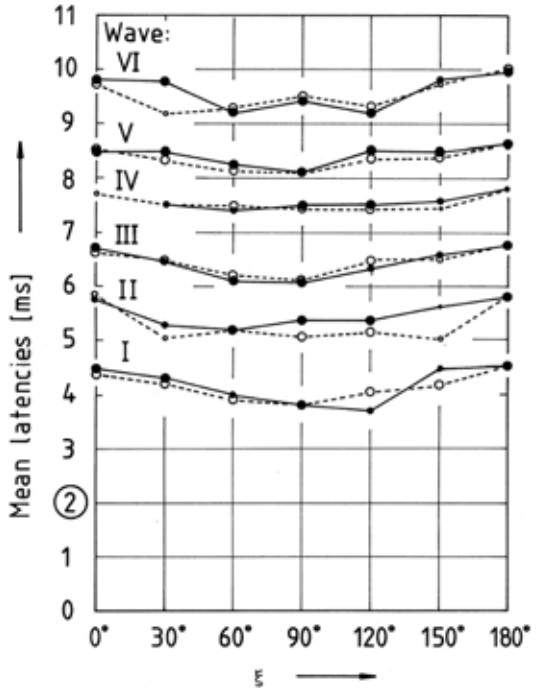

Fig. A4. Averaged latencies of ABRs for four subjects, waves I-VI. Sizes of the circle indicate the number of available data from four subjects. The latency of 2 ms indicated by ② corresponds to the distance between loudspeaker and the center of the head. Filled circles: Left ABR; Empty circles: Right ABR.

sures at the right and left ear entrances, respectively, are also plotted.

The relative latencies of the ABR peaks also change with the horizontal angle. Latencies of peaks of waves I through VI are computed relative to the time when the short click pulses were supplied to the loudspeaker. The shortest latencies were seen for angles around $\xi = 90°$ (Fig. A4). Significant differences exist ($p < 0.01$) between averaged latencies at lateral locations ($\xi = 90°$) and those in the median plane (in front $\xi = 0°$ or in back, at $\xi = 180°$). The differences are approximately 640 μs on average, which corresponds to the interaural time difference created by a sound incident at $\xi = 90°$. It is likely that the relative latency at waves I-III may be reflected by the interaural time difference, particularly at wave III. No significant differences could be seen between the latencies of the left and right of waves I-IV responding to such a relative factor.

A.2. ABR Amplitude Corresponding to the Sound Pressures and IACC

Neural ABR responses can be compared with cross-correlations derived from acoustical measurements at the two ears of a dummy head. A-weighted signals were presented and free-field sound pressure measurements were taken at the two ear entrances of a dummy head as a function of the horizontal angle of the sound source. Fig. A5 depicts the signal power at the two ears (R, L) for different angles (the zero-lag term of the ACF) and the maximum value of the IACF (ϕ),

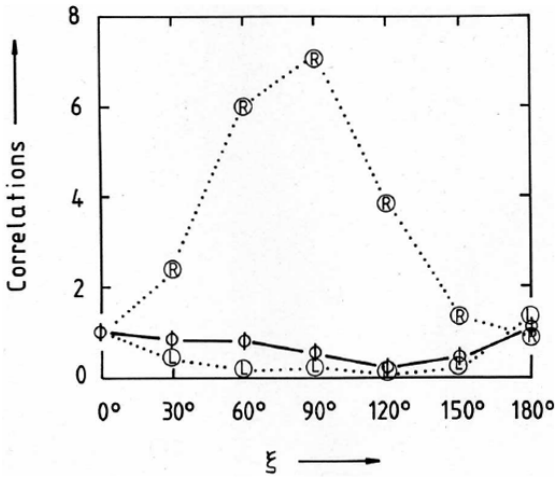

Fig. A5. Correlations of sound signals arriving at the left- and right-ear entrances of a dummy head, which are normalized by the respective values at $\xi = 0°$. (L): $\Phi_{ll}(0)$ measured at the left-ear; (R): $\Phi_{rr}(0)$ measured at the right-ear; Φ: Maximum interaural crosscorrelation, $|\Phi_{lr}(\tau)|_{max}$, $|\tau| < 1$ ms.

which are normalized only by the respective values at $\xi = 0°$. Received signal power is greatest for the ear ipsilateral to the speaker R when it is situated 90 degrees and least for the contralateral ear. These acoustic measures can be compared with the neurally generated ABR potentials (Fig. A6). Here the neural correlate of the relative power of the received signals at the left and right ears is the average of the peak amplitudes of waves IV and V (left and right), normalized to those at the frontal incidence. Similar results are obtained when amplitudes are normalized to those at . Although differences in units and scaling confound direct comparison between the results in Fig. A5 and A6, there are nevertheless qualitative similarities between these acoustic and physiological responses. The relative behavior of wave IV (l) in Fig. A6 is similar to $\Phi_{rr}(0)$ in Fig. A5, which was measured at the right-ear entrance r. Also, the relative behavior of wave IVr is similar to $\Phi_{ll}(0)$ at the left-ear entrance l. In fact, amplitudes of wave IV (left and right) are proportional to $\Phi_{rr}(0)$ and $\Phi_{ll}(0)$, respectively, due to the interchange of signal flow. The behavior of wave V is similar to that of the maximum value, $|\Phi_{lr}(\tau)|_{max}$, $|\tau| < 1$ ms. Since correlations have the dimensions of the power of the sound signals, i.e., the square of ABR amplitude , the IACC defined by Fig. 3 may correspond to,

$$P = \frac{A_V{}^2}{[A_{IV,r}\ A_{IV,l}]} \qquad \text{(A1)}$$

where A_V is the amplitude of the wave V, which may be reflected by the "maximum" synchronized neural activity ($\approx |\ \Phi_{lr}(\tau)|_{max}$) in the inputs to the inferior colliculus (see Fig. A3). And, $A_{IV,r}$ and $A_{IV,l}$ are amplitudes of wave IV from the right and left, respectively. The results obtained by Eq. (A1) are plotted in Fig. A7. It is clear that the behavior of the IACC and P are in good agreement ($r = 0.92$, $p < 0.01$).

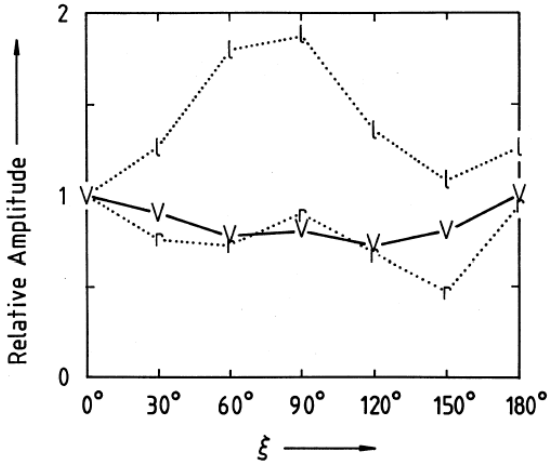

Fig. A6. Averaged amplitudes of ABR waves IV_l (symbol: l) and IV_r (symbol: r), and averaged amplitudes of waves V_l and V_r (symbol: V) normalized to the amplitudes at the frontal incidence (4 subjects).

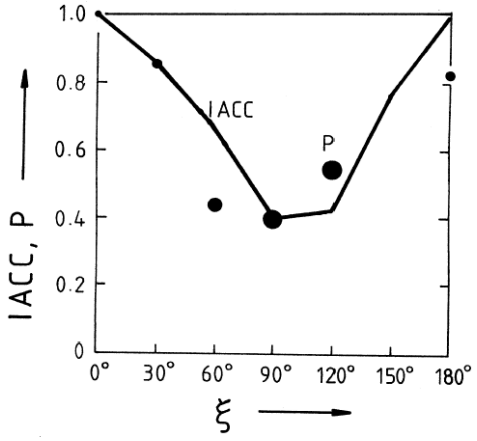

Fig. A7. Values of the IACC measured and corresponding values of P obtained by Eq. (A1). A linear relationship between the IACC and the value P is obtained ($p < 0.01$). Data at $\xi = 150^O$ were not available, except for one indicating P value far over unity, which is excluded.

A.3. Remarks

The amplitudes of the ABR clearly differ according to the horizontal angle of the incidence of sound relative to the listener (Fig. A2). In particular, it is found that the amplitudes of waves IV_l and IV_r are nearly proportional to the sound pressures at the right and left ear entrances, respectively, when the amplitude is normalized to that in front or back ($\xi = 180°$).

As far as the left- and right- amplitude behaviors of the ABR recorded here are concerned, the first interchange of the neural signal is considered to occur at the entrances of the cochlear nucleus; the second interchange may take place at the superior olivary complex and the third may be at the lateral lemniscus nucleus as shown in Fig. A3. Thompson and Thompson (1988), who used neuroanatomical tract-tracing methods in guinea pigs, found four separate pathways connecting one cochlea either with the other cochlea, or with itself, all via brainstem neurons. This relates to the first interchange at the entrance of the cochlear nucleus, in three interchanges in the auditory pathway. As has been discussed, the "maximum" of neural activity for wave V (inferior colliculus) in the auditory pathways corresponds to the IACC, which appeared in the amplitude of the ABR, around 8.5 ms after the sound signal supplied to the loudspeakers (68±1 cm). Also, the relative latency at wave III corresponds clearly to

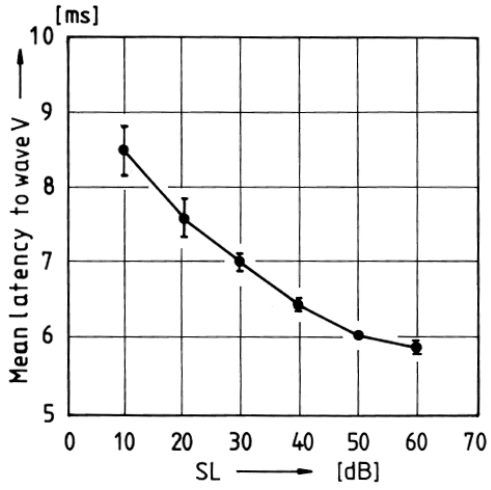

Fig. A8. Latency to wave V as a function of the sensation level (SL) (Hecox and Galambos, 1974). This may correspond to a binaural summation of sound energies from the left and right ears.

the interaural time difference (Fig. A4).

At inferior colliculus, on the other hand, Hecox and Galambos (1974) found that the latency of wave V decreases with an increasing sensation level as shown in Fig. A8. This implies binaural summation for the sound energy or the sound pressure level (SPL), which may be reflected in both $\Phi_{11}(0)$ and $\Phi_{rr}(0)$ of denominator of Eq. (4).

REFERENCES

Ando, Y. (1983). Calculation of subjective preference at each seat in a concert hall. Journal of the Acoustical Society of America, 74, 873-887.

Ando, Y. (1985). Concert hall acoustics. Springer-Verlag, Heidelberg.

Ando, Y. and Kurihara, Y. (1986). Nonlinear response in evaluating the subjective diffuseness of sound field. Journal of the Acoustical Society of America, 80, 833-836.

Ando, Y., Yamamoto, K., Nagamastu, H., and Kang, S.H. (1991). Auditory brainstem response (ABR) in relation to the horizontal angle of sound incidence. Acoustic Letters, 15, 57-64.

Ando, Y. (1992). Evoked potentials relating to the subjective preference of sound fields. Acustica, 76, 292-296.

Ando, Y. (1998). Architectural Acoustics, Blending Sound Sources, Sound Fields, and Listeners. AIP Press/Springer-Verlag, New York.

Ando, Y., Sato, S., Sakai, H. (1999). Fundamental subjective attributes of sound fields based on the model of auditory-brain system. Computational Acoustics in Architecture, Sendra, J.J. ed., WIT Press, Southampton.

Ando, Y. (2001). A theory of primary sensations measuring environmental noise. Journal of Sound and Vibration, 241, 3-18.

Ando, Y. (2002). Correlation factors describing primary and spatial sensations of sound fields. Journal of Sound and Vibration, 258, 405-417.

Ando, Y., and Pompoli, R. (2002). Factors to be measured of environmental noise and its subjective responses based on the model of auditory-brain system. Journal of Temporal Design in Architecture and the Environment, 2, 2-12. http://www.jtdweb.org/journal/

Ando, Y. (2003). Investigations on cerebral hemisphere activities related to subjective preference of the sound field, published for 1983 - 2003. Journal of Temporal Design in Architecture and the Environment, 3, 2-27. http://www.jtdweb.org/journal/

Ando, Y. (2007a). Concert Hall Acoustics Based on Subjective Preference Theory, in Handbook of Acoustics, Ed., Thomas Rossing, Springer-Verlag, New York, Chapter 10.

Ando, Y. (2007b). Neural Based Theory of Temporal and Spatial Sensations of the Visual Field. Proceedings of the 3rd International Symposium on Temporal Design, Guangzhou, China, November 2007.

Buchwald, J.A.S., and Huang, C.M. (1975). Far-field acoustic response: origins in the cat. Science, 189, 382-384.

Cariani, P.A., and Delgutte, B. (1996a). Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. Journal of Neurophysiology, 76, 1698-1716.

Cariani, P. A., and Delgutte, B. (1996b). Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase-invariance, pitch circularity, and the dominance region for pitch. Journal of Neurophysiology, 76, 1717-1734.

Cariani, P. (2001). Temporal coding of sensory information in the brain. Acoustical Science and Techechonoly, 22, 77-84.

Fujii, K. Soeta, Y., and Ando, Y. (2001). Acoustical properties of aircraft noise measured by temporal and spatial factors. Journal of Sound and Vibration, 241, 69-78.

Hanada, K., Kawai, K., and Ando, Y. (2007). A study of the timbre of an electric guitar sound with distortion. Proceedings of the 3rd International Symposium on Temporal Design, Guangzhou, November 2007 published on Journal of South China University of Technology, 35, 96-99.

Hecox, K., and Galambos, R. (1974). Brain stem auditory evoked responses in human infants and adults. Archives Otolaryngology. 99, 30-33.

Hotehama, T., Sato, S., and Ando, Y. (2002). Dissimilarity judgments in relation to temporal and spatial factors for the sound fields in an existing hall. Journal of Sound and Vibration, 258, 429-441.

Inoue, M., Ando, Y., and Taguti, T. (2001). The frequency range applicable to pitch identification based upon the auto-correlation function model. Journal of Sound and Vibration, 241, 105-116.

Lev, A., and Sohmer, H. (1972). Sources of averaged neural responses recorded in animal and human subjects during cochlear audiometry (Electro-cochleo-gram). Arch. Klin. Exp. Ohr., Nas.-u. Kehlk. Heilk., 201, 79-90.

Jewett, D.L. (1970). Volume-conducted potentials in response to auditory stimuli as detected by averaging in the cat. Electroencephalography and Clinical Neurophysiology, 28, 609-618.

Keet, M. V. (1968). The influence of early lateral reflections on the spatial impression. Proceedings of the 6th International Congress on Acoustics, Tokyo, Paper E-2-4.

Kitamura, T., Shimokura, R., Sato, S., and Ando, Y. (2002). Measurement of temporal and spatial factors of a flushing toilet noise in a downstairs bedroom. Journal of Temporal Design in Architecture and the Environment, 2, 13-19. (http://www.jtdweb.org/)

Marui, A., and Martens, W. L. (2005). Constructing individual and group timbre space for sharpness-matched distorted guitar timbres. Audio Engineering Society Convention Paper, Presented at the 119th Convention, New York.

Mouri, K., Akiyama, K., and Ando, Y. (2001). Preliminary study on recommended time duration of source signals to be analyzed, in relation to its effective duration of autocorrelation function. Journal of Sound and Vibration, 241, 87-95.

Saifuddin, K., Matsushima, T., and Ando, Y. (2002). Duration sensation when Listening to pure tone and complex tone. Journal of Temporal Design in Architecture and the Environment, 2, 42-47. http://www.jtdweb.org/journal/

Sato, S., and Ando, Y. (1996). Effects of interaural crosscorrelation function on subjective attributes. Journal of the Acoustical Society of America, 100 (A), 2592.

Sato, S., Ando, Y., and Mellert, V. (2001). Cues for localization in the median plane extracted from the autocorrelation function. Journal of Sound and Vibration, 241, 53-56.

Sato, S., Kitamura, T., and Ando, Y. (2002). Loudness of sharply (2068 dB/Octave) filtered noises in relation to the factors extracted from the autocorrelation function. Journal of Sound and Vibration, 250, 47-52.

Sato, S., Kitamura, T., and Ando, Y. (2004). Annoyance of noise stimuli in relation to the spatial factors extracted from the interaural cross-correlation function. Journal of Sound and Vibration, 277, 511-521.

Soeta, Y., Nakagawa, S., Tonoike, M., and Ando, Y. (2002). Magnetoencephalographic responses corresponding to individual subjective preference of sound fields. Journal of Sound and Vibration, 258, 419-428.

Soeta, Y., Maruo, T., and Ando, Y., (2004). Annoyance of bandpass filtered noises in relation to the factor extracted from autocorrelation function. Journal of Acoustical Society of America, 116, 3275-3278.

Soeta, Y., and Nakagawa, S. (2006). Auditory evoked magnetic fields in relation to interaural time delay and interaural correlation. Hearing Research, 220, 106-115.

Sperry, R.W. (1974). Lateral specialization in the surgically separated hemispheres. The Neurosciences: Third study program, Eds. Schmitt, F.O., and Worden, F.C. MIT Press, Cambridge, Chapter 1.

Sumioka, T., and Ando, Y. (1996). On the pitch identification of the complex tone by the autocorrelation function (ACF) model. Journal of the Acoustical Society of America, 100 (A), 2720.

Thompson, A.M., and Thompson, G.C. (1988). Neural connections identified with PHA-L anterograde and HRP retrograde tract-tracing techniques. Journal of Neuroscience Methods, 25, 13-17.

Yost, W. A. (1996). A time domain description for the pitch strength of iterated rippled noise. Journal of the Acoustical Society of America 99, 1066-1078.

Zwicker, E., and Fastl, H. (1999). Psychoacoustics. Springer-Verlag, New York.