BRAIN ORIENTED ACOUSTICS

YOICHI ANDO
Kobe University, JAPAN

Abstract – Approach of previous human related acoustics was from only our outside physics and psychology. Now, we are attempting to approach inside neural and brain science, in addition. According to two scientific approaches, we could subserve the percept correctly. The cerebral hemisphere specialization may play an important role for the independent effects on temporal and spatial sensations of the temporal and spatial factors. The temporal and spatial factors extracted from respective correlation functions well describe primitive temporal and spatial sensation, respectively, as well as over all attributes including subjective preference and annoyance. These approaches may be applied for vision and other modalities due to possible similarities in the correlation based neural information processing mechanisms. In fact, it has been discussed that the scale value of subjective preference of both the sound field and the visual field are described by the temporal and spatial factors, which may be observed by alpha rhythms in cortical populations (Ando, 2009).

1. INTRODUCTION

Since 1974, we have investigated subjective preference of the sound field (Ando, 1985; 1998). This article briefly reviews of the signal-processing model of human auditory system, theory of subjective preference and auditory primary sensations described in the book, auditory and visual sensations (Ando, 2009a). For development of current of acoustics, applications for noise and music are discussed here.

First of all, reasons why subjective preference was selected for evaluation of the sound field avoiding ill-defined adjectives are:

1) Subjective preferences are the most primitive response;
2) Preferences are the evaluative judgments;
3) Preferences guide the organism in the direction of maintaining life, thus we may observe brain activities related to preference;
4) It is deeply related to an individual and global aesthetic sense (Schroeder, 1966);
5) Preference is a representation of individual personality and further a source of creation (Ando, 2006).

2. THEORY OF SUBJECTIVE PREFERENCE AND APPLICATIONS

It has been identified four orthogonal factors of the sound field in a room (Ando, 1985; 1998; 2007). Two of those are temporal factors, i.e., the initial time delay gap between the direct sound and the first reflection (Δt_1) and the subsequent reverberation time (T_{sub}). Other two are spatial factors, i.e., the binaural listening level (LL) and the magnitude of interaural crosscorrelation (IACC).

2.1 Subjective Preference Judgments and Resulting Theory

Subjective preference judgments of the simplest sound field consisting of only the direct sound and the single reflection were first conducted. Such a single reflection was considered as a representative of multiple and reverberant reflections. From such a sound field, we expected to obtain a basic result.

2.1.1 Preferred Delay Time of a Single Reflection

The sound field consists of the frontal direct sound and a single reflection from a fixed direction (the horizontal angle from the front $\xi = 36^{O}$ and the elevation $\eta = 9^{O}$). The delay time Δt_1 was adjusted in the range of 6 - 256 ms. The paired comparison test was performed in Goettingen for all pairs in an anechoic chamber using subjects of normal hearing ability with two different music motifs, A and B (Table 2.1).

Sound source [1]	Title	Composer or writer	τ_e[2] [ms]	$(\tau_e)_{min}$[3] [ms]
Music motif A	Royal Pavane	Orlando Gibbons	127 (127)	125
Music motif B	Sinfonietta, Opus 48; IV movement	Malcolm Arnold	43 (35)	40
Speech S	Poem read by a female	Doppo Kunikida	10 (12)	

Table 2.1 Music and speech source signals used and their effective duration of the long-term ACF, τ_e measured in the early investigations (Ando, 1977; Ando and Kageyama, 1997), and the minimum value of running ACF, $(\tau_e)_{min}$ (Ando, 1998).

[1] The left channel signals of original recorded signals (Burd, 1969) were used.
[2] Values of τ_e differ slightly with different radiation characteristics of loudspeakers used; thus all of physical factors must be measured at the same condition of the hearing tests, 2T = 35 s;
[3] The value of $(\tau_e)_{min}$ is defined by the minimum value in the running or short-moving ACF, for this analyses 2T = 2 s, with a running interval of 100 ms. Recommended 2T has been investigated (Mouri, Akiyama and Ando, 2001). Subjective judgments may be made at the most active piece of sound signals.

For the purpose of obtaining the most preferred conditions, the total score is simply obtained by accumulating the scores giving +1 and –1 corresponding to positive and negative judgments, respectively, in this section. The total score is divided by S (F – 1) for normalization, where S is the number of subjects and F is the number of sound fields. The normalized scores and the percentage of preference of the sound field as a function of the delay are shown in FIGURE 2.1 (Ando, 1977; Ando and Morioka; 1981, Kang and Ando, 1985).

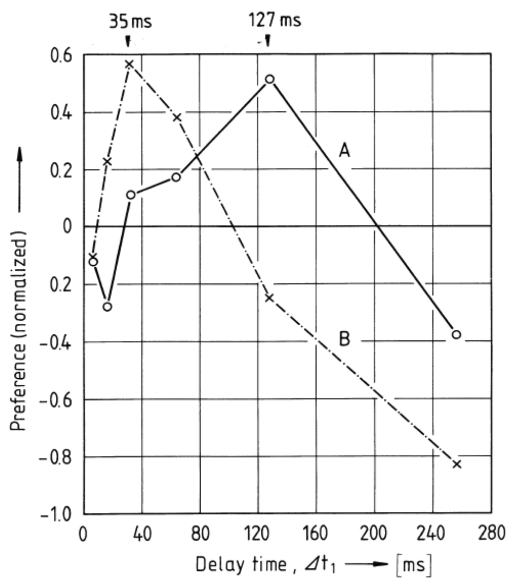


Figure 2.1 a)

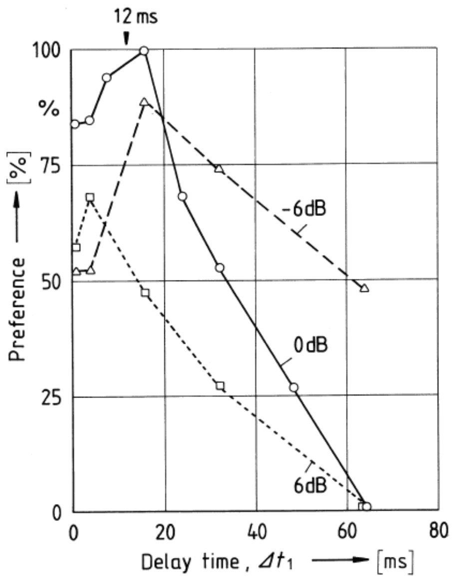


Figure 2.1 b)

Obviously, the most preferred delay time, with the maximum score, differs greatly between the two motifs and continuous speech. With an amplitude of reflection, $A_1 = 1$, the most preferred delays are around 130 ms for music motif A, 35 ms for music motif B (FIGURE 2.1a), and 16 ms for speech (FIGURE 2.1b).

It is found that this corresponds well to the effective duration of the ACF of source signals of 127 ms (motif A), 35 ms (motif B) and 12 ms (speech) as indicated in Table 2.1. After inspection, the preferred delay is found in relation to the effective duration of the ACF, defined by τ_p, such that the envelope of the ACF becomes $0.1A_1$. Thus, $\tau_p = \tau_e$ only when $A_1 = 1$. The data collected as a function of the duration τ_p are shown in FIGURE 2.2, where data from continuous speech signal of τ_e = 12 ms are also plotted. When the envelope of the ACF is exponential, then it is expressed approximately by (Ando; 1977; 1985)

$$\tau_p = [\Delta t_1]_p \approx (1 - \log_{10} A_1)\tau_e \qquad (2.1a)$$

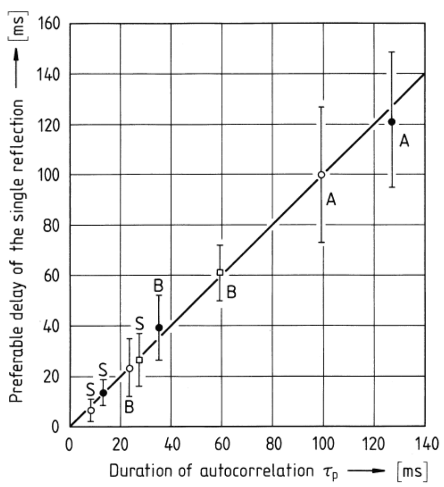


Figure 2.2

It is worth noticing that the amplitude of reflection A_1 relative to that of the direct sound in Equation (2.1) should be measured by the most accurate method, i.e., the square root of the ACF at the origin of the delay time: $[\Phi(0)]^{1/2}$ of the reflection in relative to that of the direct sound.

Two reasons can be given for explaining why the preference decreases for the short delay range of reflection, $0 < \Delta t_1 < \tau_p$ (FIGURE 2.3):

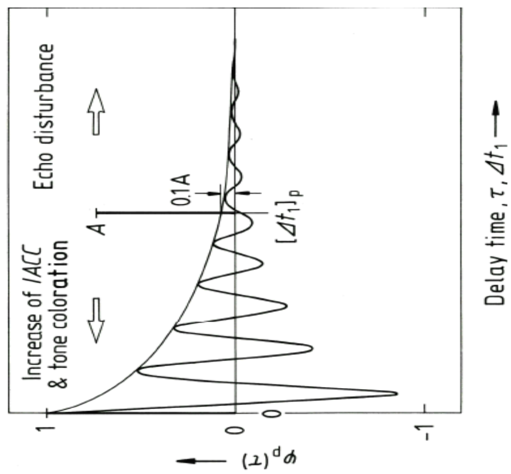


Figura 2.3

1. Tone coloration effects occur because of the interference phenomenon in the coherent time region; and

2. The IACC increases when Δt_1 is near 0.

On the other hand, echo disturbance effects can be observed when Δt_1 is greater than τ_p.

Demonstration of this experiment may be available by use of Quick Time 7, try at: http://www.andolab.org/demo.html.

2.1.2 Preferred Horizontal Direction of a Single Reflection to a Listener

The delay time of the reflection, in the experiment showing the preferred direction of the single reflection, was fixed at 32 ms. The reflection was presented by the loudspeaker located at $\xi = 18^O$ ($\eta = 9^O$), 36^O,…, 90^O, and $\xi = 0^O$ ($\eta = 27^O$).

Results of the preference test are shown in FIGURE 2.4.

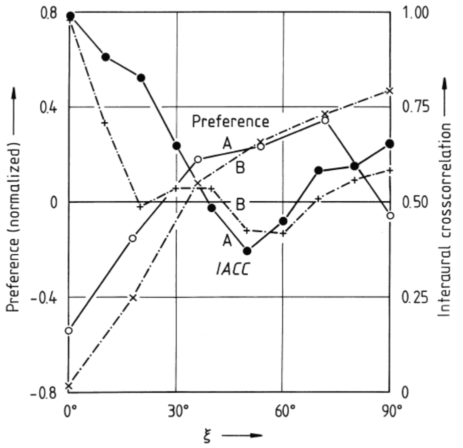


Figure 2.4

No fundamental differences are observed between curves for two music motifs in spite of the great difference between values of τ_e. The preferred score increases roughly with a decreasing IACC. The correlation coefficient between the score and the IACC is – 0.8 ($p < 0.01$). The score with motif A at $\xi = 90^{O}$ drops to a negative value, indicating that the lateral reflections, coming only from around 90^{O}, are not always preferred. The figure shows that there is a preference for angles greater than $\xi = 30^{O}$, and on average there may be an optimum range centered on about $\xi = 55^{O}$. Similar results can be seen in the data from speech signals (Ando and Kageyama, 1977). Therefore, the dissimilarity of the binaural signals is the preferred condition (Damaske and Ando, 1972; Schroeder, Gottlob and Siebrasse, 1974).

2.1.3. Optimal Conditions Maximizing Subjective Preference

A) Listening Level (LL)

The binaural listening level (LL) is, of course, the primary factor for listening to the sound field in the concert hall. The preferred listening level depends upon the music and the particular passage being performed. For the two extreme music signals as indicated in Table 2.1, the gross preferred levels obtained by 16 subjects are in the peak ranges of 77 - 79 dBA for music motif A, (Royal Pavane by Gibbons), for a slow tempo, and 79 - 80 dBA for music motif B (Sinfonietta by Arnold), with a fast tempo.

B) Early Reflections After the Direct Sound (Δt_1)

An approximation for the most preferred delay time has been expressed in terms of the effective duration of the ACF of the source signal and the total amplitude of reflections (Ando, 1985). When the envelope of the ACF has an exponential decay,

$$[\Delta t_1]_p = (\log_{10}\frac{1}{k} - c\log_{10} A)\tau_e \qquad (2.1b)$$

where k and c are constants, which depend on the subjective attributes as listed in Table 2.2.

Subjective attributes	In Equation (2.1b) k	c	Delay time to be obtained	Amplitude examined [dB]	Source signals	Authors investigated
Preference of listeners	0.1	1	Preferred delay time		Speech & music	Ando (1977)
Threshold of perception of reflection	$10^{-5/2}$	1	Critical delay time		Speech	Seraphim (1961)
50%-echo disturbance	0.01	4	Disturbed delay time		Speech	Haas (1951); Ando et al. (1973)
Coloration		2	Critical delay time		Gaussian noise	Ando and Alrutz (1982)
Preference of Alto-rec order	2/3	1/4	Preferred delay time	[illegible]	Music	Nakayama (1984)
Preference of Cello	1/2	1	Preferred delay time		Music	Sato, Ohta and Ando (2000)

Table 2.2: Constants in Equation (2.1b) related to the ACF envelope of source signals for calculating various subjective responses to the sound field with a single reflection.

And, the total pressure amplitude of reflection is given by:

$$A = \left[A_1^2 + A_2^2 + A_3^2 + \ldots\right]^{1/2} \qquad (2.2)$$

The relationship for the sound field with a single reflection only may be obtained by where $A = A_1$, $k = 0.1$ and $c = 1$, so that

$$\tau_p = [\Delta t_1]_p \approx (1 - \log_{10} A_1)\tau_e \qquad (2.3)$$

This is identical with Equation (2.1a).

Later, it has been clarified that the value of τ_e in Equations (2.1b) and (2.3) is replaced by $(\tau_e)_{min}$ of the running ACF (Ando, Okano and Takezoe, 1989; Mouri, Akiyama and Ando, 2000). The minimum value of τ_e is observed at the music piece of the most active part and it is realized by signals without any redundancy. It may contain such as vibrato and staccato in the musical flow and an impulsive sound signal. The echo disturbance, therefore, are easily perceived at the signal piece occurring $(\tau_e)_{min}$. Even for a long music composition, the musical flow might be divided into such a short piece, so that the most minimum part of $(\tau_e)_{min}$ of ACF in the whole music may be taken into consideration. This value is useful for choosing the music program to be performed in a given concert hall. The same is true for the preferred reverberation time as is given by Equation (2.4).

Methods for controlling the $(\tau_e)_{min}$ for vocal music performances have been discussed (Taguti and Ando, 1997; Kato and Ando, 2002; Kato, Fujii, Kawai, Ando and Yano, 2004). If vibrato is included during singing, for example, we may make effectively decrease the $(\tau_e)_{min}$.

C) Subsequent Reverberation Time after the Early Reflections (T_{sub})

For flat frequency characteristics of reverberation, one of the most preferred conditions, (Ando, Okano and Takezoe, 1989), the preferred subsequent reverberation time is expressed approximately by (Ando, Okura and Yuasa, 1982; Ando, Otera and Hamana, 1983)

$$[T_{sub}]_p \approx 23\tau_e \qquad (2.4)$$

Experimental conditions were that the total amplitude of reflections A was in the range of 1.1 and 4.1, which covers the usual conditions of a sound field in a room. A lecture and conference room must be designed for speech; an opera house must be designed primarily for vocal music but also orchestral music. For orchestral music, it is recommended to select the concert hall from one of two or three types of concert halls according to the effective duration of the ACF. For example, Symphony No. 41 by Mozart, "Le Sacre du Printemps" by Stravinsky and Arnold's Sinfonietta have short values of $(\tau_e)_{min}$. On the other hand, Symphony No. 4 by Brahms and Symphony No. 7 by Buckner are typical of orchestra music (B). Much longer values of $(\tau_e)_{min}$ are typical for pipe organ music, for example, by Bach. Thus, the most preferred reverberation time for each sound source given by Equation (2.4) plays an important role for the selection of the music program to be performed. Considering the fact that the value of τe is obtained at the tenth percentile (or -10 dB) delay of the ACF envelope of a source signal, the - 60 dB delay time of the ACF envelope corresponds roughly to the "reverberation time" contained in the source signal itself, given by ($6\tau e$). This means that the most preferred reverberation time of sound fields expressed by Equation (2.4) implies about 4 times the "reverberation time" containing the source signal itself.

D) Magnitude of the IACF (IACC)

All available data with listeners of normal hearing ability indicate a negative correlation between the magnitude of the IACC and subjective preference, i.e., it has been reconfirmed that a dissimilarity of signals arriving at the two ears is preferred. This holds under the condition that the maximum value of the interaural cross-correlation function (IACF) is maintained at the origin of the time delay, keeping balance of the sound field for two ears. If not, then an image shift of the source may occur. To obtain a small magnitude of IACC in the most effective manner, the directions from which the early reflections arrive at the listener should be kept within a certain range of angles from the median plane centered on $\xi = \pm 55^O$. It is obvious that the sound arriving from the median plane makes the IACC greater. Sound arriving from $\xi = \pm 90^O$ in the horizontal plane is not always advantageous, because the similar "detour" paths around the head to both ears cannot decrease the IACC effectively, particularly for frequency ranges higher than 500 Hz. For example, the most effective angles for the frequency ranges of 1 kHz and 2 kHz are about centered on $\xi = \pm 55^O$ and $\xi = \pm 36^O$, respectively. To realize these conditions simultaneously, a geometrical uneven surface for the sidewalls in the concert hall has been proposed (Ando and Sakamoto, 1988).

2.1.4. Theory of Subjective Preference

Let us now put these results into theory. Since the numbers of orthogonal acoustic factors, which are included in the sound signals at both ears, are limited as mentioned in Table 3.1, the scale value of any one-dimensional subjective response may be expressed by

$$S = f(x_1, x_2, \ldots x_I). \qquad (2.5)$$

Here after, the linear scale value of subjective preference obtained by the law of comparative judgment or the PCT is obtained (Thurstone, 1927; Mosteller, 1951; Gullikson, 1956; Torgerson, 1958) for a number of subjects as well as for the individual subjective preference (Ando, 1998). It has been verified, through a series of experiments, that four objective factors act independently of the scale value when changing two of the four factors simultaneously. Results indicate that units of scale values obtained by a series of experiments with different source signals were appeared to be almost constant, so that we may add scale values to obtain the total scale value (Ando, 1983),

$$\begin{aligned} S &= f(x_1) + f(x_2) + f(x_3) + f(x_4) \\ &= S_1 + S_2 + S_3 + S_4 \end{aligned} \qquad (2.6)$$

where Si , i = 1, 2, 3, 4 is the scale value obtained relative to each objective factor. Equation (2.6) indicates a four-dimensional continuity.

From the nature of the scale value, it is convenient to place a value of zero at the most preferred conditions, as shown in this figure. These results of the scale value of subjective preference obtained from the different experimental series; using different music programs, yield the similar, when each factor is normalized by the most preferred value. The following common formula is given:

$$S_i \approx - \alpha_i \,|\, x_i \,|^{3/2}, \; i = 1, 2, 3, 4 \qquad (2.7)$$

where x_i is the normalized factor, and the values of α_i are weighting coefficients as listed in Table 2.3. If α_i is close to zero, then a lesser contribution of the factor x_i on subjective preference is signified.

i	x_i	α_i $x_i > 0$	$x_i < 0$
1	$20\log P - 20\log[p]_p$ (dB)	0.07	0.04
2	$\log(\Delta t_1/[\Delta t_1]_p)$	1.42	1.11
3	$\log(T_{sub}/[T_{sub}]_p)$	0.45 + 0.75A	2.36 - 0.42A
4	IACC	1.45	---

Table 2.3: Four orthogonal factors of the sound field, and its weighting coefficients α_i in Equation (2.7), which was obtained with a number of subjects.

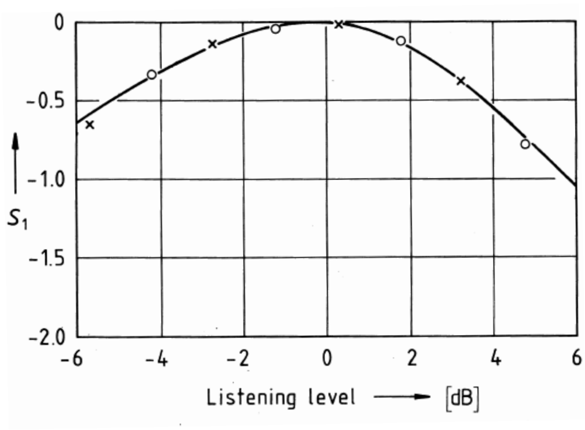


Figure 2.5a

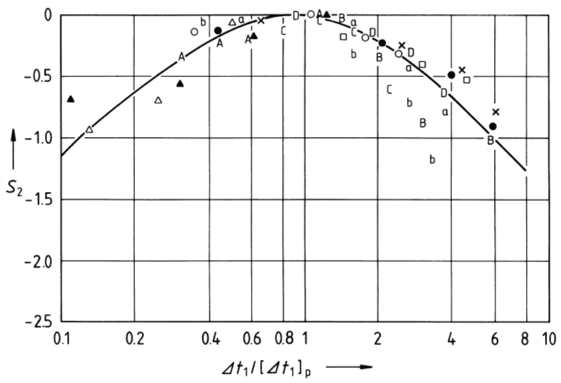


Figure 2.5b

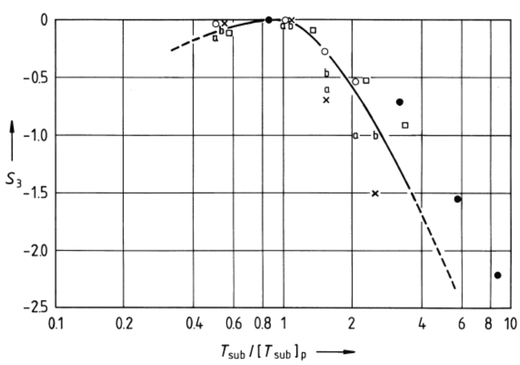


Figure 2.5c

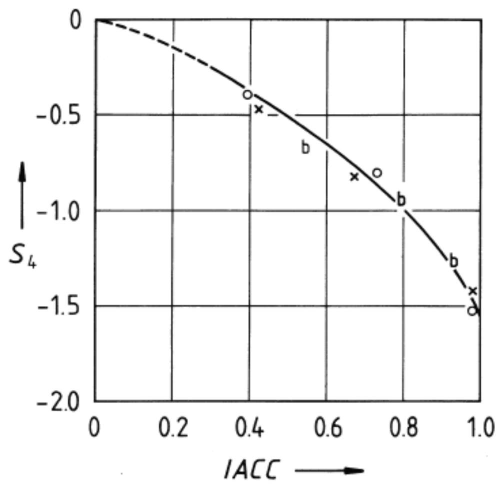


Figure 2.5d

The factor x_1 is given by the sound pressure level difference, measured by the A-weighted network, so that,

$$x_1 = 20\log P - 20\log[P]_p \tag{2.8}$$

P and $[P]_p$ being the sound pressure or the listening level (LL) at a specific seat and the most preferred sound pressure that may be assumed at a particular seat position in the room under investigation;

$$x_2 = \log(\Delta t_1 / [\Delta t_1]_p) \tag{2.9}$$

$$x_3 = \log(T_{sub} / [T_{sub}]_p) \tag{2.10}$$

$$x_4 = \text{IACC} \tag{2.11}$$

The values of $[\Delta t_1]_p$ and $[T_{sub}]_p$ may be calculated by Equations (2.2) and (2.4), respectively.

The scale value of preference has been formulated approximately in terms of the 3/2 power of the normalized factor, expressed in the logarithm for the normalized factors, x_1, x_2 and x_3. The remarkable fact is that the spatial binaural factor x_4 = IACC is expressed in terms of the 3/2 powers of its real values, indicating a greater contribution than those of the temporal parameters. The scale values are not greatly changed in the neighborhood of the most preferred conditions, but decrease rapidly outside of this range. Since the experiments were conducted to find the optimal conditions, this theory holds in the range of preferred conditions tested for the four orthogonal factors. Under the conditions of fixing Δt_1 and T_{sub} around the preferred conditions, for example, the scale value of subjective preference calculated by Equation (2.6) for the LL and the IACC is demonstrated in FIGURE 2.6. Agreement between calculated values and observed ones are satisfactory, so that the independency of the LL and the IACC on the scale value is achieved (Ando and Morioka, 1981). The same is true for the other two factors (Ando, 1985).

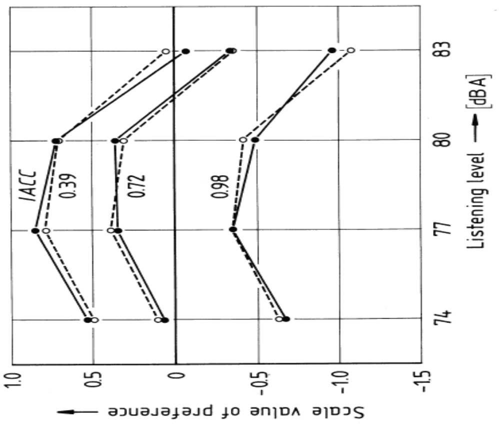


Figure 2.6

2.2 Design Study of a Concert Hall

First of all, as a purpose of the concert hall to be designed, specific music programs classified by the value of $(\tau_e)_{min}$, the temporal factor relating to its dimensions and the absorption coefficient of walls(Δt_1 and T_{sub}) should be carefully selected. Using the scale value of subjective preference in the four orthogonal factors of the sound field obtained by a number of listeners, the "principle of superposition" expressed by Equation (2.6) can be applied for calculating the scale value at each seat. Comparison of the total preference values for different configurations for a concert hall allows a designer to choose the best scheme. In this section, we discuss mainly the optimum space form of the hall due to the spatial factor, the IACC and LL.

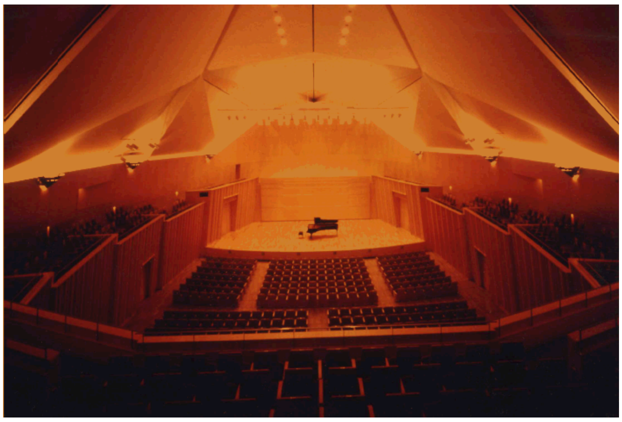


Photo 2.1

2.2.1. Design Studies by Genetic Algorithms (GA) Optimizing the Shape

The theory of subjective preference described in Section 2.1.4 allows us to evaluate the sound field in terms of the four orthogonal acoustical factors. The linear scale value of subjective preference has been obtained by using the law of comparative judgment. The units of the scale value derived from a series of experiments with different sound sources and different subjects were almost constant, so the scale values may be added as expressed by Equation (2.6). Genetic algorithms were applied to the problem of optimizing the shape of a concert hall. The first model was an optimization of the proportions of a hall of the typical shoebox type. Results show that the optimized form is similar to the Grosser Musikvereinsaal in Vienna. The second model is the optimization of the shape based on the first results with a number of portions to be modified. The typical result of the maximization of audience subjective preference values in relation to listening level and spatial clarity is a kind of leaf-shaped plan.

Our goal is to design a structure that has temporal and spatial factors that best satisfy both the left and right human cerebral hemispheres for each listener. Genetic algorithms (Holland, 1975), a form of evolutionary computing, have been applied to the design of concert halls (Sato, Otori, Takizawa, Sakai, Ando and Kawamura, 2002; Sato, Hayashi, Takizawa, Tani, Kawamura and Ando, 2004). In architectural optimization using genetic algorithms, an initial set of “parent” building structures is specified by a set of “genetic” parameters that are clustered together on “chromosomes.” Through use of a “pattern-grammar,” each set of genes specifies the construction of some virtual structure (i.e. room form, dimensions, materials). A population of individual, genetically unique structures is generated, and then the fitness of each structure vis-à-vis some function is evaluated (via some calculated “fitness function”). In this case, the fitness of a given architectural structure can be calculated from the ensemble of seat preference values that are generated by acoustical simulations and the auditory signal-processing model. Those individuals with the highest fitness values are selected for mating and reproduction, to be the parents of the next generation. Variability is introduced into the genes of “offspring” via operations of “mutation” and “crossover”. A new population of individuals based on a reshuffling of the genes of the fittest individuals of the previous generation is created, and the variation-evaluation-selection cycle continues on until either a satisfactory level of performance is reached or no further improvements can be made with the options at hand.

Here the GA system was used to generate alternative schemes for evolutionary selection (seen on the left hand side of FIGURE 2.7), and these alternative schemes can be evaluated in terms of the preference scale values in relation to the spatial factors associated with the right hemisphere that they would be expected to produce. Thus, the GA can search through the complex space of possible room shapes, simulate their acoustic properties and their consequent effects on audience satisfaction levels, and find ever better acoustic architectural designs.

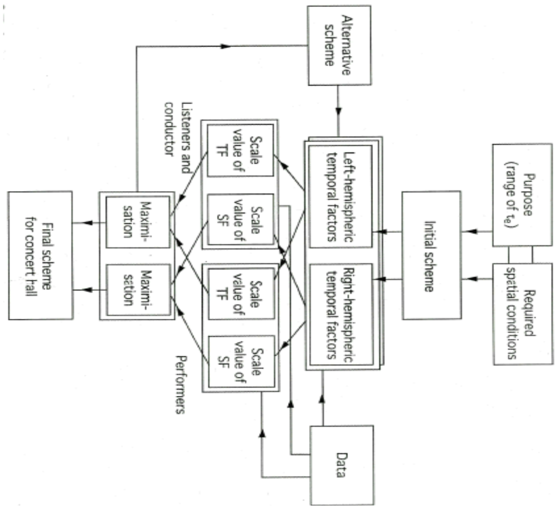


Figure 2.7

The behavior of the scale value in relation to each orthogonal factor S_i is given by Equations (2.6) through (2.11). Here, the parameters x_i and coefficients α_i are listed in Table 2.3. In this calculation, linear scale values of subjective preference S_1 and S_4 are employed as fitness function due to the LL and IACC, respectively, because the geometrical shape of a hall is directly affected by these spatial factors. The spatial factor for a source on the stage was calculated at a number of seating positions. For the sake of convenience, the single omni-directional source was assumed to be at the center of the stage, 1.5 m above the stage floor. The receiving points that correspond to the ear positions were 1.1 m above the floor of the hall. The image method was used to determine the amplitudes, delay times, and directions of arrival of reflections at these receiving points. Reflections were calculated up to the second order. In fact, there was no change in the relative relationship among the factors obtained from calculations performed up to the second, third, and fourth order of reflection. The averaged values of the IACC for five music motifs (Motifs A through E: Ando, 1985) were used for calculation.

Those hall shapes that produced greater scale values are selected as parent chromosomes. An example of the encoding of the chromosome is given in FIGURE 2.8. The first bit indicated the direction of motion for the vertex. The other (n-1) bits indicated the range over the vertex moved. To create a new generation, the room shapes are modified and the corresponding movement of the vertices of the walls is encoded in the chromosomes, i.e., binary strings. After GA operations that includes crossover and mutation, new offspring are created. The fitness of the offspring is then evaluated in terms of the scale value of subjective preference. This process is repeated until the end condition of about two thousand generations is satisfied.

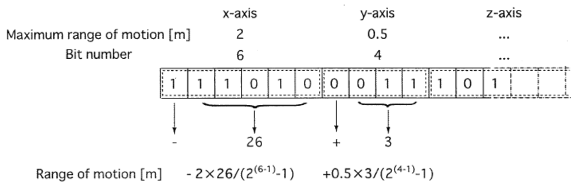


Figure 2.8

First, the proportions of the shoebox hall were optimized (Model 1). The initial geometry is shown in FIGURE 2.9. In its initial form, the hall was 20-m wide, the stage was 12-m deep, the room was 30-m long, and the ceiling was 15 m above the floor. The point source was located at the center of the stage and 4.0 m from the front of the stage and seventy-two listening positions were selected. The range motion for each sidewall and the ceiling was ±5 m from the respective initial positions, and the distance through which each was moved was coded on the chromosome of the GA. Scale values at the listening positions other than those within 1 m of the sidewalls were included in the averages (S_1 and S_4). In this calculation, the most preferred listening level, $[LL]_p$, was chosen at the frontal seat near the stage.

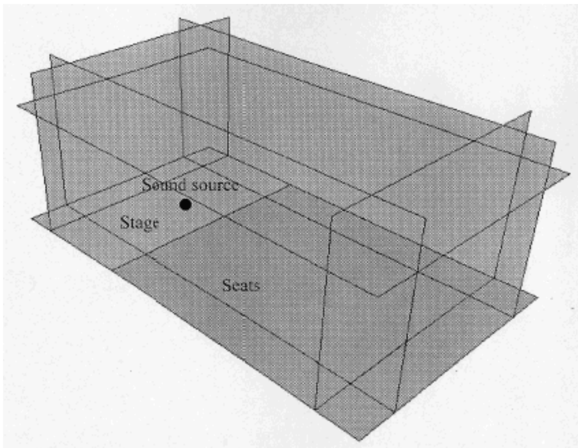


Figure 2.9

Results of optimization of the hall for S_1 and S_4 are shown in FIGURE 2.10 (a) and (b), respectively. The width and length were almost the same in the two results, but the respective heights indicated quite different. The height of the ceiling that maximizes S_1 was as low as possible within the allowed range of motion to obtain a constant LL (FIGURE 2.10a). The height that maximizes S_4, on the other hand, was at the upper limit of the allowed range of motion to obtain small values of IACC (FIGURE 2.10b).

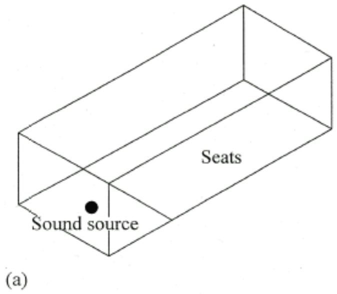


Figure 2.10a

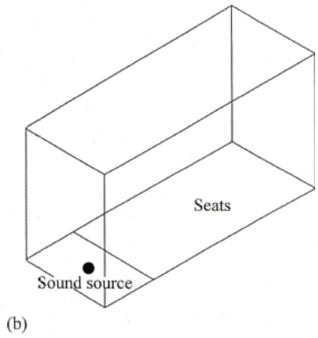


Figure 2.10b

The optimization for Model 1 produced optimized proportions for the shoebox form. Table 8.2 shows the comparison of the proportions we obtained and those of the Grosser Musikvereinsaal, which is an example of an excellent concert hall. The length/width ratios are almost the same. For the ceiling of the hall, the height that maximized S_1 was the lowest within the allowed range of motion (FIGURE 2.10a). This is due to the fact that more energy should be provided from the ceiling to the listening position. To maximize S_4, on the other hand, the ceiling took on the maximum height in the possible range of motion (FIGURE 2.10b). The reflection from the sidewalls were useful decreasing the IACC, however, this was not true of the reflections from the flat ceiling. The reflection from the flat ceiling was not useful to decrease the IACC, but from the sidewalls.

Next, to obtain more excellent sound fields, a slightly more complicated form (Model 2) as shown in FIGURE 2.11 was examined. The floor plan optimized from the above results was applied as a starting point. The hall in its initial form was 14-m wide, the stage was 9-m deep, the room was 27-m long, and the ceiling was 15 m above the stage floor. The sound source was again 4.0 m from the front of the stage, but was 0.5 m to one side of the centerline and 1.5 m above the stage floor. The front and rear walls were vertically bisected to obtain two faces, and each stretch wall along the side of the seating area was divided into four faces. The walls were kept vertical (i.e., tilting was not allowed) to examine only the plan of the hall in terms of maximizing S_1 and S_4. Forty-nine listening positions distributing throughout the seating area on a 2 x 4 m grid were selected. In the GA operation, the sidewalls were moved so that any of these listening positions were not excluded. The moving range of each vertex was 2 m in the direction of the line normal to the surface. The coordinates of the two bottom vertices of each surface were encoded on the chromosomes for the GA. In this calculation, the most preferred listening level was set for a point on the hall's long axis (central line), 10 m from the source position.

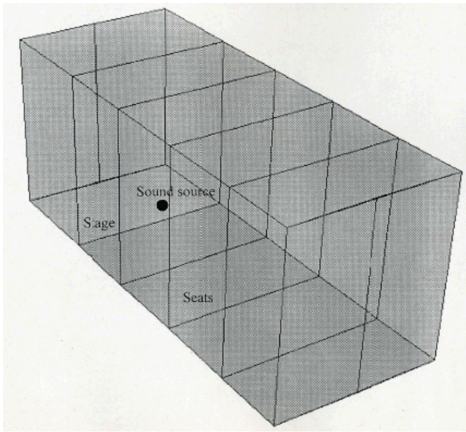


Figure 2.11

The result of optimizing the hall for S_1 is shown in FIGURE 2.12 and contour lines of equal S_1 values are shown in FIGURE 2.13. To maximize S_1, the rear wall of the stage and the rear wall of the audience area took on concave shapes. The result of optimizing for S_4 is shown in FIGURE 2.14 and contour lines of equal S_4 values are shown in FIGURE 2.15. To maximize S_4, on the other hand, the rear wall of the stage and the rear wall of the audience area took on convex shapes. Any way, it has been approved that a "leaf-shape" plan of concert hall is one of the most optimized schemes, which had been applied in design of the Kirishima International Concert Hall, which opened in 1994.

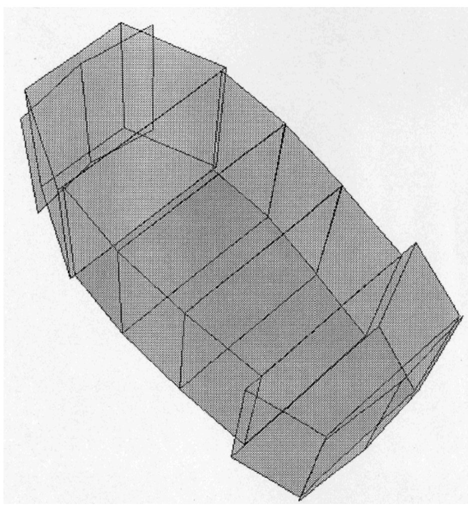


Figure 2.12

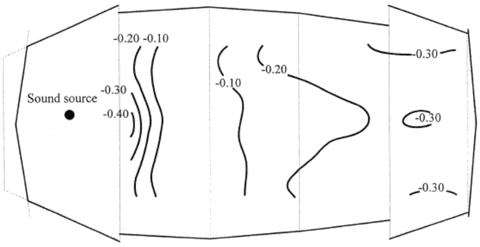


Figure 13

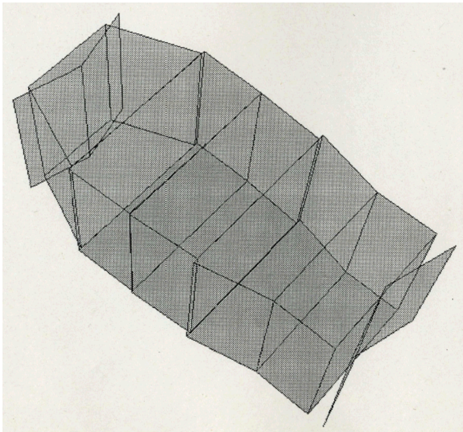


Figure 14

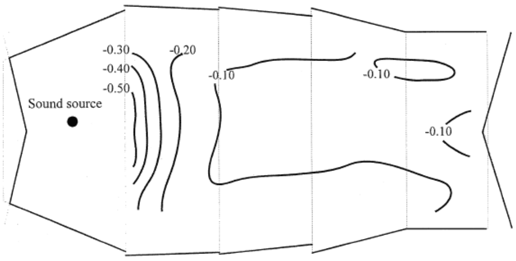


Figure 15

As for the conflicting requirements for S_1 and S_4, the maximization of S_4 may take a higher priority than S_1 because the preference increases with a decreasing IACC for all subjects tested (Singh, Ando and Kurihara, 1994), while there is a large individual difference in the preferred LL (Sakai, Singh and Ando, 1997). Thus, listeners can choose the seat with respect to their preferred LL, when they buy tickets for concert.

As a temporal design, if the conductor or music director is aware of the acoustic of a concert hall, they can plan a program of music that will sound best in that hall in terms of the temporal factors. This relates to the minimum value of effective duration of the ACF of source signals, $(\tau_e)_{min}$. It has been found, for example, that music with rapid sound movements or vibrato can decrease the value of $(\tau_e)_{min}$ and best fit to a concert hall with a short initial time delay gap Δt_1, and a short subsequent reverberation time T_{sub}. Music with a slow tempo usually sounds best in a hall with relatively long values for the factor Δt_1 and T_{sub}. An ideal application of this principle would allow the architect, concert hall manager, and music director to collaborate and actually change the configuration of a given concert hall to suit a specific music program. A "sound coordinator" could select a program of music blending acoustic to the given concert hall.

2.2.2. Actual Design Studies

An example of applying this design theory was performed in the Kirishima International Concert Hall, in cooperation with the architect Fumihiko Maki in 1992 as shown in FIGURE 2.16 (Maki, 1997, Ando, Sato, Nakajima and Sakurai, 1997; Nakajima and Ando, 1997). Acoustic design elements were as follows (Ando, 1998, 2007): 1) A leaf-shape plan was applied, 2) the sidewalls were tilted, and 3) the ceiling consisted of triangular plates. These realized a small value of the IACC at nearly every seat (Ando, 1998; 2007).

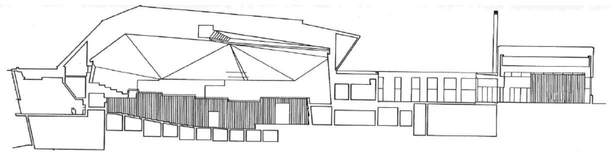


Figure 2.16a

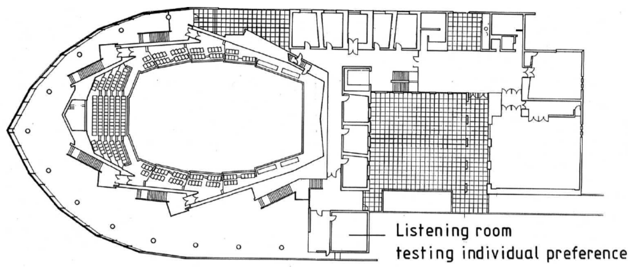


Figure 2.16b

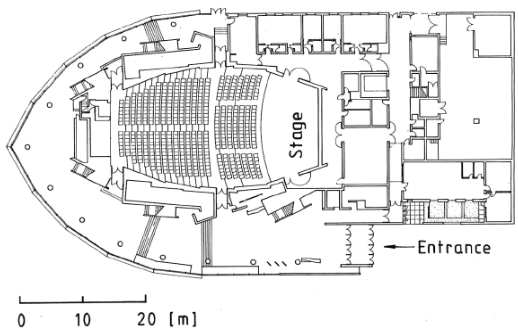


Figure 2.16c

Another example is the Tsuyama Music Cultural Hall (Suzumura, Sakurai, Ando, Yamamoto, Iizuka and Owaki, 2000) with the similar shape of the Kirishima Hall. An additional design, introduced 52 columns (30-cm diameters) as a design element in the hall (PHOTO 2.2). The columns provide surfaces for scattering reflected sound waves for the higher frequency range above 1 kHz. This brings avoid the image shift of sound source on the stage, and about a small value of IACC for the frequency range near the columns and at the seating-center area close to the stage. Since columns weakened the specular reflection from the sidewalls by scattering effects (Fuji, Hotehama, Kato, Shimokura, Okamoto, Suzumura and Ando, 2004), the value of SPL and the A-value are decreased near the sidewalls. And, □t_1 is prolonged throughout the hall.

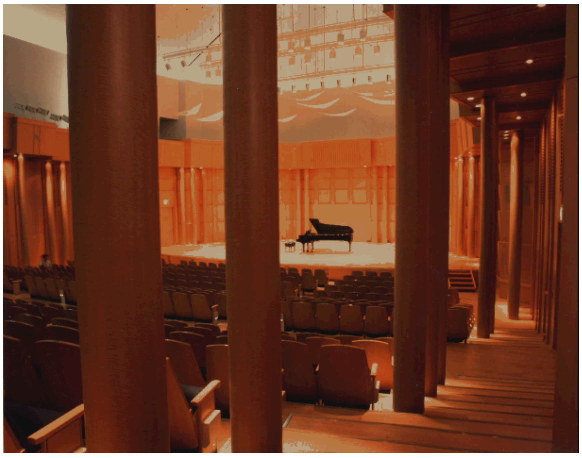


Photo 2.2

As a temporal design, the height of triangular reflectors installed above the stage may be adjusted for performers, according to the effective duration $(\tau_e)_{min}$ of the ACF for music in a program. The acoustic environment inside the hall is well suited to chamber music, with a $(\tau_e)_{min}$ in the range of 50 - 90 ms, because the subsequent reverberation time T_{sub} measured was almost constant about 1.7 s with audience. This canopy plays an important role in decreasing the IACC at the seating audience area close to the stage also (Nakajima, Ando and Fujita, 1992).

2.3 Seat Selection in a Hall Maximizing Individual Preference

In order to maximize the individual subjective preference for each listener, a special facility for seat selection, testing each listener's own subjective preference, was first introduced at the Kirishima International Concert Hall in 1994. The sound simulation is realized based on the system with multiple loudspeakers. The system used arrows for testing subjective preference of sound fields for listeners at the same time. Since the four orthogonal factors of the sound field influence the preference judgments almost independently, as was discussed above, a single orthogonal factor is varied, while the other three are fixed at the most preferred condition for the average listener. Results of testing acousticians who participated in the International Symposium on "Music and Concert Hall Acoustics" (MCHA95), which was held in Kirishima, in May 1995, are presented here (Ando and Noson, 1997).

2.3.1 Individual Subjective Preference

The music source was orchestral, the "Water Music" by Handel; the effective duration of the ACF, τ_e, was 62 ms. The total number of listeners participating for the individual PCT was 106 (Ando and Singh, 1996; Ando, 1998). Typical examples of the test results, as a function of each factor, for listener BL are shown in FIGURE 2.17.

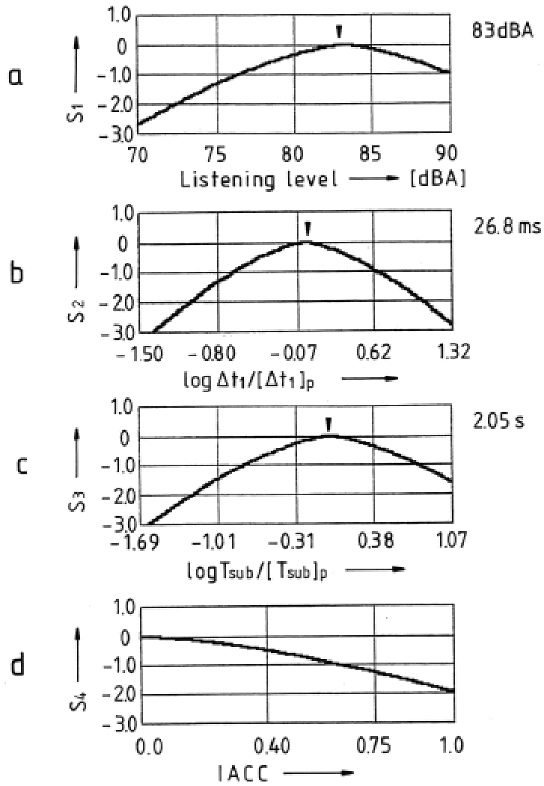


Figure 2.17

The scale value of this listener was close to the average for subjects previously collected: the most preferred $[LL]_p$ is 83 dBA, $[\Delta t_1]_p$ is 26.8 ms (the preferred value calculated by Equation (3.9) was 24.8 ms, where $[\Delta t_1]_p = (1 - \log_{10}A)\,\tau_e$, A = 4), and the most preferred reverberation time was 2.05 s (the preferred value calculated by Equation (3.15) is 1.43 s). Thus, the center area of seats was preferred for listener BL similar to the calculated value at the design stage, as shown in FIGURE 2.18.

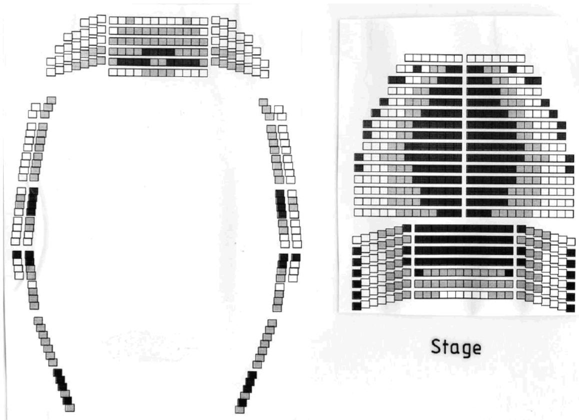


Figure 2.18

With regard to the IACC, it was the result for all listeners that the scale value of preference increased with decreasing IACC value. Since listener KH preferred a very short delay time of the initial reflection, his preferred seats were located close to the boundary walls as shown in FIGURE 2.19.

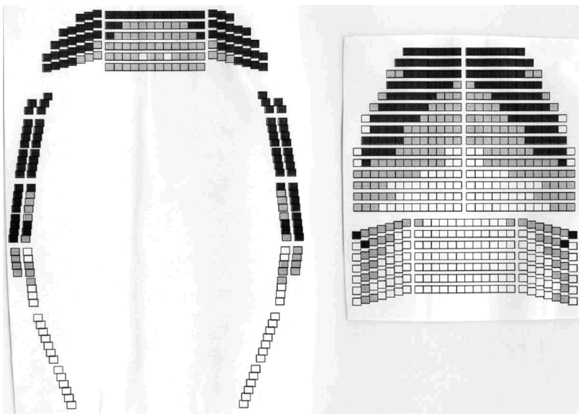


Figure 2.19

Listener KK indicated a preferred listening level exceeding 90 dBA. For this listener, the front seating areas close to the stage were preferable, as shown in FIGURE 2.20.

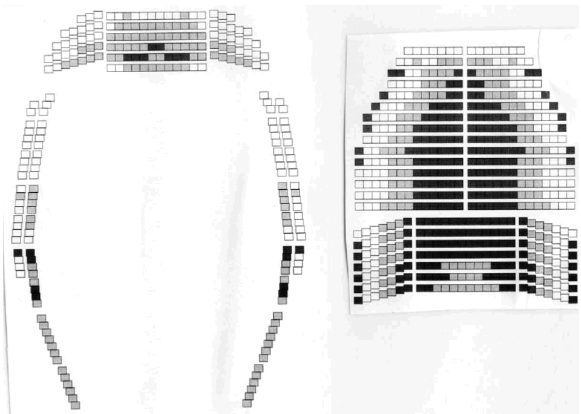


Figure 2.20

On the contrary, for listener DP, whose preferred listening level was a rather weak (76.0 dBA) and preferred initial delay time short (15.0 ms), so that the preferred seats are in the rear part of hall as shown in FIGURE 2.21.

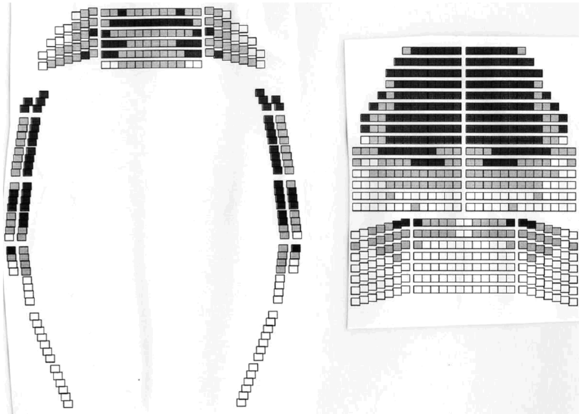


Figure 2.21

The preferred initial time delay gap for listener AC exceeds 100.0 ms, but was not critical. Thus, any initial delay times are acceptable, but the IACC is critical. Therefore, the preferred area of seats was located only in the rear part, as is shown in FIGURE 2.22.

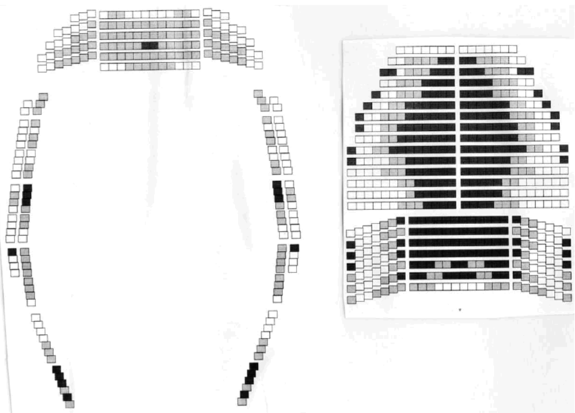


Figure 2.22

2.3.2 Preferred Conditions for Each Individual

Cumulative frequencies of the preferred value with 106 listeners are shown in FIGURE 2.23 through FIGURE 27 for three factors.

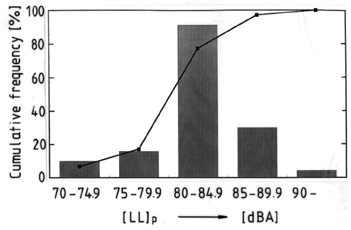


Figure 2.23

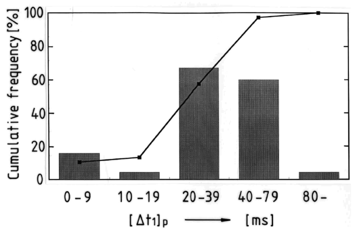


Figure 2.24

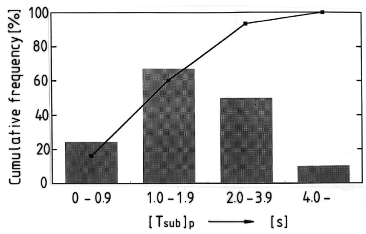


Figure 2.25

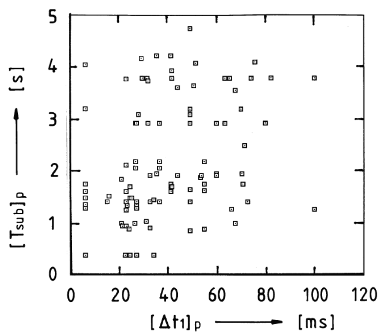


Figure 2.26

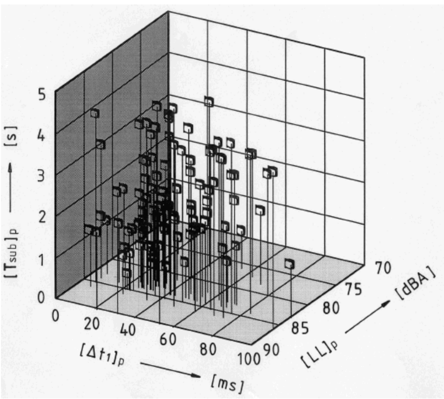


Figure 2.27

As indicated in FIGURE 2.23, about 60 % of listeners preferred the range of 80 to 84.9 dBA in listening to music, but some of listeners indicated that the most preferred LL was above 90 dBA, and the total range of the preferred LL was scattered, exceeding a 20 dB range. As shown in FIGURE 2.24, about 45 % of the listeners preferred the initial delay times 20 to 39 ms, which were around the calculated preference of 24.8 ms (Equation 2.3); some of listeners indicated 0 to 9 ms and others more than 80 ms. With regard to the reverberation time, as shown in FIGURE 2.25, about 45 % of listeners preferred 1.0 to1.9 s which are centered on the calculated preferred value of 1.43 s, but some listeners indicated preferences less than 0.9 s or more than 4.0 s.

It was thought that both the initial delay time and the subsequent reverberation time appear to be related to a kind of "liveness" of the sound field. And, it was assumed that there is a great interference effect on subjective preference between these factors for each individual. However, as shown in FIGURE 2.26, there is little correlation between preference values of $[\Delta t_1]_p$ and $[T_{sub}]_p$ (correlation is 0.06). The same is true for the correlation between values of $[T_{sub}]_p$ and $[LL]_p$ and for that between values of $[LL]_p$ and $[\Delta t_1]_p$, a correlation of less than 0.11. FIGURE 2.27 shows the three-dimensional plots of the preferred values of $[LL]_p$, $[\Delta t_1]_p$ and $[T_{sub}]_p$ excluding the consensus factor of IACC. Looking at a continuous distribution in preferred values, no specific groupings of individuals can be classified to emerge from the data.

In calculation with Equation (2.6), each term S_i is independent, and there is no correlation between weighting coefficients α_i and α_j, $i \neq j$, (i and j = 1, 2, 3, 4) also (Ando, 1998). A listener indicating a relatively small value of one factor will not always indicate a relatively small value for another factor. Thus, a listener can be critical about preferred conditions as a function of certain factors, while insensitive to other factors, resulting in individual characteristics distinct from other listeners. This is an indication of the individual difference.

3. ACF/IACF MODEL OF THE HUMAN AUDITORY SYSTEM

The workable central auditory signal processing model is proposed, which describes both temporal sensations and spatial sensations in relation to the temporal and spatial factors, respectively. Based on the model, the temporal factors associated with the left cerebral hemisphere may be extracted from the ACF processors. And, the spatial factors associated with the left hemisphere may be extracted from the IACF processor for the signal arriving at the two ear entrances. Thus, any subjective attributes may be described by the use of these temporal and the spatial factors as well as the specialization of the human cerebral hemispheres, which represents independent influence on the subjective responses.

3.1 Central Auditory Signal Processing Model

The central auditory signal processing model is based on the following evidences as discussed above.

A. **Physical System**

First, it is interesting to note 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. The transfer function of such a cascade system including the human head and pinna, the external canal and the eardrum and the born chain and excluding a feedback system (Ando, 1985; 1998). For the sake of practical convenience, the A-weighting network may be utilized representing the ear sensitivity.

B. **ABRs from the left and right auditory pathway**

The records of auditory brainstem responses (ABR) as demonstrated in FIGURE 3.1 imply:

1. Amplitudes of wave $I_{l,r}$ & $III_{l,r}$ correspond roughly to the sound pressure level at the two ear entrances as a function of the horizontal angle of incidence to the listener (ξ) as shown in FIGURE 3.2a and 3.2c.

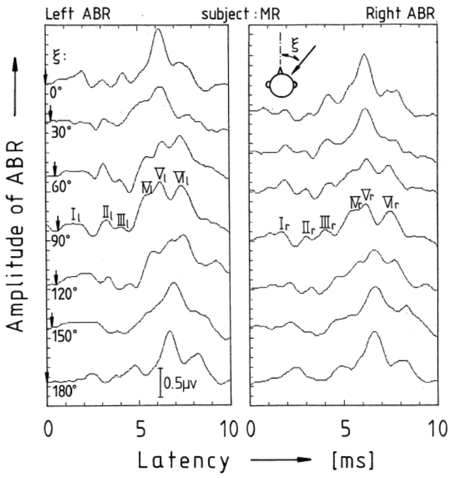


Figure 3.1

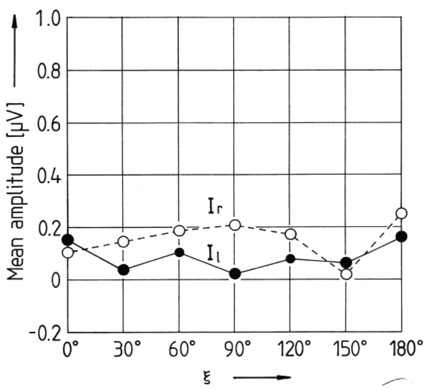


Figure 3.2a

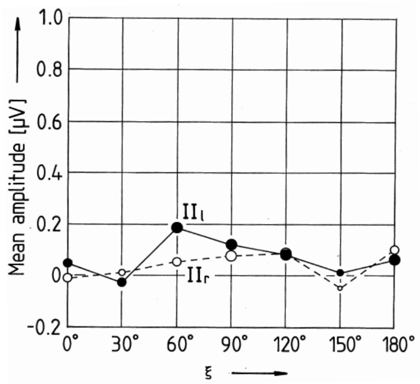


Figure 3.2b

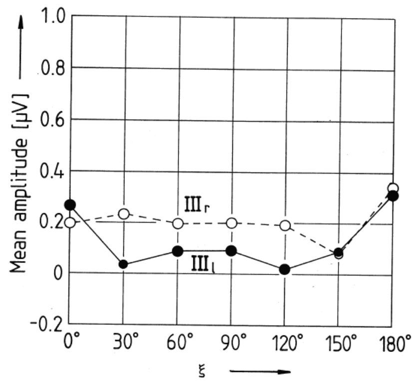


Figure 3.2c

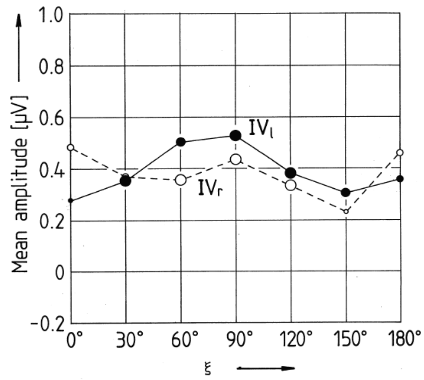


Figure 3.2d

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 (-ξ) as shown in FIGURE 3.2b and 3.2d. This evidence implies three-times interchanges of neural information flow between the left and right auditory pathways (FIGURE 3.3).

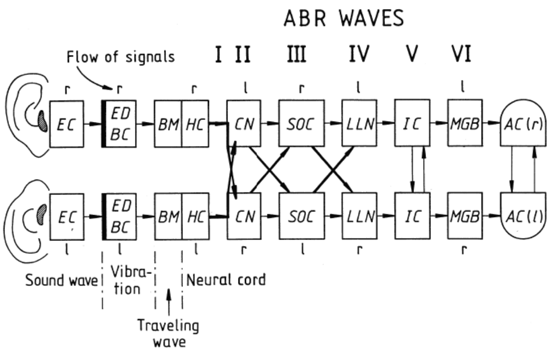


Figure 3.3

Note that latencies III typically correspond to interaural time differeces (FIGURE 3.4).

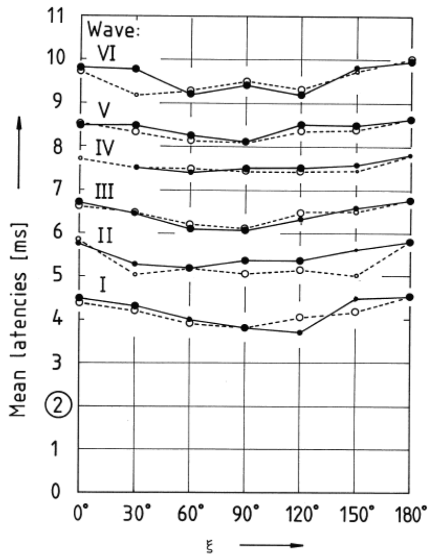


Figure 3.4

3. Results of analysis of the ABRs indicate possible neural activities at the inferior colliculus, which correspond well with the value of IACC (FIGURE 3.5 - 3.7).

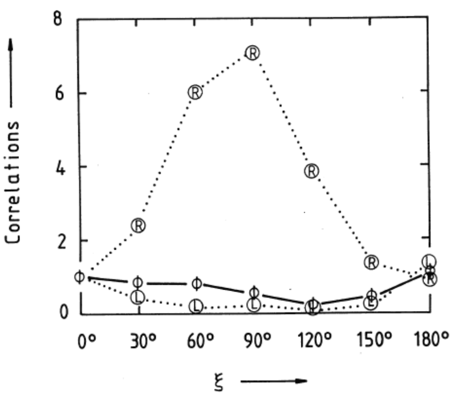


Figure 3.5

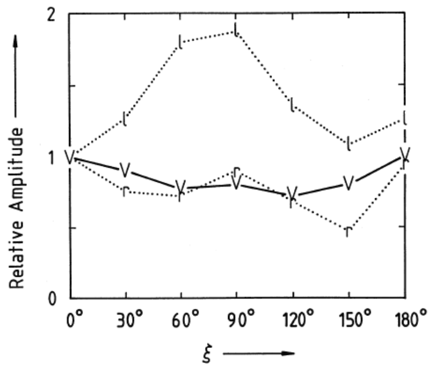


Figure 3.6

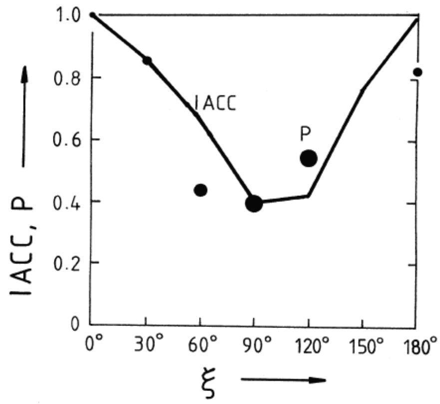


Figure 3.7

C. **SVR on the left and right hemisphere**

Recording the left and right SVR as demonstrated in FIGURE 3.9 has discovered following:

4. The left and right amplitudes of the early SVR, $A(P_1 - N_1)$ indicate that the left and right hemispheric dominance, respectively, are due to the temporal factor: Δt_1 (FIGURE 3.10) and spatial factors: the sensation level, SL (FIGURE 3.11) and IACC (FIGURE 3.12). At first, it was considered that the SL or the LL is classified as a temporal-monaural factor from a physical viewpoint. However, results of the SVR indicate that SL is the right hemisphere dominant. It is quite natural that the LL is classified as a spatial factor, which is measured by the geometric average of sound energies arriving at the two ears.

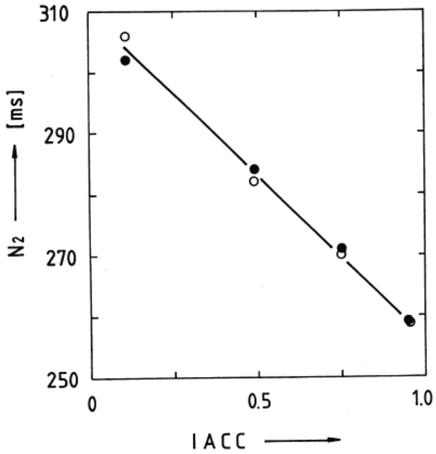


Figure 3.8a

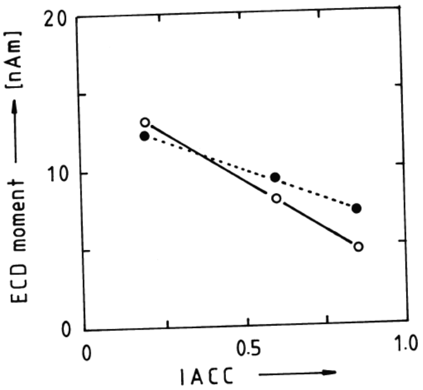


Figure 3.8b

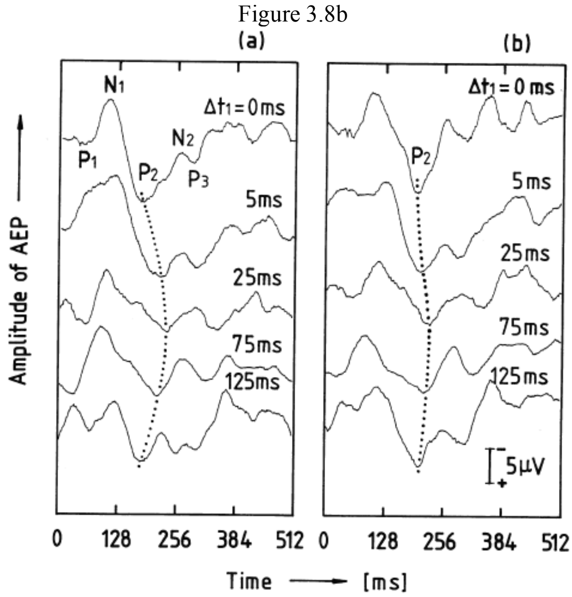


Figure 3.9

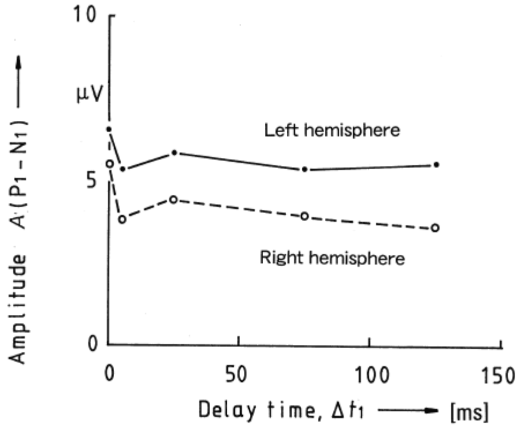


Figure 3.10

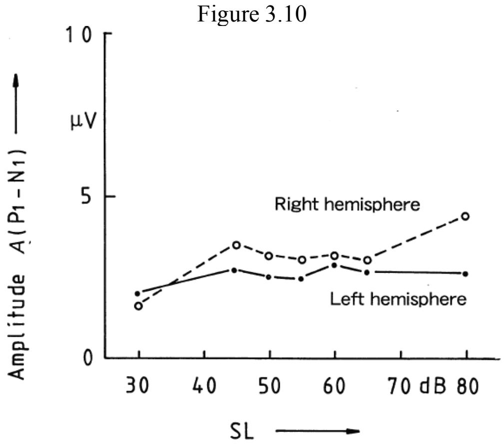


Figure 3.11

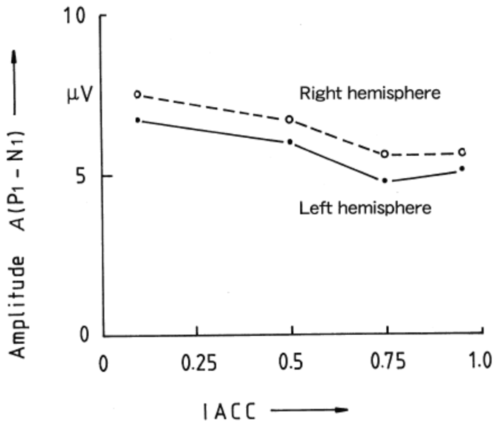


Figure 3.12

5. Both left and right latencies of N_2 correspond well to the IACC (FIGURE 3.8), and thus these are related to the scale value of subjective preference as an overall-primitive response.

D. **EEG on left and right hemispheres**

Analysis of an EEG recorded from the left and right cerebral hemispheres reconfirm that:

6. Δt_1 and T_{sub} indicate left hemisphere dominance (FIGURE 3.13 through 3.16), and the IACC response indicates right hemisphere dominance (FIGURE 3.17 through 3.19). It is considered; therefore, a high

degree of independence between the left and right hemispheric factors was resulted.

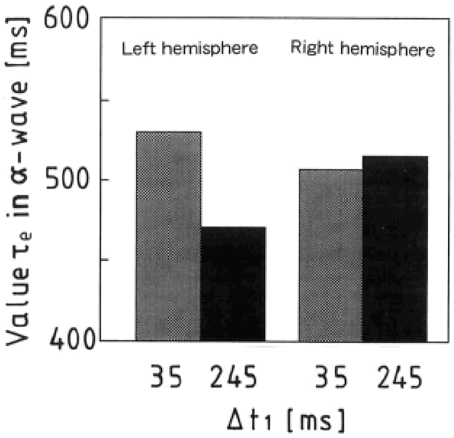


Figure 3.13

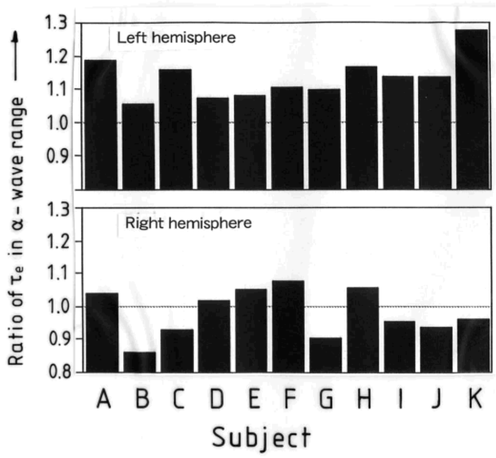


Figure 3.14

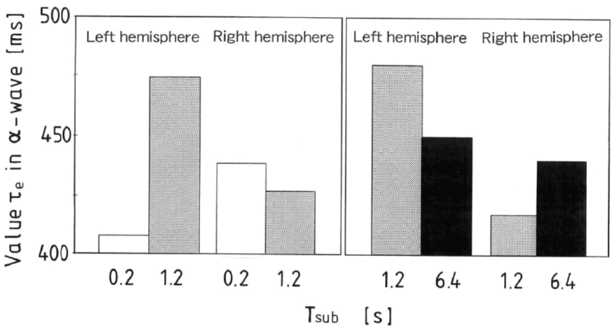


Figure 3.15

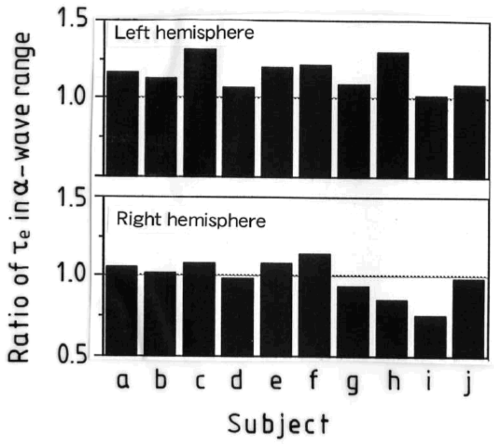


Figure 3.16a

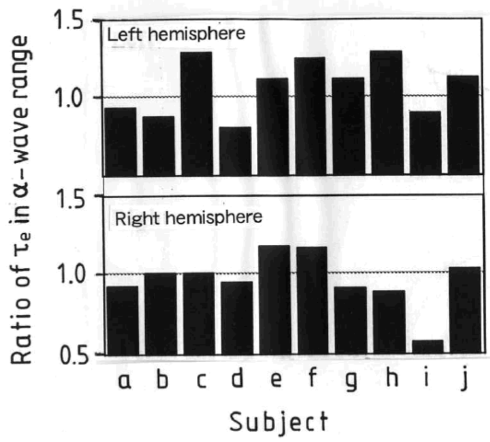


Figure 3.16b

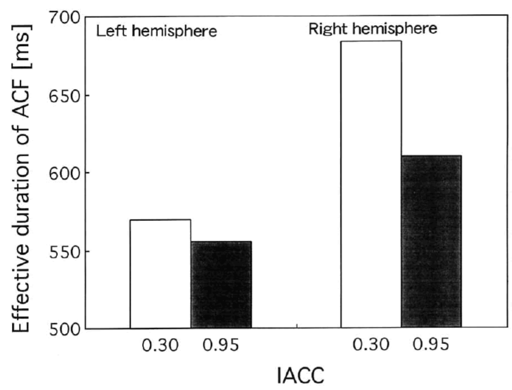


Figure 3.17

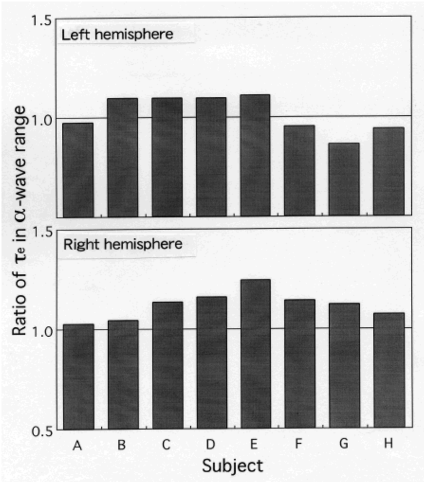


Figure 3.18

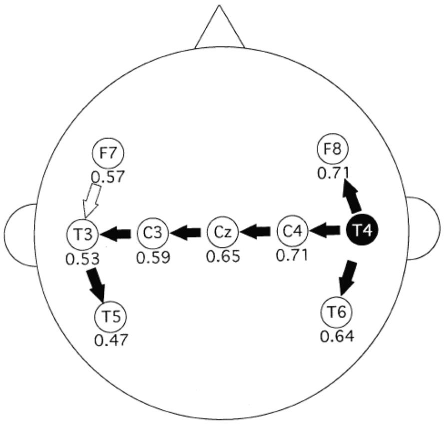


Figure 3.19

7. The scale value of subjective preference corresponds well to the value of τ_e extracted from the ACF of the α-wave, when temporal and spatial factors of the sound field are varied.

E. **MEG on the left and right hemisphere**

In addition, analysis of MEG recorded on left and right cerebral hemispheres reconfirm that:

8. Amplitudes of MEG recorded when Δt_1 is changed show left hemisphere specialization (Soeta, 2002).

9. The scale value of individual subjective preference is directly related to the value of τ_e extracted from the ACF of the α-wave of the MEG in the left hemisphere.

4. MODEL AND ALPHA RHYTHMS CORRESPONDING TO SUBJECTIVE PREFERENCE AND ANNOYANCE

4.1. Model and Specialization of Human Cerebral Hemispheres

Based on the above-mentioned physical system and physiological responses, a central auditory signal processing model may be formed for the comprehensive temporal and spatial factors (Ando, 1985). The model consists of the autocorrelation mechanisms, the interaural cross-correlation mechanism between the two auditory pathways, and the specialization of human cerebral hemispheres for temporal and spatial factors of 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, a model may be proposed as shown in FIGURE 4.1. In this figure, a sound source p(t) is located at r_0 in a three-dimensional space and a listener is sitting at r which is defined by the location of the center of the head, $h_{l,r}(r|\, r_0,t)$ being the impulse responses between r_0 and the left and right ear-canal entrances. The impulse responses of the external ear canal and the bone chain are $e_{l,r}(t)$ and $c_{l,r}(t)$, respectively. The velocity of the basilar membrane is expressed by $V_{l,r}(x, \omega)$, x being the position along the membrane.

The action potentials from the hair cells are conducted and transmitted to the cochlear nuclei, the superior olivary complex including the medial superior olive, the lateral superior olive and the trapezoid body, and to the higher level of the two cerebral hemispheres. The input power density spectrum of the cochlea I(x') can be roughly mapped at a certain nerve position x' (Katsuki, Sumi, Uchiyama and Watanabe, 1958, Kiang, 1965), as a temporal activity. Amplitudes of waves (I - IV) of the ABR reflect to the sound pressure levels at both ears as a function of the horizontal angle of incidence to a listener (Section 3.1.1.). Such neural activities, in turn, include sufficient information to attain the ACF, probably at the lateral lemniscus as indicated by $\Phi_{ll}(\sigma)$ and $\Phi_{rr}(\sigma)$. In fact, the time domain analysis of firing rate from auditory nerve of cat reveals a pattern of ACF rather than the frequency domain analysis (Secker-Walker and Searle, 1990). Pooled interspike interval distributions resemble the short time or the running ACF for the low-frequency component as shown in FIGURE 4.2 (Cariani and Delgutte, 1996a). It traces a change of the missing fundamental or pitch as a function of the time. And, the pooled interval distributions for sound stimuli consisting of the high-frequency component resemble the envelope to running ACF.

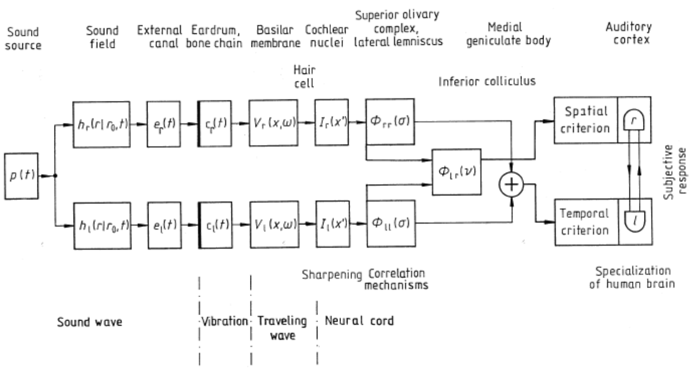


Figure 4.1

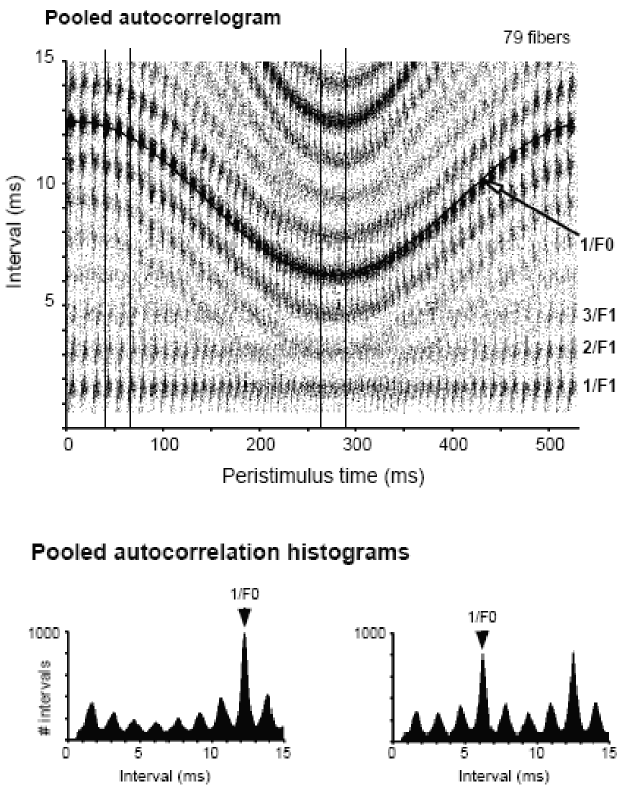


Figure 4.2

From a viewpoint of the missing fundamental or pitch of the complex tone judged by humans, the running ACF must be processed in the frequency components below about 5 kHz (Inoue, Ando and Taguti, 2001). The missing fundamental or pitch may be perceived less than about 1.2 kHz, which may cover most of musical signals. A tentative model of the running ACF processor is illustrated in FIGURE 4.3. The output of the ACF processor may be dominantly connected with the left cerebral hemisphere.

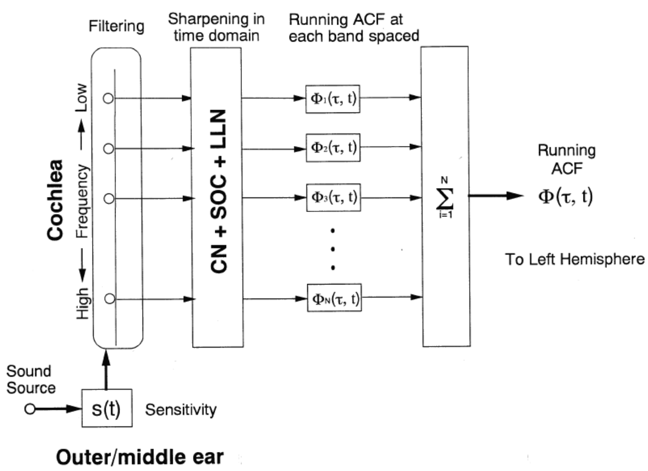


Figure 4.3

As is also discussed, the neural activity (wave V together with waves IV_l and IV_r) may correspond to the IACC as shown in FIGURE 3.7. Thus, the interaural cross-correlation mechanism may exist at the inferior colliculus. It is concluded that the output signal of the interaural cross-correlation mechanism including the IACC may be dominantly connected to the right hemisphere. Also, the binaural sound pressure level expressed by a geometrical average of the ACFs for the two ears at the origin of time ($\sigma = 0$) and in fact appears in the latency at the inferior colliculus, may be processed in the right hemisphere. The neural process has been developed realizing a minimum of effort and a maximum of efficiency, so that only information of criteria extracted from the ACF and IACF are transmitted into the left and right hemispheres, respectively.

It is concluded that the listening level (LL) and the IACC are associated with the right cerebral hemisphere, and the temporal factors, $\Delta t1$ and Tsub, the sound field in a room are associated with the left (Table 4.1). 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" may well be explained by such specialization of the human brain, because speech is processed in the left hemisphere, and independently the spatial information is mainly processed in the right hemisphere.

Based on the model, we well describe temporal and spatial sensations, in turn, any subjective attributes of the sound fields in term of processes in the auditory pathways and the specialization of two cerebral hemispheres.

Factors changed	AEP (SVR) $A(P_1 - N_1)$	EEG, ratio of ACF τ_e values of α-wave	AEP (MEG) N1m	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 (band noise)	R > L (music) [2]	R > L (band noise) [3]	
τ_{IACC}			R > L (band noise) [3]	
Head related transfer functions			R > L (vowels) [4]	

TABLE 4.1 Hemispheric specializations of temporal and spatial factors determined by analyses of the auditory evoked potentials (AEP), EEG and MEG.

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

[2] Flow of EEG α-wave from the right hemisphere to the left hemisphere for music stimulus 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.

[4] Palomaki, Tiitinen, Makinen, May and Alku, 2002.

4.2 MEG Alpha Wave Corresponding to Subjective Preference and Annoyance

In this section, we demonstrate the effective duration of autocorrelation function of alpha wave in MEG corresponding well to individual subjective preference (7-8 subjects).

4.2.1. Brain Response Corresponding to Subjective Preference

Measurements of responses on MEG were performed in a magnetically shielded room using a 122-channel whole-head neuromagnetometer (Neuromag-122™, Neuromag Ltd., Finland, see PHOTO 4.1) (Soeta, Nakagawa, Tonoike and Ando, 2002).

Photo 4.1

The source signal was the word, "piano," with a 0.35 s duration (FIGURE 4.4(a)). The minimum value of the moving τ_e, i.e., $(\tau_e)_{min}$, was about 20 ms as shown in FIGURE 4.4b. It is worth noticing that this value is close to the most preferred delay time of the first reflection of sound fields with continuous speech (Ando and Kageyama, 1977; FIGURE 2.1(b)). In the present experiment, the delay time of the single reflection (Δt_1) was set at five levels (0, 5, 20, 60, and 100 ms). The direct sound and a single reflection were mixed and the amplitude of the reflection was the same as that of the direct sound ($A_0 = A_1 = 1$). The auditory stimuli were binaurally delivered through plastic tubes and earpieces into the ear canals. The sound-pressure level, which was measured at the end of the tubes, was fixed at 70 dBA.

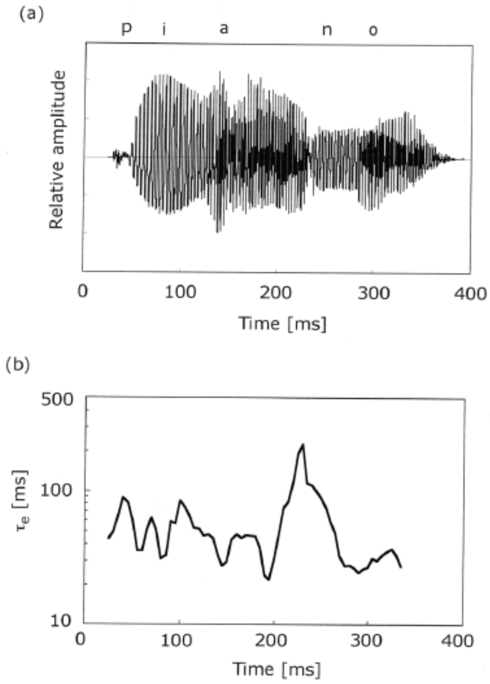


Figure 4.4a/b

Eight, 23to 25-year-old subjects participated in the experiment. All had normal hearing. In accordance with the PCT, each subject compared ten possible pairs per session, and a total of ten sessions were conducted for each subject. Measurements of magnetic responses were performed in a magnetically shielded room. The paired-auditory stimuli were presented in the same way as in the subjective preference test. During measurements, the subjects sat in a chair with their eyes closed. To compare the results of the MEG measurements with the scale values of the subjective preference, combinations of a reference stimulus ($\Delta t_1 = 0$ ms) and test stimuli ($\Delta t_1 = 0$, 5, 20, 60, and 100 ms) were presented alternately 50 times, and the MEG's were analyzed. The magnetic data was recorded continuously with a filter of 0.1-30.0 Hz and digitized with a sampling rate of 100 Hz. FIGURE 4.5 shows an example of recorded MEG alpha waves. Eight channels, that had larger amplitude of N1m response in each hemisphere, were selected for the ACF analysis. We analyzed the MEG alpha wave for each of the paired stimuli, for each subject. FIGURE 4.6 show examples of the value of τ_e obtained by the straight line for 5 dB from the top of the normalized ACF expressed in logarithm. Obviously, for the preferred condition at $\Delta t_1 = 5$ ms of the sound field the value of $\tau_e \approx 0.5$ s, and for the condition of echo disturbance ($\Delta t_1 = 100$ ms), the value of $\tau_e \approx 0.3$ s is much smaller than that.

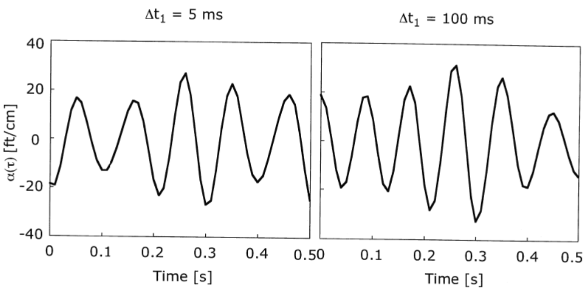


Figure 4.5

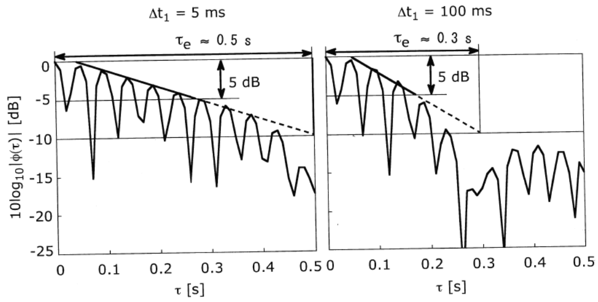
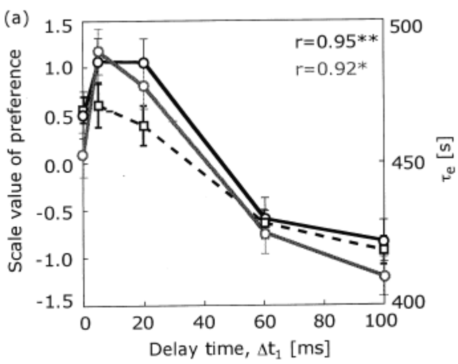


Figure 4.6

The results from the eight subjects confirm a linear relationship between the averaged τ_e values of alpha wave and the averaged scale values of subjective preference. Their correlation coefficients were 0.95 ($p < 0.01$) in the left hemisphere, and 0.92 ($p < 0.05$) in the right hemisphere (FIGURE 4.7a). Since the left hemisphere dominates Δt_1, reconforming the aforementioned studies of SVR and EEG, the results of the individual level from the left hemisphere are analyzed. An almost direct relationship between individual scale values of subjective preference and the τ_e values over the left hemisphere as found in each of the eight subjects. Results for each of the eight subjects are shown in FIGURE 4.8. Remarkably, the correlation coefficient, r, was achieved more than 0.94 for all subjects. However, as shown in FIGURE 4.7b, there is little relationship between the scale values of subjective preference and the amplitude of α wave, Φ (0), in both hemispheres ($r < 0.37$).



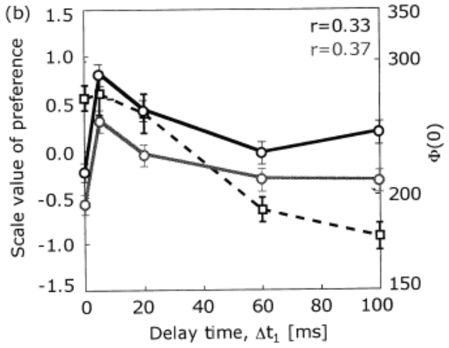


Figure 4.7

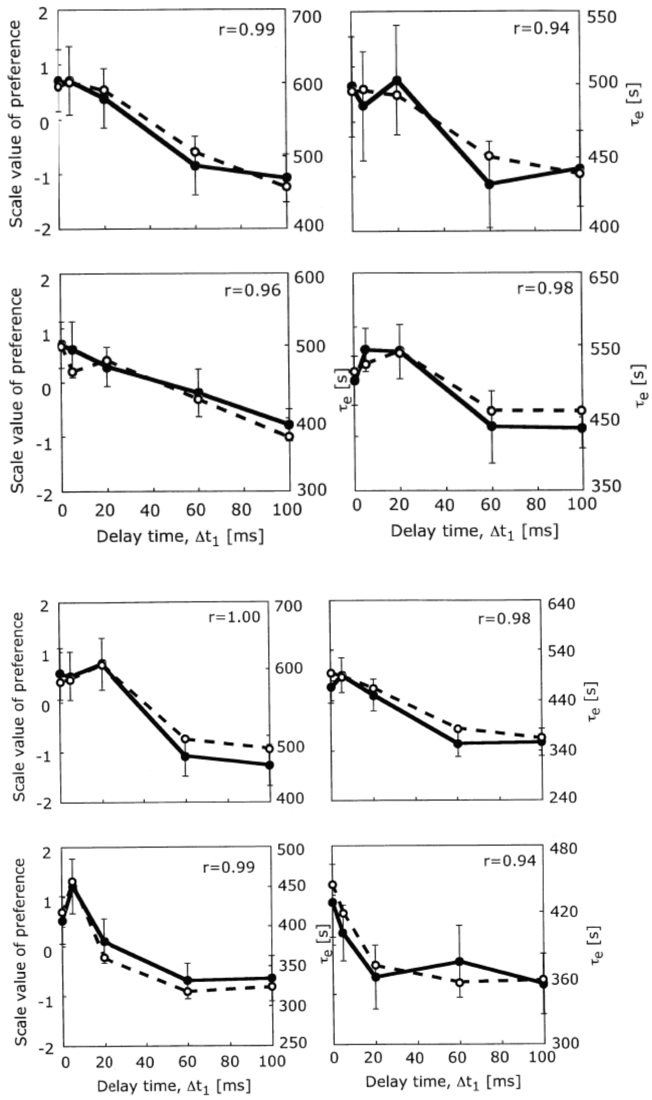


Figure 4.8

The value of τ_e is the degree of similar repetitive features included in alpha waves, so that the brain repeats a similar rhythm under the preferred conditions. This tendency for a larger τ_e under the preferred condition is much more significant than the results on EEG alpha waves as mentioned above.

4.2.2. Cross-Correlation Function (CCF) Analysis of the Alpha Wave Over the Scalp

The magnetic responses were analyzed by a cross-correlation function (CCF) between 36 reference channels and 35 test channels. In MEG measurements using speech (the word, “piano”) as the source signal, combinations of a reference stimulus ($\Delta t_l = 0$ ms) and test stimuli ($\Delta t_l = 0$, 5, 20, 60, and 100 ms) were presented 50 times alternately at a constant 1-s interstimulus interval (Soeta, Nakagawa, Tonoike and Ando, 2003). Eight, 23 to 25-year-old subjects participated in the experiment. The scale value of the subjective preference of each subject was obtained by the PCT also. Results showed that: 1. The maximum amplitude of the CCF, $|\Delta (t)|_{max}$, between alpha waves (8-13 Hz) recorded at the two different channels is increased when increasing the scale value of subjective preference. 2. The maximum amplitude of the CCF decreases when increasing the channel distance. These imply that the brain repeats a similar temporal rhythm in the alpha-wave range over such a wider area of the scalp under the preferred sound field.

Also it has been reconfirmed by experiments on MEG with the same speech signal in change of the IACC (0.27, 0.61 and 0.90) that the values of τ_e and the maximum amplitude of the CCF were increased when deceasing the IACC (Soeta, Nakagawa and Tonoike, 2005).

4.2.3. Brain Response Corresponding to Annoyance

In a manner similar to the previous section, measurements of responses on MEG were performed by use of a pure tone and band-pass noise with the center frequency of 1000 Hz as the source signals for judgment of individual annoyance (Soeta, Nakagawa, Tonoike and Ando, 2004). In order to control the ACF of the source signal, the bandwidth of the noise centered on 1000 Hz was varied at five levels (0, 40, 80, 160 and 320 Hz) by a sharp filter with the slope of 2000 dB/Oct. (Sato, Kitamura and Ando, 2002). Such a sharp filter was applied here due to the sharpening effect observed at the inferior colliculus in the auditory pathway particularly for the higher frequency band than 1000 Hz (Katsuki, Sumi, Uchiyamada and Watanabe, 1958). It is worth noticing that a filter, with its slope of about 60 dB/octave, is too small to apply for any acoustic measurement. The filter “0 Hz” means an actual filter, which is setting cutoff frequencies of high-pass filter and low-pass filter at 1000 Hz, so that it consists of only the slope component. The sound pressure level of all source signals was fixed at 74 dBA by the measurement of the ACF, Φ (0). The source signals were characterized by the ACF temporal factors, τ_1, ϕ_1 and τ_e. The dependent factors, ϕ_1 and τ_e, may be controlled by the bandwidth of the source signal. Seven 22-to 28-year-old subjects participated in this experiment. The PCT was performed for all combination of 15 pairs as one session. A total of ten sessions were conducted for each subject. The signal duration was 2 s with a rise and fall times of 10 ms. They were asked to judge which of the two sound signals was more annoying. The scale value of annoyance for each subject was obtained (Ando, 1998).

The same subjects, who participated in the annoyance tests, also participated in recording the MEG response and paired-stimuli, which were presented in the same way as in the annoyance tests. Combinations of the reference pure-tone stimulus and a test noise stimulus were presented alternatively 30 times to be integrated and recorded in MEG responses. Eighteen channels located around the temporal area in each hemisphere were selected for the ACF and CCF analysis of alpha wave 8 – 13 Hz. Examples of recorded MEG are shown in FIGURE 4.9.

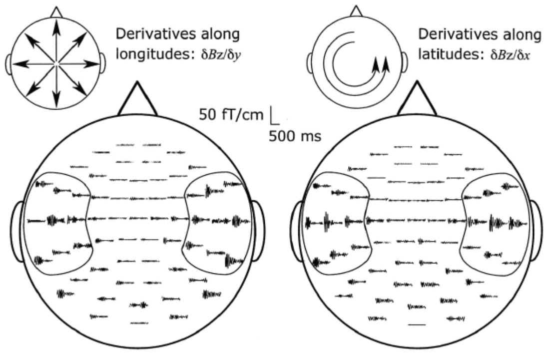


Figure 4.9

A two-way ANOVA showed significant effects of Stimulus ($p < 0.01$) and Subject ($p < 0.01$). Thus, we first discuss the effects of stimulus and individual difference. A remarkable finding is that the difference of scale values of annoyance and the ratio of τ_e value are inversely correlated for each subject ($r = -0.83$), as shown in FIGURE 4.10. This tendency of annoyance is just opposite of subjective preference, in which the τ_e value was increased with increasing scale values of subjective preference as is discussed in previous sections. Thus, the value of τ_e was shorter during the presentation of annoying stimuli. Also, it is found that the magnitude of CCF, $|\phi(\tau)|_{max}$, was decreased with increasing annoyance. Thus, the brain is not relaxed either temporally or spatially under presentation of an annoying stimulus. Previous studies on EEG and MEG show that the value of τ_e increased significantly with increasing scale value of preference. This signifies the brain is repeating a similar rhythm over a wider area under the preferred conditions. It is remarkable that the sites that signify the highest correlation between the scale values of annoyance and the values of τ_e were observed in the right hemisphere for all of subjects who participated.

This implies right hemisphere dominance against noise or non-verbal stimuli (Chon, 1970).

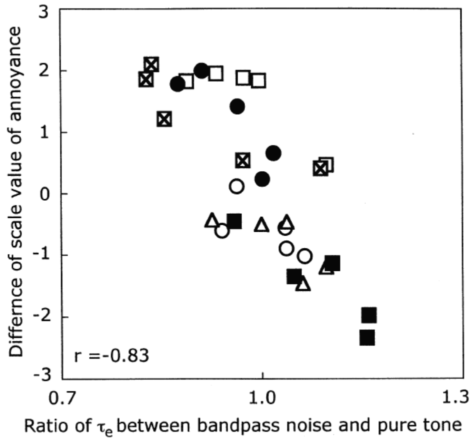


Figure 4.10

Although there is such a deep relationship between annoyance and the value of τ_e of the MEG-alpha wave in each subject, annoyance itself differs between individuals. As shown in FIGURE 4.11, large individual differences in scale values of annoyance resulted as a function of bandwidth within the critical band.

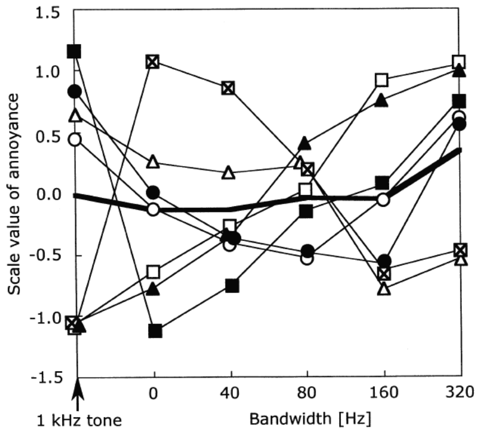


Figure 4.11

It is worth noticing that, results of evoked magnetic response shows that the N1m amplitude corresponds well to the loudness value in the frequency range between 250 and 2000 Hz (Soeta, Nakagawa and Matsuoka, 2006).

5. PRIMARY TEMPORAL SENSATIONS

It shall show that temporal and spatial sensations may be described by the ACF temporal factors and the IACF spatial factors, respectively. First of all, the temporal window analyzing the running ACF is discussed.

5.1 Temporal Window for the ACF Processing

In analysis of the running ACF, the so-called the "auditory-temporal window" 2T in must be carefully determined. The initial part of the ACF within the effective duration τ_e of the ACF contains important information of the signal. In order to determine the auditory-temporal window, successive loudness judgments in pursuit of the running LL have been conducted (Mouri, Akiyama and Ando, 2001). Results are shown in FIGURE 5.1, so that the recommended signal duration $(2T)_r$ to be analyzed is approximately given by,

$$(2T)_r \approx 30(\tau_e)_{min} \tag{5.1}$$

where $(\tau_e)_{min}$ is the minimum value of τ_e obtained by analyzing the running ACF. This signifies an adaptive temporal window depending on the temporal characteristics of the sound signal in the auditory system. Therefore, the temporal window differs according to the music pieces [$(2T)_r$ = 0.5 – 5 s] and to the vowels [$(2T)_r$ = 50 – 100 ms], and consonants [$(2T)_r$ = 5 – 10 ms] in the continuous speech signals. For example, the time constant represented by "fast" or "slow" in the present system of sound level meter should be replaced by this temporal window. The running step (R_s), which signifies a degree of overlap of the signal to be analyzed, is not so critical. It may be selected as $K_2(2T)_r$, K_2 being chosen, in the range of 1/4 – 1/2.

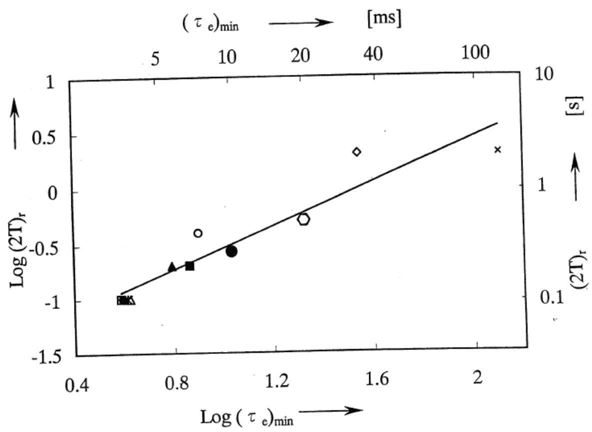


Figure 5.1

The factors extracted from ACF of a sound signal are defined in FIGURE 5.2 and FIGURE 5.3. Following all of four temporal primary sensations may be described by these temporal factors. Let us look experimental results.

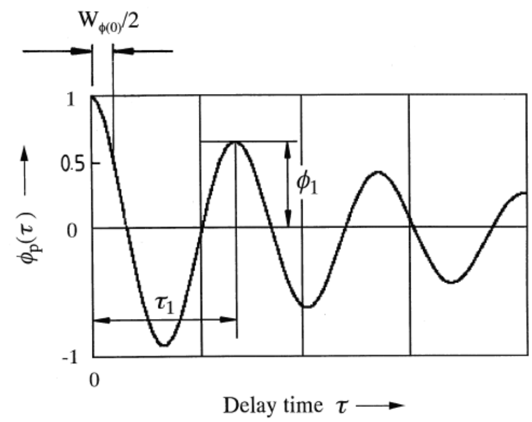


Figure 5.2

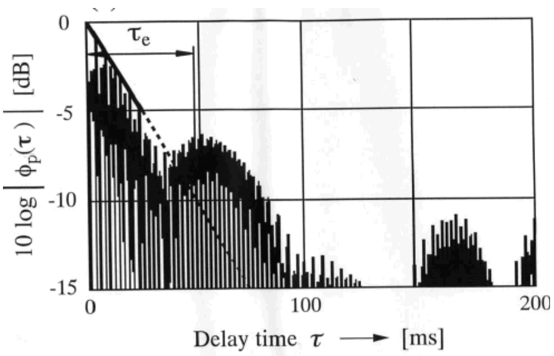


Figure 5.3

5.2 Pitch (Missing Fundamental)

Most of the sounds in tonal music that constitute the notes of melodies and harmonies are harmonically complex tones rather than pure tones. A harmonic complex tone consists of a series of partials whose frequencies (f_1, f_2, f_3... f_m) are integer multiples (n = 1, 2, 3... m) of its fundamental frequency (F_0). Such harmonic complexes produce the strongest pitches at their fundamentals, so long as these periodicities lie in the existence region of musical tonality (roughly 30-5000 Hz). Other, weaker pitches can also be heard that correspond to individual partials, especially the first five (harmonic number n < 6). 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) yet still can produce strong "low" pitch 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. In order to save the notion of the auditory system as a general Fourier processor, it becomes necessary to postulate a complicated central harmonic analyzer.

As a result of these difficulties, some auditory theorists (Seebeck, 1844; Wever, 1949; Licklider, 1951: Rose, 1980) have instead sought temporal explanations for pitch, pointing to the elegance with which time-domain representations cope with the phenomenon of the missing fundamental. In the ACF, the positions of major peaks, which reflect the fundamental, are unchanged. Temporal theories have the advantage of explaining pitch perception of both low-frequency pure tones and complex tones in terms of the same central representations and mechanisms. They account for those pitch phenomena most important for music and speech, i.e., for periodicities between 30 and 4000 Hz. These explanations notwithstanding, it is also clear that temporal representations cannot account for high frequency hearing, and the (atonal) pitches evoked by pure tones with frequencies above ~5000 Hz. Moreover, most auditory centers throughout the pathway have spatially ordered frequency maps that mimic the rough tonotopic organization of the cochlea.

For these reasons, 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).

The first autocorrelation model developed to account for the pitch of the missing fundamental phenomenon was therefore originally formulated as a ‘duplex’ model (Licklider, 1951). Licklider's time-delay neural network architecture was similar in many respects to the Jeffress (1948) model of binaural cross-correlation that had been proposed three years earlier. Licklider used a network of delay lines and coincidence counters arranged along the axes of frequency and delay to compute both a central spectrum and a central global temporal autocorrelation representation. Cherry and Sayers (1956) have emphasized the usefulness of cross-correlation in degree of aural fusion including directional hearing.

Following a series of turns in the evolution of pitch theory (de Boer, 1976 for a historical review), temporal models were neglected in favor of spectral pattern approaches. In the wake of the difficulties with Schouten's temporal theory, spectral pattern recognition models were proposed to explain the strong low pitches produced by low, perceptually resolved harmonics (Wightman, 1973a; Goldstein. 1973; Terhardt, 1974). Two mechanisms were assumed, a spectral pattern mechanism for strong pitches of perceptually resolved low harmonics, and a temporal mechanism for weak pitches of perceptually unresolved high harmonics. Since the best models for low-frequency pure tone pitch discrimination use interspike interval information, some theorists (Goldstein, 1973) left open the possibility that central representations of frequency might be based on interspike interval information in early auditory stations. Explicit temporal representations were thus marginalized to pitches produced by unresolved harmonics; phenomena that are largely irrelevant for pitch in music and speech.

Beginning in the 1980's, temporal models for pitch that were based on first-order interspike intervals (times between successive spikes produced by a given neuron) in the auditory nerve were proposed (Moore, 2003; van Norden, 1982). In these models interspike interval information was pooled together from all regions of the auditory nerve to form a temporal population code for frequency and periodicity. By the end of the decade, temporal autocorrelation models for pitch were revived and tested using computer simulations of the cochlea and auditory nerve (Meddis and Hewitt, 1991a; 1991b). These autocorrelation models are based on all-order interspike intervals (times between all spikes produced by a neuron, consecutive and nonconsecutive) rather than first-order intervals. Soon after, neurophysiologic studies of temporal discharge patterns in the cat auditory nerve (Cariani and Delgutte, 1996a, b, Cariani, 1999, 2001) were conducted to test the temporal models. Taken together the computer simulations and neurophysiologic studies showed that the temporal autocorrelation models based on interspike interval distributions could predict a very wide range of pitch phenomena: pitch of the missing fundamental, pitch equivalence between pure and complex tones, level- and phase- invariance, pitch shift of inharmonic complex tones, pitch dominance, octave similarity, and the non-spectral pitch of amplitude modulated noise.

Analogous phenomena have also been observed for non-periodic, inharmonic complex tones as well as non-stationary sounds (noises). It is important to note that more advanced temporal models go well beyond autocorrelation operations on the stimulus itself to include cochlear filtering and neuronal dynamics. Another line of research in temporal models for pitch has focused on the role of cochlear filtering on the temporal structure of the resulting signals. These studies (Yost, Hill and Perez-Falcon, 1978; Yost, 1996a; 1996b) used rippled noise stimuli to probe pitch strength, peripheral weighting, and the effects of the dominance region for pitch (Ritsuma, 1967). Time-domain cancellation models involving an array of delay lines and inhibitory gating neurons have also been proposed, and these generally behave in a manner similar to those based on autocorrelation (Cheveigne, 1998) (reference de Cheveigne's new pitch review).

Here we propose a model for pitch that is based on a central autocorrelation function (ACF). Pitch can be calculated by the value of the first major lag τ extracted from the ACF, while pitch strength relies on the 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. The main purpose of the present experiment is to find the frequency range applicable to pitch identification by the ACF model. First, a pitch-matching test, comparing 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 FIGURE 5.4. As test signals, the two waveforms of complex tones, (a) in-phases and (b) random-phases, were applied as shown in FIGURE 5.5. 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 FIGURE 5.6, the normalized ACF (NACF) of these stimuli were calculated at the integrated interval 2T = 0.8 s. Though the waveforms differ greatly from each other, as shown in FIGURE 5.5, their NACF are identical. The time delay at the first maximum peak of the NACF, τ_1 equals to 5 ms (200 Hz) corresponding to the fundamental frequency. Five, 20-to 26-years old musicians participated as subjects in the experiment. 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.

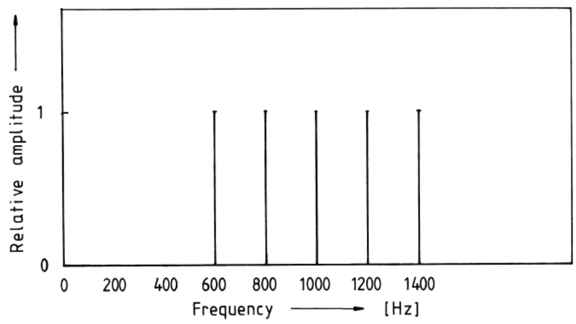


Figure 5.4

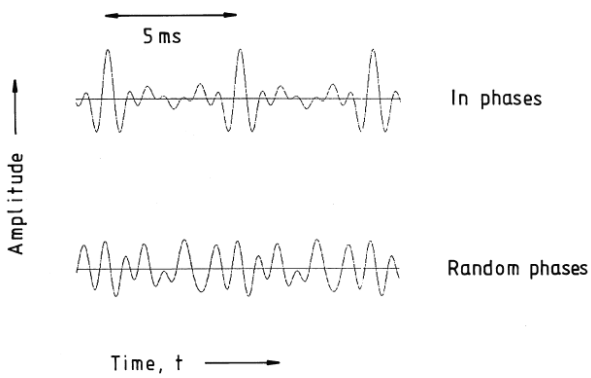


Figure 5.5

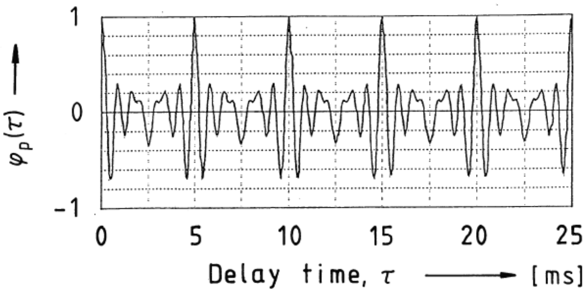


Figure 5.6

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 FIGURE 5.7. 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 NACF. For both in-phase and random-phase conditions, about 60% of the responses clustered within a semitone of the fundamental. There are no fundamental differences in the distributions of pitch-matching data between the two conditions. 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 NACF, τ_1.This result reconfirmed those obtained by Yost (1996a) who demonstrated that pitch perception of iterated rippled noise is greatly affected by the first major ACF peak of the stimulus signal. Pitch as one of temporal sensations may be expressed by

$$S = S_L = f_L(\tau_1) \approx 1/\tau_1 \text{ [Hz]} \qquad (5.2)$$

When $\phi_1 = 1$.

This equation holds for the pitch frequency range below about 1.2 kHz (Inoue et al. 2001).

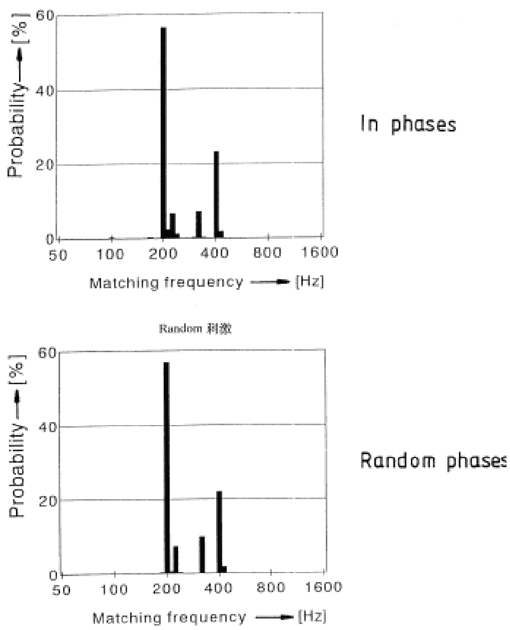


Figure 5.7

5.3 Loudness

This study examines correspondences between the perceived loudness of band-pass noise and properties of the ACF for center frequencies of 250, 500 and 1000 Hz. The bandwidth of the source signal was controlled using a 2068 dB/octave sharp filter that parametrically altered the ACF of the filtered, source signal. The scale value of loudness was obtained using the paired-comparison method. Results show that loudness of a pure tone is greater than that of sharply filtered noises. Loudness of band pass noise increases with increasing effective duration of the ACF (τ_e) of the source signal, which reflects the degree of repetitive structure in the signal. Thus, the loudness of band pass noise inside the critical band is not constant.

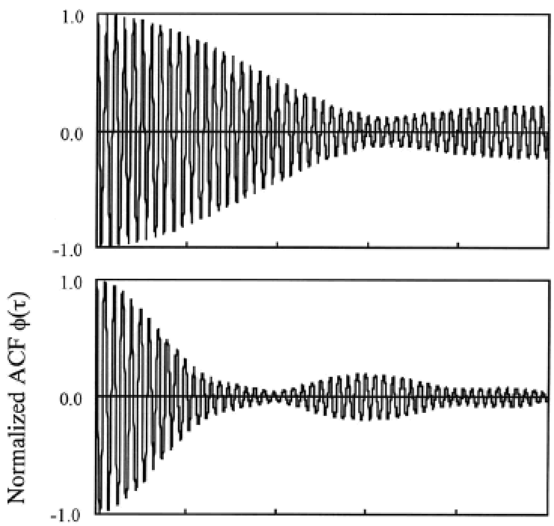
Previous studies on the relationship between loudness and the bandwidth of noise have concluded that for sounds having the same sound pressure level, loudness remains constant as bandwidth increases, up until the bandwidth reaches the "critical band". For bandwidths larger than the critical band, loudness increases with bandwidth (Zwicker, Flottorp and Stevens, 1957). The spectral characteristics of the filters used in those studies were not specified, except by Greenwood (1961a, 1961b). Mathews and Pfafflin (1965) suggested that loudness of band-pass noises might differ between that using an actual filter, and that using an ideal (rectangular shape) filter. An actual filter passes not only frequencies within the band defined by the –3 dB attenuation at the low and high cut-off frequencies, but also at frequencies outside the band. The outside bandwidth response of the filter greatly affects the repetitive feature of the signal, represented by τ_e extracted from the ACF processed in the auditory system (FIGURE 4.1: Ando, 1998; Ando, Sato and Sakai, 1999). Due to the sharpening effect at least in the high frequency range, such a sharp filter may exist in the auditory system (Katsuki, Sumi, Uchiyama and Watanabe, 1958), a roll-off of more than 1000 dB/octave is required. It is considered that loudness of a sharply (1080 dB/octave) filtered noise increases as the effective duration of the normalized ACF (τ_e) increases, even if the bandwidth of the signal is within the critical band. It is worth noticing that tendency was observed that the subsequent reverberation time (T_{sub}) of a sound field increases, the τ_e also increases (Ando, 1998).

The purpose of this study was to examine the loudness of the band-pass noise in terms of factors extracted from the ACF. It is assumed that when the sound pressure level is fixed at a constant value, the scale value of loudness S is expressed by,

$$S = S_L = f_L(\tau_1, \phi_1, \tau_e, D) \qquad (5.3)$$

where the factors are defined in FIGURE 5.2 and FIGURE 5.3, $W_{\phi(0)}$ is excluded in above equation because of the center frequency of the noise is fixed and is represented by τ_1, and D is the duration of the sound signal. As is well known, the sampling frequency of the sound wave should be more than the twice of the maximum audio frequency. Thus, the value $10\log\Phi(0)/\Phi(0)_{ref}$ is far more accurate than any factor based on the envelope of the waveform, $\Phi(0)_{ref}$ being the reference. The difference between them is prominent for an impulsive sound. It is worth noting that loudness does not depend on the IACC under conditions in which the sound pressure level at both ear entrances is fixed. This confirms the results obtained using headphone reproduction (Chernyak and Dubrovsky, 1968; Dubrovskii and Chernyak, 1969).

How do signal and filter parameters affect the shape of the ACF and the features derived from it? A random generator produced the white noise and then filtered it. The source signal of band pass noises is characterized in terms of their ACF as shown in FIGURE 5.8. Bandwidth (Δf) was changed by using a sharp filter with the cut-off slope of 2068 dB/octave, which was realized by a combination of two filters. Factors of τ_1, τ_e and ϕ_1 analyzed are shown in FIGURE 5.9. In fact, the filter bandwidth of 0 Hz included 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, Φ (0). As one can readily see, while filter bandwidth has absolutely no effect on the signal's dominant periodicity (as reflected by τ_1), it has a profound effect on the slope of the ACF envelope (effective duration, as reflected by τ_e) and a lesser effect on the relative height of the peak associated with the dominant periodicity ϕ_1.



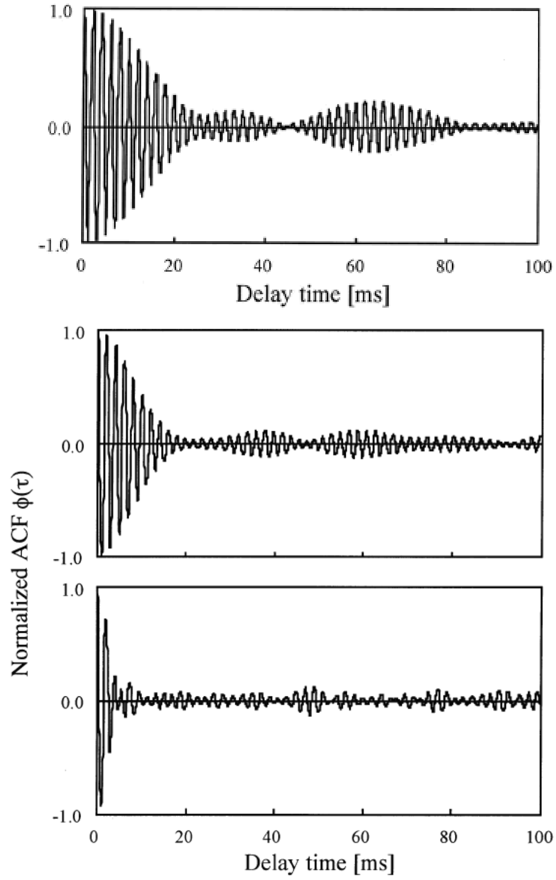


Figure 5.8

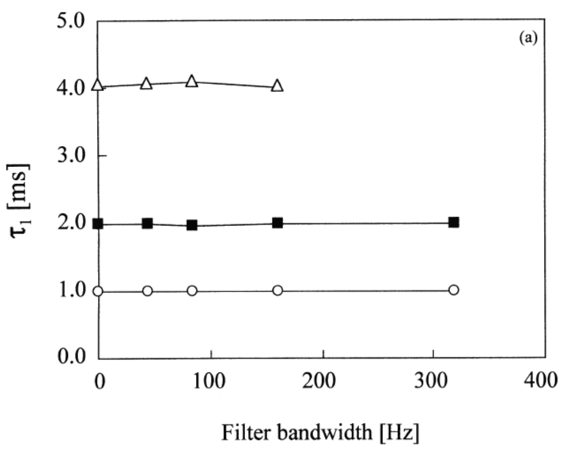


Figure 5.9a

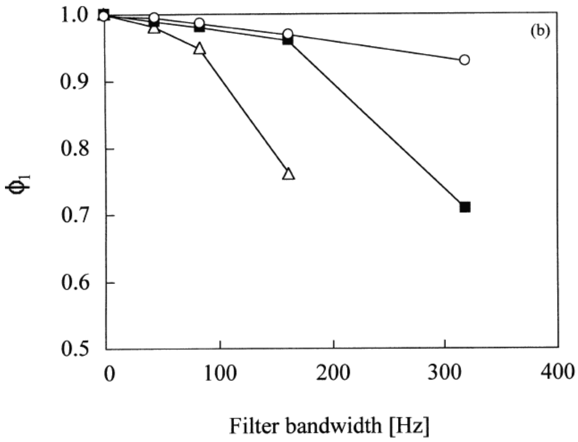


Figure 5.9b

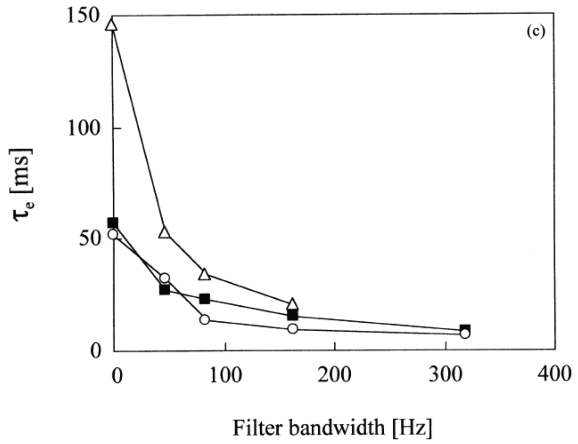


Figure 5.9c

The loudness judgment was performed by the 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. Five subjects with normal hearing participated in the experiment. They were seated in an anechoic chamber and asked to judge which of two paired sound signals were perceived to be louder. Stimulus durations were 1.0 s, rise and fall times were 50 ms, and silent intervals between the stimuli were 0.5 s. A silent interval of 3.0 s separated each pair of stimuli, and the pairs were presented in random order. Fifty responses (5 subjects x 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 ($p < 0.05$) agreement among all subjects. A scale value of loudness was obtained by applying the law of comparative judgment (Thurstone's case V) and was confirmed by goodness of fit.

The relationship between the scale value of loudness and the filter bandwidth is shown in FIGURE 5.10. 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 large bandwidths, with minima at smaller 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 almost 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 ($p < 0.01$). When the results of loudness with changes in the reverberation time T_{sub} of the sound field are taken into account, the fact is that the factor τ_e, a measure of repetitive features of the sound signal, may contribute to the loudness (Merthayasa, Hemmi and Ando, 1994).

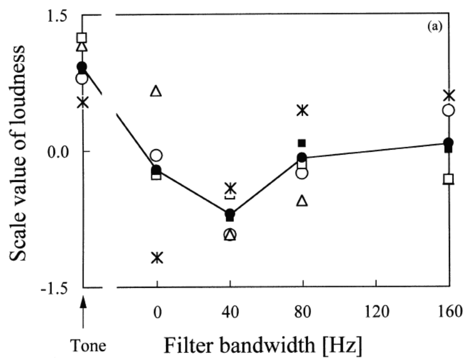


Figure 5.10a

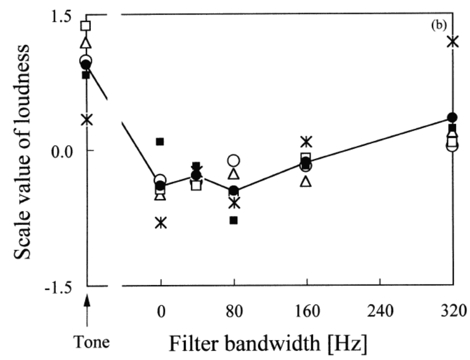


Figure 5.10b

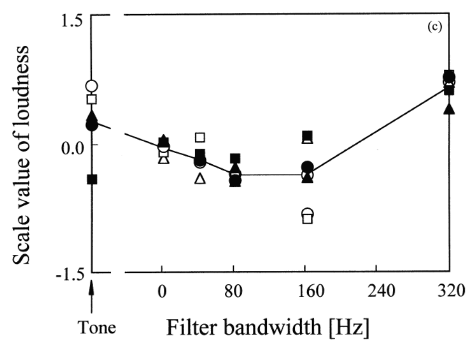


Figure 5.10c

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. Therefore, Equation (5.3) within the critical band may be reduced by

$$S = S_L = f_L(\tau_1) + f_L(\tau_e) \qquad (5.4)$$

In fact, records of auditory-evoked magnetic fields showed that the N1m magnitude decreases with increasing bandwidth when the bandwidth is less than the critical bandwidth. And, it increases with increasing bandwidth beyond the critical band (Soeta, Nakagawa and Matsuoka, 2005).

5.4 Duration Sensation

The sensation of temporal duration (DS) is introduced here, and is also part of a musical note. We will discuss timbre in the following section. Perceptually, the duration sensation depends on the physical signal duration, D, of course. In terms of internal auditory representations, we shall show that it is influenced by one of temporal factors, τ_1 extracted from the ACF, which corresponds to pitch (Ando, Saifuddin and Sato, 2002).

In this section, the perception of duration for complex tones is discussed (Saifuddin, Matsushima and Ando, 2002). An experimental study for pure and complex tones was performed by the PCT. The sound pressure level was fixed at 80 dBA throughout this investigation. 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 threshold 3 dB below the steady level. The perceived durations of the two-component complex tones (3000 and 3500 Hz) having a fundamental at 500 Hz were compared with those evoked by a 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. There were thus 9 stimulus conditions, and 36 pair-wise stimulus combinations. The source stimuli were presented in a darkened soundproof chamber from a single loudspeaker at the horizontal distance of 74 (±1) cm from the center of the seated listener's head. Ten, 22-to 36-years old with normal hearing levels participated in the experiment. 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 FIGURE 5.11. 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, F_0 = 500 Hz) were almost identical, while durations for pure tones of different frequencies (500 vs. 3000 Hz) differed significantly, by approximately 10 ms (judging from equivalent scale values, the 500 Hz pure tone appeared about 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 (fundamental frequency: 500 Hz; τ_1 = 2 ms). Also, the scale values of DS between the two pure tones: τ_1 = 2 ms (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, Equation (5.2) may be reduced to

$$S = S_L = f_L(\tau_1,D) = f_L(\tau_1) + f_L(D) \qquad (5.5)$$

where τ_1 is extracted from the stimulus ACF. FIGURE 5.12 shows the normalized stimulus ACF. Here τ_1 corresponds to the missing fundamental, which is the pitch that can be heard for fundamental periodicities below roughly 1.2 kHz.

The significant results of this study are summarized below.

Apparent stimulus duration DS depends primarily on the duration of the signal, and secondarily on signal periodicity τ_1 (pure tone frequency or complex tone fundamental frequency).

Effects of the τ_1 extracted from the ACF on DS are almost the same on the scale value for the pure-tone (τ_1 = 2 ms) and complex-tone (τ_1 = 2 ms) stimuli. The apparent duration DS of the pure-tone stimulus (τ_1 = 0.33 ms = 1/3000 Hz) with the higher pitch is significantly shorter than that of the pure-tone and complex-tone stimuli with the lower pitch (τ_1 = 2 ms = 1/500 Hz).

3. While apparent duration DS can be readily expressed as a function of D and τ_1 for both pure and complex tones.

4.

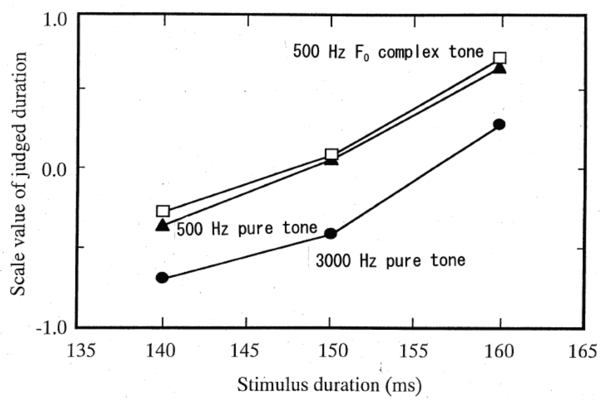


Figure 5.11

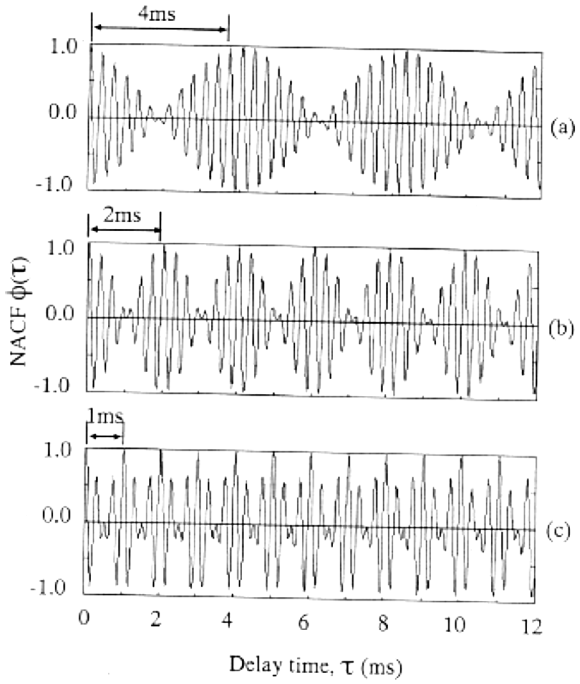


Figure 5.12

5.5 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 FIGURE 5.2, a factor $W_{\phi(0)}$ is defined by the delay time at the first 0.5 crossing of the normalized ACF, $\phi(\tau)$. 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 Wartens (2005) investigated timbre variations by use of three types of nonlinear distortion processors with 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.

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 computer. The distortion of music signal p(t) was processed by a computer program, such that: when $|p(t)| \le C$

$$p(t) = p(t), \qquad (5.6a)$$

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

$$p(t) = +C,\ p(t) \ge C;\quad p(t) = -C,\ p(t) \le -C \qquad (5.6b)$$

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

$$CL = 20\log_{10}(C/|p(t)|_{max}) \qquad (5.7)$$

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

The value of CL was varied as 0 to – 49 dB (7 dB step), so that eight stimuli were applied for test signals. Pitch, signal duration and listening level were fixed. Subjects participated were 19 students (male and female of 20 years of age). Subjects listened to three stimuli, and judged 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. This value is different from the scale value obtained by the law of comparative judgment.

We analyzed contributions to the scale value of other factors, for example, the mean value of $W_{\phi(0)}$, the decay rate of SPL [dBA/s] and the mean value of ϕ_1 (pitch strength). It was found that the most significant factor contributing to the scale value was the mean value of $W_{\phi(0)}$. Certain correlations between the mean value of $W_{\phi(0)}$ and other factors were found, so that the mean value of $W_{\phi(0)}$ is considered as the representative. The scale value as a function of the mean value of $W_{\phi(0)}$ is shown in FIGURE 5.13. The correlation between the SV and the value of $W_{\phi(0)}$ is 0.98 ($p < 0.01$).

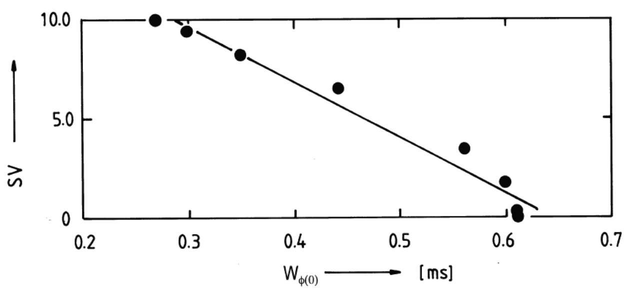


Figure 5.13

6. PRIMARY SPATIAL SENSATIONS

6.1. Sound Localization

Spatial sensations include the localization, the apparent source width, and the subjective diffuseness (envelopment) in the sound field. These are described by the multiple spatial factors extracted from the IACF for the signal arriving at the two ear entrances.

Spatial factors (LL, IACC and τ_{IACC}) are associated with the right hemispheres, which have been found by the analysis of SVR, EEG and MEG as mentioned above.
It is considered that the perceived direction of a sound source in the horizontal plane is expressed in terms of the spatial factors extracted from the IACF, such that

$$L_{Holizontal} = S_R = f_R(\Phi_{ll}(0), \Phi_{rr}(0), IACC, \tau_{IACC}, W_{IACC}) \quad (6.1)$$

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 Equation (6.1) is the inter-aural delay time, τ_{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. On the other hand, subjective diffuseness or no spatial directional impression corresponds to a low value of IACC < 0.15 (Damaske and Ando, 1972) and a wide delay time of W_{IACC} due to the low frequency components.

6.2. Apparent Source Width (ASW)

As discussed in the book (Ando, 2009a), ASW may be described by the spatial factors extracted from the IACF, such that

$$S = S_R = f(IACC) + f(W_{IACC}) + f(LL) \approx \alpha(IACC)^{3/2} + \beta(W_{IACC})^{1/2} + \gamma(LL)^{3/2}$$
((6.2)

where $\alpha \approx -1.64$, $\beta \approx 2.42$, $\gamma \approx 0.005$.

For example, the scale value of the ASW has been obtained by the PCT with ten subjects (Sato and Ando, 1996). In order to control the value of W_{IACC}, the center frequency of the 1/3-octave band-pass noises was changed as 250 Hz, 500 Hz, 1 kHz, and 2 kHz. The value of IACC was adjusted by controlling the sound pressure ratio between reflections ($\xi = \pm 54^o$) and the direct sound ($\xi = 0^o$). To avoid effects of the listening level on the 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. Subjects judged which of two sound sources they perceived to be wider.
Results of the analysis of variance for the scale value S_R indicate that both factors the IACC and W_{IACC} are contribute to S_R independently ($p < 0.01$), so that

$$S = S_R = f_R(IACC) + f_R(WIACC) \approx \alpha(IACC)^{3/2} + (WIACC)^{1/2} \quad (6.3)$$

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 FIGURE 6.1. This holds under the condition of $\tau_{IACC} = 0$. Obviously, as shown in FIGURE 6.2, the calculated scale values by Equation (6.3) and measured scale values are in good agreement ($r = 0.97$, $p < 0.01$).

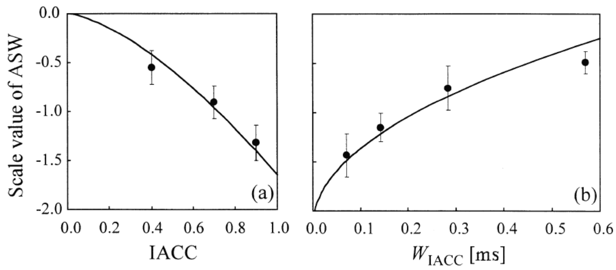


Figure 6.1

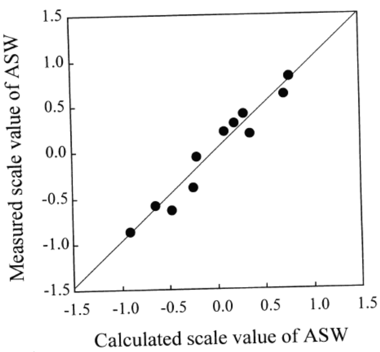


Figure 6.2

For each individual listener, the scale value can be calculated by Equation (6.3) in the similar manner. Coefficients α and β in the equation for each listener are obtained by the multiple regression analysis as well as indicated in Table 6.1. FIGURE 6.3 shows the relationship between the measured scale values and the calculated scale values with the constants for each of ten subjects. The different symbols indicate the scale values of different subjects. The correlation coefficient between the measured and calculated S(ASW) is 0.91 ($p < 0.01$).

Individual	α	β	Corre lation coefficient
SH	-1.21	2.58	0.88
TS	-1.50	3.18	0.97
CC	-1.05	2.82	0.97
SY	-0.94	2.92	0.91
MK	-2.21	2.09	0.92
ST	-2.57	1.94	0.94
TH	-2.04	1.32	0.87
FK	-0.99	3.27	0.89
NK	-1.79	2.14	0.80
OS	-2.09	2.14	0.94
Average	-1.64	2.44	0.97

TABLE 6.1 Coefficients α and β in Equation (6.3) for calculating ASW of each individual, and the correlation coefficient between the measured and calculated scale values of the ASW.

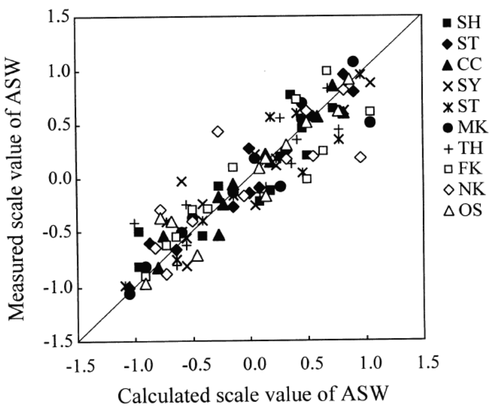


Figure 6.3

6.3. Subjective Diffuseness

The scale value of subjective diffuseness of the sound is described by the representative spatial factor, the IACC.
In order to obtain the scale value of subjective diffuseness, the PCT with the 1/3 octave 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. A remarkable finding is that the scale values of 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.,

$$S = S_R \approx -\alpha(IACC)^{\beta} \qquad (6.4)$$

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

The results of the scale value from the PCT together calculated values by Equation (6.4) as a function of the IACC, are shown in FIGURE 6.4. There is a great variation of data in the range of the IACC < 0.5, however, no essential difference may be found in results with different frequencies between 250 Hz - 4 kHz.

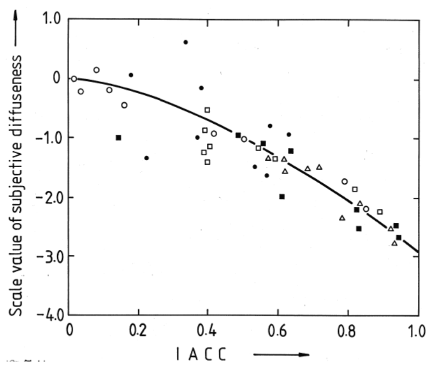


Figure 6.4

The most effective horizontal angles of reflections minimizing IACC depend on the frequency range as shown in FIGURE 6.5. These are about $\pm 90^{O}$ for the low frequency range less than 500 Hz, and around $\pm 55^{O}$ for the 1 kHz range, which is the most important angle for the music, and smaller than 18^{O} for the 2 kHz and 4 kHz ranges. The control of such a directional reflection for each frequency range may be realized by means of a Fractal structure of the wall surface (Ando, 1998), for example.

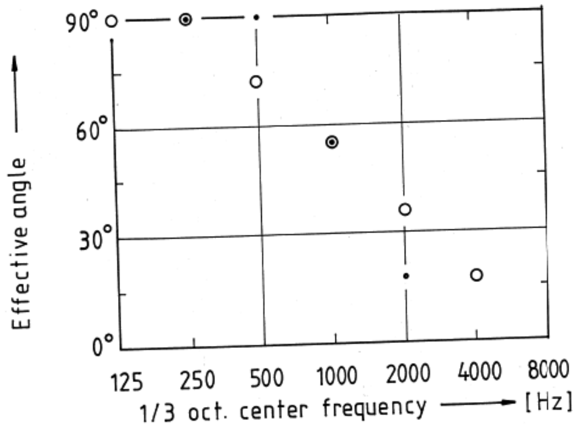


Figure 6.5

7. APPLICATIONS FOR MEASUREMENT AND EVALUATION OF NOISE, MUSIC AND SPEECH

A basic procedure for measurement and evaluations of noise, music and speech is shown in FIGURE 7.1, and the system is shown in FIGURE 7.2. In this section, only two examples of application for noise and music are discussed. Applications for speech identification have been discussed (Ando, 2009a). It is most promising that if investigations are continued for this area, simultaneous speech interpretation system in a mobile phone could be available for existing about 2000 different languages.

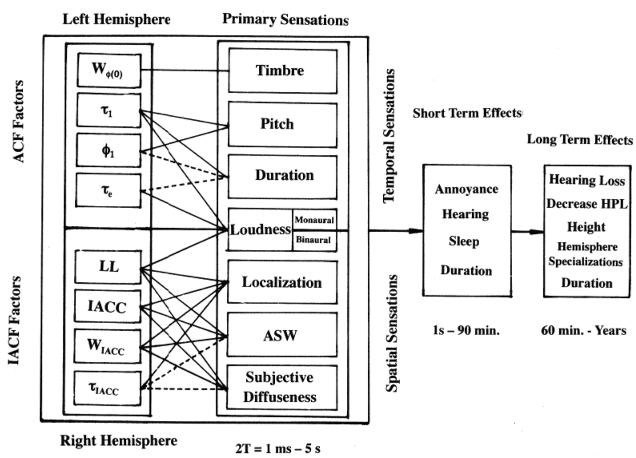


Figure 7.1

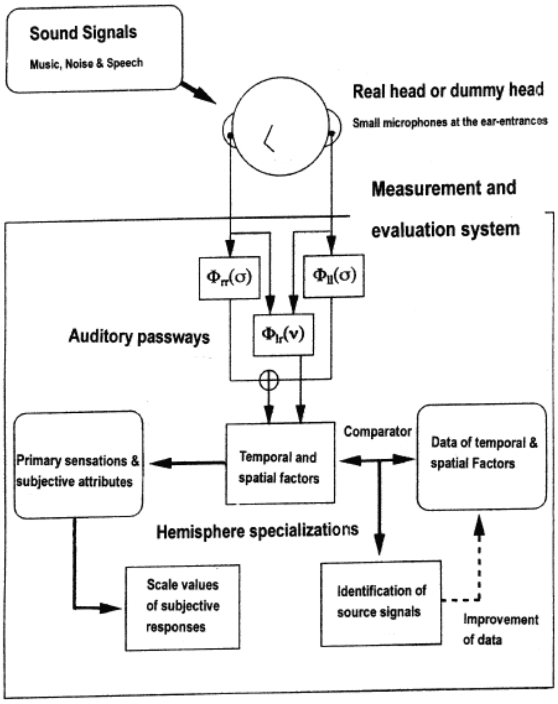


Figure 7.2

7.1. Annoyance of Traffic Noise

The purpose of this section is to discuss the annoyance of traffic noise in relation to only temporal factors, which are extracted from the ACF of noise. In order to avoid effects of spatial factors, the single loudspeaker located in front of a subject reproduced the noise recorded in a listening room.

The measured SPL and three ACF factors of the noise are shown in FIGURE 7.3 as a time function (Fujii, Atagi and Ando, 2002). Thick lines and thin lines show two extremes of noise: one has a clear pitch as a tonal, and the other has a weak pitch as a non-tonal because of a small value ϕ_1 near to 0.2. The value of τ_1 varied discretely 1 ms and 10 ms, meaning that perceived pitch varied 1000 Hz and 100 Hz for both noises. For the tonal noise, the ϕ_1 value reaches maximum around 0.7. Thus, a strong tonal noise is heard having a pitch of 100 Hz (τ_1 = 10 ms). When the τ_1 value varies with a high ϕ_1 value, we perceive a variation of the pitch. If the ϕ_1 value for the un-tonal noise remained constant around 0.2, in spite of the variation of τ_1, then perceived pitch for the un-tonal noise is weak enough, and it is hard to discriminate pitch fluctuation.

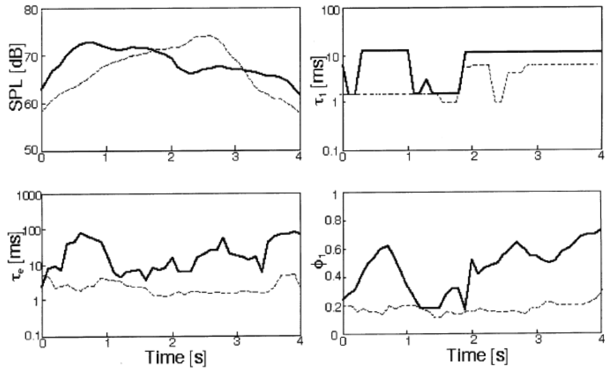


Figure 7.3

In addition, the so-called previous standard measures may be extracted from the running ACF. (1) LL = mean SPL [dBA] due to $\Phi_{ll}(0)$ and $\Phi_{rr}(0)$, (2) variance σ^2 of the SPL, (3) maximum SPL, (4) minimum SPL, (5) the SPL values exceeded 10 % of the time (L_{10}), (6) 50 % of the time (L_{50}), (7) 90 % of the time (L_{90}), and (8) equivalent sound level LA_{eq}. As indicated in Table 7.1, however, most of these standard measures were highly inter-correlated, and thus not orthogonal. Clearly, all of these factors contain information about the overall sound level and its variability. Therefore, only the median (L_{50}) as a representative of the mean SPL and LA_{eq}, and the variance of the SPL (Var_SPL) as a representative of maximum SPL, minimums SPL, L_{10}, L_{90} were selected in the subsequent analysis.

Factor	max	min	mean	σ^2	L_{50}	L_{eq}	L_{90}	L_{10}
max	1							
min	-0.69*	1						
mean	-0.24	0.62	1					
σ^2	0.89**	-0.91**	-0.53	1				
L_{50}	0.06	0.24	0.84**	-0.11	1			
L_{eq}	0.57	-0.31	0.44	0.48	0.82**	1		
L_{90}	-0.85**	0.87**	0.66*	-0.97**	0.27	-0.32	1	
L_{10}	0.70*	-0.46	0.28	0.63*	0.70**	0.98**	-0.49	1

** $p < 0.01$, * $p < 0.05$

Table 7.1. Correlations between previously defined eight standard noise measures for nine different traffic noises.

The calculated power spectrum without any filter, and the ACF after passing through the A-weighting network for the tonal and the non-tonal noise is shown in FIGURE 7.4. For the tonal noise, there are several discrete peaks in the spectrum, but it is difficult to identify which corresponds to the clear pitch. The strong initial peak in the ACF signifies the periodicity corresponding to the missing fundamental or pitch. Minor peaks within the first peak in the normalized ACF give information about the higher-frequency components or timbre of the sound (Midis and Hewitt, 1991a; 1991b, Cariani and Delgutte, 1996a: 1996b). The information can be used to identify the noise source. Also, the envelope of the ACF represented by the value of τ_e is a good measure of the repetitive feature of the sound signal.

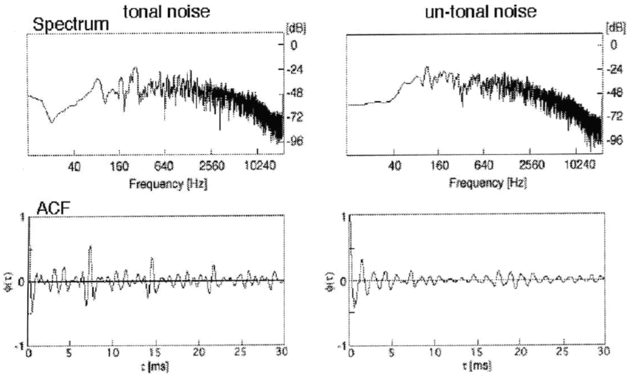


Figure 7.4

Nine recordings of noise were used in the annoyance experiment. Each stimulus was a 4-s duration from a single vehicle's passage. The maximum level near the middle of the sound was adjusted to be equal (73 ± 2 dB). To make the envelope of noise equal, a 0.5-s rise and fall time was added to all stimuli. To characterize the acoustical properties of a stimulus, we used the median and variance of each factor.

In order to find the effects of such fluctuations of noise level and its quality, we added the variance of SPL and ACF factors to the variables for describing the annoyance. The single loudspeaker was used to keep the spatial properties of the sound field constant. The subjects sat 1.0 m in front of the loudspeaker. Ten subjects (nine males and one female) participated in the experiment. They were between the ages of 23 and 27, in good health, with normal auditory acuity. Annoyance was judged by the PCT. All possible pairs from the nine sounds (36 pairs) were presented to the subject in a random order in one session. After the presentation of paired stimuli, the subject was asked to judges, which of the two noises were more annoying. All subjects had four series of sessions, giving a total of 144 comparisons.

The scale value of annoyance for all the subjects was averaged, and the correlation coefficients were calculated between the annoyance and the median value and variance of the ACF factors. The correlation matrix between the ACF factors and SV is shown in Table 7.2. Contrary to the assumption, perceived annoyance was not correlated to the SPL in this study due to the fact that the limited range of the SPL among the stimuli was 5.0 dBA. Instead, the variance of the SPL had greatly affected on annoyance. In other words, Var.-SPL, τ_e and Var-τ_1 are more effective on the scale value of annoyance than SPL in the range of about 5 dB. The values of τ_e and ϕ_1 were significantly correlated to annoyance ($r = 0.56$ and 0.57, respectively, $p < 0.05$). This result shows that the noise having a strong tonal component was perceived to be more annoying than the non-tonal_noise. The subjects' comments also indicated that they judged a sound having a clear pitch to be more annoying.

Factor	SPL	τ_1	ϕ_1	τ_e	Var_SPL	Var_τ_1	Var_ϕ_1	Var_τ_e
SPL	1							
τ_1	-0.66	1						
ϕ_1	-0.29	0.82**	1					
τ_e	0.34	0.33	0.74**	1				
Var_SPL [1]	-0.11	0.02	0.22	0.03	1			
Var_ τ_1	-0.57	0.46	0.37	0.35	-0.04	1		
Var_ ϕ_1	-0.09	0.50	0.30	0.78**	-0.35	0.12	1	
Var_ τ_e	-0.15	0.59*	0.77**	0.78**	0.13	0.33	0.58*	1
annoyance	0.11	0.30	0.57*	0.56*	0.64*	0.39	0.20	0.67*

** $p < 0.01$, * $p < 0.05$

[1] Symbol Var represents the variance defined by $\sigma^2 = 1/n\,\Sigma\,(x_i\text{-}m)^2$, where x_i is the value of data i, m is the mean value, and n is the number of the data.

Table 7.2. Correlations between the ACF factors (median and variance) and annoyance

Previously, on the evaluation of the perceived noise level for the tonal noise as used in the experiment, a number of tone corrections were used. For the sake of convenience, a value was added to the perceived noise level (PNL) to give the tone corrected perceived noise level (PNLT). However, the calculation for this correction is lengthy, and their accuracy is not well established (May, 1978). Instead, as is discussed here by using the value of τ_e and ϕ_1, effects of the tonal component on perceived annoyance may be clearly described. As indicated in Table 7.2,

1)Var_SPL,

2)τ_e as a representative of ϕ_1, Var._ ϕ_1 and Var_τ_e, and

3)Var_τ_1 as a representative of SPL and τ_1,

are considered independent factors affecting annoyance. A linear combination of the three variables obtaining the scale value of annoyance may be made approximately, such that,

$$S = S_L \approx a\mathrm{Var\_SPL} + b\tau_e + c\mathrm{Var\_}\tau_1 \qquad (7.1)$$

where coefficients obtained are: $a \approx 0.64$, $b \approx 0.50$ and $c \approx 0.36$. Using these tentative constants in Equation (7.1), the total correlation coefficient 0.91 was obtained with the significance level $p < 0.05$, as shown in FIGURE 7.5. Similarly, the noise whose pitch and timbre fluctuates is more annoying than one having a constant quality (Molino, 1979).

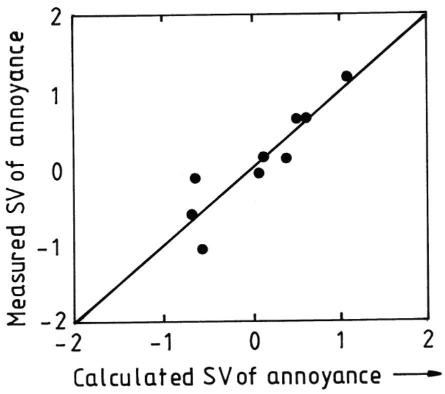


Figure 7.5

This result shows that the combination of the ACF factors was sufficient and simple to calculate perceived annoyance. In addition to the fluctuation in SPL, tonality of the noise and pitch fluctuation is the important factor for annoyance. These factors may be extracted from the ACF may be applied for any environmental noise evaluations (Sato, You and Jeon, 2007).

7.2. Preferred Delay Time of the Single Reflection for Cellists

The present study evaluates the subjective preferences, with regard to ease of performance, of five cello soloists for the delay time of a single reflection. The scale values of preference for the delay time of a single reflection were obtained using a paired comparison method. The scale values of preference for both individuals and for global cellists with regard to the delay time of reflection can be expressed by a single approximate formula with different constants, normalizing the delay time by the most-preferred delay time observed for different music motifs. A notable finding is that the most-preferred delay time of a single reflection for each cellist can be calculated from the amplitude of the reflection and the minimum value of the effective duration $(\tau_e)_{min}$ of the running ACF of the music motifs played by each cellist.

In order to realize an excellent concert, we need to design the sound fields not only in the audience area but also in the stage area for performers. The primary issue is that the stage enclosure should be designed to provide a sound field in which performers can play easily. Marshall, Gottlob and Alrutz (1978) investigated the effects of stage size on the playing of an ensemble. The parameters related to stage size in their study were the delay time and the amplitude of reflections. Gade (1989) performed a laboratory experiment to investigate the preferred conditions for the total amplitude of the reflections of performers. On the other hand, the preferred delay time of a single reflection for listeners can be calculated by the effective duration of the long-time ACF of the source signal and the amplitude of reflections (Ando, 1977).

When music signals contain a large fluctuation in tempo, it is more accurately expressed by the minimum value of the effective duration $(\tau_e)_{min}$ of the running ACF of source signal (Ando, Okano and Takezoe, 1989). Nakayama (1984) showed that the amplitude of the reflection and the duration of the long-time ACF of the source signal as a similar manner could determine the preferred delay time of a single reflection for alto-recorder soloists for listeners (see also, Ando, 1998). Noson, et al. also reported that the most preferred condition of the single reflection for individual singer may be described by the $(\square_e)_{min}$ and a modified amplitude of reflection by the over estimate and bone conduction effect (Noson, Sato, Sakai and Ando, 2000; 2002).

The present study examines whether or not the amplitude of the reflection can calculate the preferred delay time of a single reflection for each cello soloist and the minimum value of the effective duration of the running ACF of the music motifs played by that cellist (Sato, Ohta and Ando. 2000).

Music motifs may be characterized in terms of the running ACF of the source signal played. The same music motifs (motifs I and II) used in the experiments conducted by Nakayama were applied here. The tempo of motif I was faster than that of motif II as shown in FIGURE 7.6. A microphone in front of the cellist picked up the music signal performed by each of five cellists. The distance between the microphone and the center of the cello body was 50 ± 1.0 cm. The music tempo was maintained with the help of a visual and silent metronome. Each music motif was played three times by each cellist. The minimum value of the effective duration $(\tau_e)_{min}$ of the running ACF of a music signal is the most active part of the music signal, containing important information and influencing the subjective attributes related to the temporal factors. It was analyzed after passing through the A-weighted network with the integration interval, 2T = 2.0 s, which was chosen according to Equation (5.3). Usually, the envelope decay of the initial part of the ACF can be fitted by a straight line in the range from 0 dB to -5 dB to obtain the effective duration $\square_e$ by the extrapolation at -10 dB. Examples of effective durations of the running ACF for music motif I played by subjects B and E are shown in FIGURE 7.7 and FIGURE 7.8, respectively. The minimum value of the effective duration $(\tau_e)_{min}$ of the running ACF for each cellist and each session are listed in Table 7.3. For all cellists, the effective durations $(\tau_e)_{min}$ for music motif I were about a half of those for music motif II. Mean values of $(\tau_e)_{min}$ were 46 ms for music motif I and 84 ms for music motif II, and for both motifs the ranges of $(\tau_e)_{min}$ are within ± 5 ms. Individual differences in the effective durations of the running ACF may depend on the performer's style.

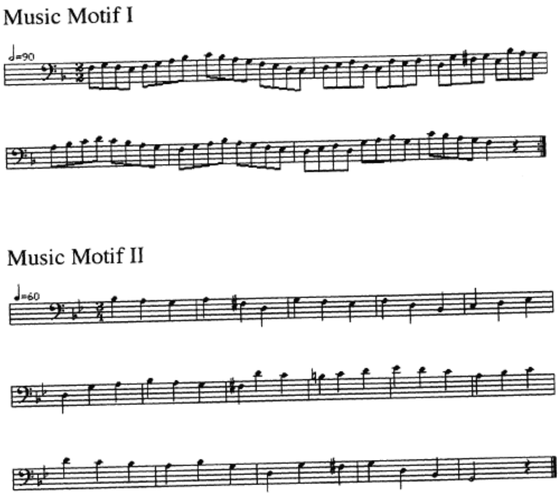


Figure 7.6

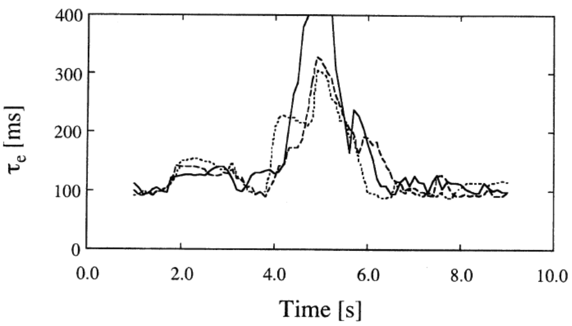


Figure 7.7a

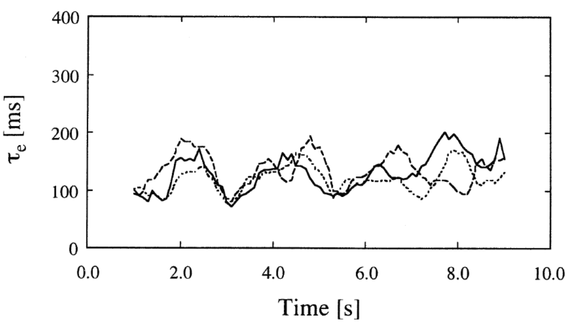


Figure 7.7b

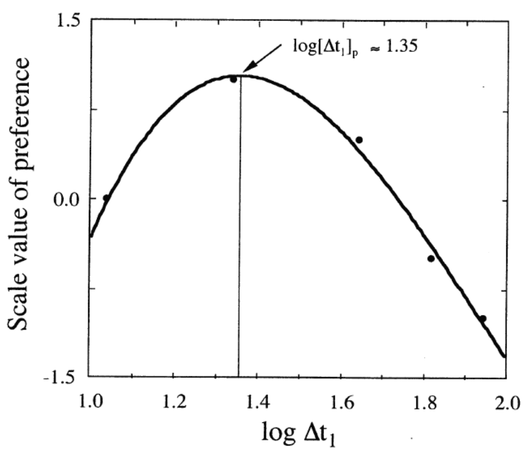


Figure 7.8

A [dB]	A'[dB] (= A+10)	A'	Cellist	Judged $[\Delta t_1]_p$ [ms] Motif I	Motif II	Calculated $[\Delta t_1]_p$[ms] Motif I	Motif II
			A	16.2	47.9	16.3	38.5
			B	< 12.0	73.8	35.2	62.7
-15	-5	0.56	C	< 12.0	60.8	21.3	51.3
			D	22.6	38.2	35.1	53.9
			E	17.6	63.6	17.3	35.2
			Global	18.0	48.3	24.3	47.5
			A	18.1	48.4	21.8	51.5
			B	61.2	105.0	59.3	105.6
-21	-11	0.28	C	---	77.9	---	80.6
			D	74.6	86.8	56.9	87.4
			E	< 14.0	42.2	24.8	50.2
			Global	30.4	71.8	37.6	73.4

Table 7.3 Judged and calculated preferred delay times of a single reflection for each cello soloist. Calculated values of $[\Delta t_1]_p$ are obtained by (10.15) using the amplitude of the reflection A'_1 and $(\tau_e)_{min}$ for music signals performed by the cellist.

The single reflection from the back wall in the stage enclosure was simulated in an anechoic chamber by a loudspeaker 80 ± 1.0 cm from the head of the cellist. The sound signal was picked up by a 1/2-inch condenser type microphone at the entrance of the performer's left ear and was reproduced by the loudspeaker after passing through a digital delay device. The amplitudes of reflection A_1, relative to that of the direct sound measured at the entrance of the performer's left ear, was kept constants at -15 dB or -21 dB when the cellist played a musical note 'a' (442 Hz).

Paired comparison tests were conducted for the five sound fields. The delay times of the reflection was changed, because the preferred delay time of a single reflection might depend on the $(\tau_e)_{min}$ of the running ACF of source signal. The five subjects were asked which of two sound fields was easier for them to perform in. The test consisted of 10 pairs (N(N-1)/2, N = 5) of stimuli in total, and for all subjects the test was repeated three times interchanging the order of the pairs. It took about 20 minutes for each cellist and for each music motif. Fifteen responses (5 subjects x 3 repeats) to each sound field were obtained and were confirmed by consistency tests. The scale values of preference for each cellist were obtained by applying the paired comparison method as described (Ando and Singh, 1996; Ando 1998).

FIGURE 7.9 shows an example of the regression curve for the scale value of preference and the method of estimating the most preferred delay time $[\Delta t_1]_p$. The peak of this curve denotes the most-preferred delay time. The most-preferred delay times for individual cellists and the global preference results are listed in Table 7.3. Global and individual results (except for that of subject E) for music motif II were longer than those for music motif I.

It is considered that the most-preferred delay time of a single reflection also is described by the duration τ'_p of the ACF as similar to that of listeners, which is expressed by

$$[\Delta t_1]_p = \tau'_p \quad (7.2)$$

such that

$$\left|\phi_p(\tau)\right|_{envelope} \approx kA'^c \quad \text{at } \tau = \tau'_p, \quad (7.3)$$

The values k and c are constants that depend on the subjective attributes. The value of A' is the amplitude of the reflection being defined by A' = 1 relative to -10 dB of the direct sound as measured at the ear's entrance. This is due to the over-estimation of the reflection by the performer. If the envelope of the ACF is exponential, Equation (7.3) simply yields

$$\tau'_p = (\log_{10}1/k - c\log_{10}A')\,(\tau_e) \quad (7.4)$$

According to a previous study the effective duration τ_e of the long-time ACF in Equation (7.4) is replaced by the minimum value of the effective duration $(\tau_e)_{min}$ of the running ACF of music used for judgments, so that

$$[\Delta t_1]_p = \tau'_p = (\log_{10}1/k - c\log_{10}A')\,(\tau_e)_{min} \quad (7.5)$$

Using the Quasi-Newton method, the values $k \approx 1/2$ and $c \approx 1$ are obtained. (It is worth noting that the coefficients k and c for alto-recorder soloists were respectively 2/3 and 1/4 and for listeners was respectively 0.1 and 1.) After setting k = 1/2, we obtained the coefficient c for each individual as listed in Table 7.4. The average value of the coefficient c for the five cellists obtained was about 1.

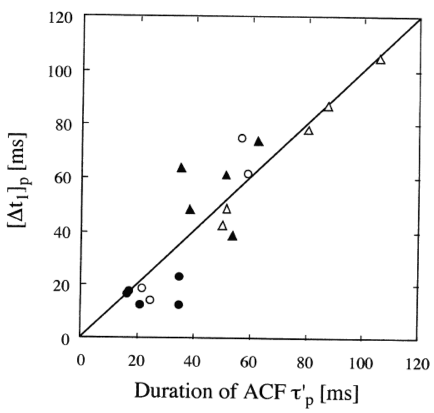


Figure 7.9

	Cellist A	B	C	D	E	Averaged (Global)
Coefficient c'	0.47	1.61	1.10	1.30	0.67	≈ 1.0

Table 7.4 Coefficient c' for each cellist in (10.15) calculating the preferred delay time of the reflection for individual results with global one, when k' = 1/2 (fixed).

The relation between the most-preferred delay time $[\Delta t_1]_p$ obtained by preference judgments and the duration τ'_p of the ACF calculated by Equation (7.5) using $(\tau_e)_{min}$ is shown in FIGURE 7.9. Different symbols indicate the values obtained in different test series. The correlation coefficient between calculated values of $[\Delta t_1]_p$ and measured values is 0.91 ($p < 0.01$). The scale values of preference for each of the five cellists as a function of the delay time of the single reflection normalized by the calculated $[\Delta t_1]_p$ are shown in FIGURE 7.10. Different symbols indicate the scale values obtained in different test series. Each symbol has 25 data (5 subjects × 5 sound fields) except for the amplitude of -15 dB for music motif I (for which there were 20 data because consistency tests did not indicate a significant ability to discriminate preference in the results of Subject C). The scale values obtained in different test series are consistent with each other. The regression curve is expressed by

$$s \approx -\alpha |x|^{\beta} \quad (7.6)$$

where $x = \log\Delta t_1/[\Delta t_1]_p$, the power of x is fixed by β= 3/2 and the weighting coefficient α is 2.3 for $x \geq 0$ and 1.0 for $x < 0$.

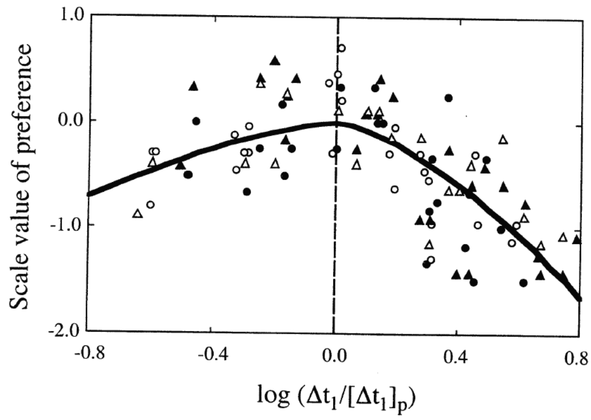


Figure 7.10

As an application, adjusting the height of the reflectors above the stage can control the delay time of a reflection. The optimum distance between the performer and the rigid reflector above the stage in relation to the minimum value of the effective duration $(\tau_e)_{min}$ of the running ACF of the music program to be performed can be calculated. Concequently, the most-preferred delay time of a single reflection for each cellist can be calculated by Equation (7.5) with the amplitude of the reflection and the minimum value of the effective duration $(\tau_e)_{min}$ of the running ACF of the music motifs played by each cellist. The scale values of preference for both individual cellists and thus for global cellists with regard to the delay time of a single reflection can be expressed by the simple Formula (7.6), normalizing the delay time by the most-preferred delay time calculated $[\Delta t_1]_p$ by Equation (7.5) with $(\tau_e)_{min}$ for each music signal performed.

7. CONCLUSIONS

As discussed above, resulting theory obtained here is so "simple and beauty," and followings are conclusions:

1. The temporal and spatial factors extracted from respective correlation functions existing human auditory-passways system well describe primary sensations consisting of temporal and spatial sensations as well as subjective preference and annoyance for both global and individual levels.
2. The cerebral hemisphere specialization may play an important role for the independent effects on temporal and spatial sensations of the temporal and spatial factors, respectively.
3. The scale values of subjective preference of the sound field are described by both temporal and spatial factors, which may be directly observed by alpha rhythms (τ_e) in cortical populations.
4. Possible applications on acoustics are, for example, noise evaluations, music performance and instruments as well as speech recognition. All these systems might be available by use of a "computer tip", for example, in such as like a mobile phone.

It is worth noticing that brain oriented theory might be apply for other modarities including visual field (Ando, 2009a), and a theory of temporal and spatial environmental design has been proposed (Ando, 2009b).

8. REFERENCES

Ando, Y., Shidara, S., Maekawa, Z., and Kido, K. (1973). Some basic studies on the acoustic design of room by computer. Journal of the Acoustical Society of Japan, **29**, 151-159 (in Japanese with English abstract).

Ando, Y. and Kageyama, K. (1977). Subjective preference of sound with a single early reflection. Acustica, **37**, 111-117.

Ando, Y. (1977). Subjective preference in relation to objective parameters of music sound fields with a single echo. Journal of the Acoustical Society of America,**62**, 1436-1441.

Ando, Y., and Morioka, K. (1981). Effects of the listening level and the magnitude of the interaural cross-correlation (IACC) on subjective preference judgment of sound field. Journal of the Acoustical Society of Japan, **37**, 613-618 (in Japanese with English abstract).

Ando, Y., and Alrutz, H. (1982). Perception of coloration in sound fields in relation to the autocorrelation function. Journal of the Acoustical Society of America,**71**, 616-618.

Ando, Y., Okura, M., and Yuasa, K. (1982). On the preferred reverberation time in auditoriums. Acustica, **50**, 134-141.

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., Otera, K., and Hamana, Y. (1983). Experiments on the universality of the most preferred reverberation time for sound fields in auditoriums. Journal of the Acoustical Society of Japan, **39**, 89-95 (in Japanese with English abstract).

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., Kang, S. H., and Nagamatsu, H. (1987). On the auditory-evoked potentials in relation to the IACC of sound field. Journal of the Acoustical Society of Japan, (E), **8**, 183-190.

Ando, Y., Kang, S.H., and Morita, K. (l987). On the relationship between auditory-evoked potential and subjective preference for sound field. Journal of Acoustical Society of Japan (E), **8**, 197-204.

Ando, Y. (1988). Effects of daily noise on fetuses and cerebral hemisphere specialization in children. Journal of Sound and Vibration, **127**, 411-417.

Ando, Y., and Sakamoto, M. (1988). Superposition of geometries of surface for desired directional reflections in a concert hall. Journals of the Acoustical Society of America, **84**, 1734-1740.

Ando, Y., Okano, T., and Takezoe, Y. (1989). The running autocorrelation function of different music signals relating to preferred temporal parameters of sound fields. Journal of the Acoustical Society of America, **86**, 644-649.

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., Johnson, B., and Bosworth, T. (1996). Theory of planning environments incorporating spatial and temporal values. Memoirs of Graduate School of Science and Technology, Kobe University, **14-A**, 67-92.

Ando, Y., and Singh, P.K. (1996). A simple method of calculating individual subjective responses by paired-comparison tests. Memoirs of Graduate School of Science and Technology, Kobe University, **14-A**, 57-66.

Ando, Y. and Noson, D. (1997). Eds. Music and Concert Hall Acoustics, Conference Proceedings of MCHA 1995. Academic Press, London.

Ando, Y., Sato, S., Nakajima, T., and Sakurai, M. (1997). Acoustic design of a concert hall applying the theory of subjective preference, and the acoustic measurement after construction. Acustica acta Acustica, **83**, 635-643.

Ando, Y. (1998). Architectural Acoustics, Blending Sound Sources, Sound Fields, and Listeners. AIP Press/Springer-Verlag, New York.
Ando, Y., Okano, T., and Takezoe, Y. (1989). The running autocorrelation function of different music signals relating to preferred temporal parameters of sound fields. Journals of the Acoustical Society of America, 86, 644-649.
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., Tsuruta, H., Motokawa, A., Matsushita, T., and Saifuddin, K. (1999). Subjective duration of every three-year period for 3 to 18 years of age, estimated by students. Journal of Human Ergology, **28**, 33-37.
Ando, Y. (2001). Differential effects of noise and music signals on the behavior of children. Journal of Sound and Vibration, **241**, 129-140.
Ando, Y., Saifuddin, K., and Sato, S. (2002). Duration sensation when listening to bandpass noises. Journal of Sound and Vibration, **250**, 31-40.
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/
Ando, Y. (2004). Spatial distribution of acoustic parameters in concert halls: comparison of different scattered reflections. Journal of Temporal Design in Architecture and the Environment, **4**, 59-68. http://www.jtdweb.org/journal/
Ando, Y. (2004). On the temporal design of environments. Journal of Temporal Design in Architecture and the Environment, **4**, 2-14. http://www.jtdweb.org/journal/
Ando, Y. (2004). On the temporal design of environments, Journal of Temporal Design in Architecture and the Environment, **4**, 2-14, http://www.jtdweb.org/
Ando, Y. (2006). Reviews on the temporal design for three stages of human life. Most unlikely"time is money," but"time is life."Journal of Temporal Design in Architecture and the Environment, **6**, 2-17, http://www.jtdweb.org/
Ando, Y., (2007). Concert hall acoustics based on subjective preference theory, in Springer Handbook of Acoustics, T. D. Rossing (ed.).
Ando, Y. (2009a). Auditory and visual sensations, Springer-Verlag, New York.
Ando, Y. (2009b). Theory of temporal and spatial environmental design, in McGraw-Hill Yearbook of Science & Technology 2009, McGraw-Hill, New York, p. 384-389.
Buchwald, J.A.S., and Huang, C.M. (1975). Far-field acoustic response: origins in the cat. Science, **189**, 382-384.
Burd, A.N. (1969). Nachhallfreir Musik fuer akustische Modelluntersuchungen. Rundfunktechn. Mitteilungen, 13, 200-201.
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.
Chernyak, R.I., and Dubrovsky, N.A. (1968). Pattern of the noise images and the binaural summation of loudness for the different interaural correlation of noise. Proceedings of the 6th International Congress on Acoustics, Tokyo, Paper A-3-12.
Cherry, E. C., and Sayers, B. M. A. (1956). "Human 'cross-correlator'" – A technique for measuring certain parameters of speech perception. Journal of the Acoustical Society of America, **28**, 889-895.
Cheveigne, A. (1998). Cancellation model of pitch perception. Journal of the Acoustical Society of America, **103**, 1261-1271.
Chon, R. (1970). Differential cerebral processing of noise and verbal stimuli. Science, **172**, 599-601.
Damaske, P., and Ando, Y. (1972). Interaural cross-correlation for multichannel loudspeaker reproduction. Acustica, **27**, 232-238.
de Boer, E. (1976). On the "residue" and auditory pitch perception, in Auditory System (Handbook of Sensory Physiology), Eds. W.D. Keidel and W.D. Neff, Springer Verlag: Berlin. 479-583.
Dubrovskii, N.A., and Chernyak, R.I. (1969). Binaural loudness summation under varying degrees of noise correlation. Soviet Physics - Acoustics, **14**, 326-332.
Fujii, K., Hotehama, T., Kato, K., Shimokura, R., Okamoto, Y., Suzumura, Y., and Gade, A. C. (1989). Investigations of musicians' room acoustic conditions in concert halls. Part I: Methods and laboratory experiments. Acustica, **69**, 193-203.
Greenwood, D. D. (1961a). Auditory masking and critical band. Journal of Acoustical Society of America, **33**, 484-502.
Greenwood, D. D. (1961b). Critical bandwidth and the frequency of the basilar membrane. Journal of Acoustical Society of America, **33**, 1344-1356.
Goldstein, J. L. (1973). An optimum processor theory for the central formation of the pitch of complex tones. Journal of the Acoustical Society of America **54**, 1496-1516.
Gullikson, H. (1956). A least square solution for paired comparisons with incomplete data. Psychometrika, **21**, 125-134.
Fujii, K., Atagi, J., and Ando, Y. (2002). Temporal and spatial factors of traffic noise and its annoyance, Journal of Temporal Design in Architecture and the Environment, **2**, 33-41. http://www.jtdweb.org/journal/

Haas, H. (1951). Ueber den Einfluss eines Einfachechos auf die Hoersamkeit von Sprache. Acustica, **1**, 49-58.
Holland, J. H. (1975). Adaptation in Natural and Artificial Systems, The University of Michigan Press.
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.
Jeffress, L.A. (1948). A place theory of sound localization, Journal of comparative and physiological psychology. **41**, 35-39.
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.
Kang, S.H., and Ando, Y. (1985). Comparison between subjective preference judgments for sound fields by different nations. Memoirs of Graduate School of Science and Technology, Kobe University, **3-A**, 71-76.
Kato, K., and Ando, Y. (2002). A study of the blending of vocal music with the sound field by different singing styles, Journal of Sound and Vibration, **258**, 463-472.
Kato, K., Fujii, K., Kawai, K., Ando, Y., and Yano, T. (2004). Blending vocal music with the sound field – the effective duration of the autocorrelation function of Western professional singing voices with different vowels and pitches, Proceedings of the International Symposium on Musical Acoustics, ISMA 2004, Nara.
Katsuki, Y., Sumi, T., Uchiyama, H. , and Watanabe, T. (1958). Electric responses of auditory neurons in cat to sound stimulation. Journal of Neurophysiology, **21**, 569-588.
Keet, M. V. (1968). The influence of early lateral reflections on the spatial impression. Proc. 6th Intern. Congr. Acoust., Tokyo, Paper E-2-4.
Kiang, N.Y.-S. (1965). Discharge pattern of single fibers in the cat's auditory nerve. MIT Press, Cambridge, MA.
Lev, A., and Sohmer, H. (1972). Sources of averaged neural responses recorded in animal and human subjects during cochlear audiometry (Electro-cochleo-gram). European Archives of Oto-Rhino-Laryngology (Arch. Klin. Exp. Ohr., Nas.-u. Kehlk. Heilk.), **201**, 79-90.
Licklider, J.C.R. (1951). A duplex theory of pitch perception, Ecxperientia, **VII**, 128-134.
Maki, F. (1997). Sound and figure: concert hall design, Ando and Noson Eds. Music and Concert Hall Acoustics, Conference Proceedings from MCHA 1995. Chaper 1.
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.
Marshall, A. H., Gottlob, D., and Alrutz, H. (1978). Acoustical conditions preferred for ensemble, Journal of the Acoustical Society of America, **64**, 1437-1442.
May, D. N. (1978). Basic subjective responses to noise. Ed. May, D. N., Handbook of noise Assessment. Van Nostrand Reinhol, New York.
Meddis, R., and Hewitt, M. J. (1991a). Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification. The Journal of the Acoustical Society of America **89**, 2866-2882.
Meddis, R., and Hewitt, M. J. (1991b). Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II: Phase sensitivity. Journal of the Acoustical Society of America, **89**, 2883-2894.
Merthayasa, Ide. N., Hemmi, H., and Ando, Y. (1994). Loudness of a 1 kHz pure tone and sharply (1080 dB/Oct.) filtered noises centered on its frequency, Memoirs of Graduate School of Science and Technology, Kobe University **12A**, 147-156.
Molino, J. A. (1979). Annoyance and noise, Ed. Harris, C. M., Handbook of noise control. McGraw-Hill. New York, Chapter 16.
Moore, B.C. J. (1982). Introduction to the Psychology of Hearing, Academic, London, 2nd. ed.
Mosteller, F. (1951). Remarks on the method of paired comparisons. III. Psychometrika, **16**, 207-218.
Mouri, K., Akiyama, K., and Ando, Y., (2000). Relationship between subjective preference and the alpha-brain wave in relation to the initial time delay gap with vocal music. Journal of Sound and Vibration, **232**, 139-147.
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.
Nakajima, T., and Ando, Y. (1997). Calculation and measurement of acoustic factors at each seat in the Kirishima International Concert Hall, Ando and Noson Eds. Music and Concert Hall Acoustics, Conference Proceedings from MCHA 1995. Chap. 5, 39-49.
Nakajima, T., Ando, Y., and Fujita, K. (1992). Lateral low-frequency components of reflected sound from a canopy complex comprising triangular plates in concert halls. Journal of the Acoustical Society of America, **92**, 1443-1451.
Nakayama, I. (1984). Preferred time delay of a single reflection for performers. Acustica **54**, 217-221.
Noson, D., Sato, S., Sakai, H., Ando, Y. (2000). Singer responses to sound fields with a simulated reflection. Journal of Sound and Vibration, **232**, 39-51.
Noson, D., Sato, S., Sakai, H., and Ando, Y. (2002). Melisma singing and preferred stage acoustics for singers, Journal of Sound and Vibration, **258**, 473-485.

Palomaki, K., Tiitinen, H., Makinen, V., May, P., Alku, P., (2002). Cortical processing of speech sounds and their analogues in a spatial auditory environment. Cognitive Brain Research, **14**, 294-299.
Ritsuma, R.J. (1967). Frequencies dominant in the perception of pitch complex sounds. Journal of the Acoustical Society of America **42**, 191-198.
Rose, J.E. (1980). Neural correlates of some psychoacoustical experiences, in Neural Mechanisms of Behavior, Ed. D. McFadden, Springer Verlag, New York. 1-33.
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/
Sakai, H., Singh, P.K., and Ando, Y. (1997). Inter-individual differences in subjective preference judgments of sound fields. Music and Concert Hall Acoustics, Conference Proceedings of MCHA 1995, Eds. Ando, Y., and Noson, D., Academic Press, London, Chapter 13.
Sato, S., and Ando, Y. (1996). Effects of interaural cross-correlation function on subjective attributes. Journal of the Acoustical Society of America,**100** (A), 2592.
Sato, S., Ohta, S., and Ando, Y. (2000). Subjective preference of cellists for the delay time of a single reflection in a performance, Journal of Sound and Vibration, **232**, 27-37.
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., Otori, K., Takizawa, A., Sakai, H., Ando, Y., and Kawamura, H. (2002). Applying genetic algorithms to the optimum design of a concert hall. Journal of Sound and Vibration, **258**, 517-526.
Sato, S., Hayashi, T., Takizawa, A., Tani, A., Kawamura, H., and Ando, Y. (2004). Acoustic design of theatres applying genetic algorithms. Journal of Temporal Design in Architecture and the Environment, **4**, 41-51. http://www.jtdweb.org/journal/
Sato, S. You, J., and Jeon, J. Y. (2007). Sound quality characteristics of refrigerator noise in real living environments with relation to psychoacoustical and autocorrelation function parameters. Journal of the Acoustical Society of America, **122**, 314-325.
Schroeder, M. R. (1966). Architectural Acoustics. Science, **151**, 1355-1359. Seebeck, A. (1844). Über die Definition des Tones. Ann. Phys. Chem., **63**, 353-368.
Secker-Walker, H.E. and Searle, C.L. (1990). Time domain analysis of auditory-nerve-fiber firing rates. Journal of the Acoustical Society of America, **88**, 1427-1436.
Seraphim, H.P. (1961). Ueber die Wahrnehmbarkeit mehrerer Rueckwuerfe von Sprachshall. Acustica, **11**, 80-91.
Singh, P. K., Y. Ando, and Kurihara, Y. (l994). Individual subjective diffuseness responses of filtered noise sound fields. Acustica, **80**, 471-477.
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., Nakagawa, S., Tonoike, M., and Ando, Y., (2003). Spatial analysis of magnetoencephalographic alpha waves in relation to subjective preference of a sound field, Journal of Temporal Design in Architecture and the Environment, **3**, 28-35. http://www.jtdweb.org/journal/
Soeta, Y., Nakagawa, S., Tonoike, M., and Ando, Y. (2004). Magnetoencephalographic responses correspond to individual annoyance of bandpass noise, Journal of Sound and Vibration, **277**, 479-489.
Soeta, Y., Hotehama, T., Nakagawa, S., Tonoike, M., and Ando, Y. (2004). Auditory evoked magnetic fields in relation to interaural cross-correlation of band-pass noise, Hearing Research, **196**, 109-114.
Soeta, Y., Nakagawa, S., and Tonoike, M. (2005). Magnetoencephalographic activities related to the magnitude of the interaural cross-correlation function (IACC) of sound fields. Journal of Temporal Design in Architecture and the Environment, **5**, 5-11. http://www.jtdweb.org/journal/
Soeta, Y., Nakagawa, S., and Matsuoka. K. (2005). Effects of the critical band on auditory evoked magnetic fields, NeuroReport, **16**, 1787-1790.
Soeta, Y., Nakagawa, S., and Matsuoka, K. (2006). The Effect of center frequency and bandwidth on the auditory evoked magnetic field, Hearing Research, **218**, 64-71.
Soeta, Y., and Nakagawa, S. (2006). Auditory evoked magnetic fields in relation to interaural time delay and interaural crosscorrelation, Hearing Research, **220**, 106-115.
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.
Suzumura, Y., Sakurai, M., Ando, Y., Yamamoto, I, and Iizuka, T., and Oowaki, M., (2000). An evaluation of the effects of scattered reflections in a sound field, Journal of Sound and Vibration, **232**, 303-308.
Taguti, T., and Ando, Y. (1997). Characteristics of the short-term autocorrelation function of sound signals in piano performances. Music and Concert Hall Acoustics , Conference Proceedings of MCHA 1995, Eds. Ando, Y., and Noson, D., Academic Press, London, Chapter 23.
Terhardt, E. (1974). Pitch, consonance, and harmony. The Journal of the Acoustical Society of America, **55**, 1061-1069.

Thurstone, L.L. (1927). A law of comparative judgment. Psychological Review, **34**, 273-289.

Torgerson, W.S. (1958). Theory and methods of scaling. Wiley, New York.

van Noorden, L. (1982). Two channel pitch perception, in Music, Mind and Brain, M. Clynes, ed, Plenum: New York. p. 251-269.

Wever, E. G. (1949). Theory of Hearing. Wiley, New York.

Wightman, F.L., (1973a). Pitch and stimulus fine structure. Journal of the Acoustical Society of America, **54**, 397-406.

Wightman, F.L., (1973b). The pattern-transformation model. Journal of the Acoustical Society of America **54**, 407-416.

Yost, W. A. Hill, R., and Perez-Falcon, T. (1978). Pitch and pitch discrimination of broadband signals with rippled power spectra. Journal of the Acoustical Society of America **63**, 1166-1173.

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

Yost, W. A. (1996b). Pitch of iterated noise. Journal of the Acoustical Society of America, **100**, 511-518.

Zwicker, E., Flottorp, G., and Stevens, S.S. (1957). Critical band width in loudness summation. Journal of the Acoustical Society of America, 29, 548-557.

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