Chapter 2
Temporal and Spatial Aspects of Sounds and Sound Fields

Temporal sequences of sounds are transformed into neural activity patterns that propagate through auditory pathways where more centrally-located processors concurrently analyze their form and interpret their meaning and relevance. Thus, a great deal of attention is paid here to analyzing the signal in the time domain. This chapter mainly treats technical aspects of the running autocorrelation function (ACF) of the signal, which contains the envelope and its finer structures as well as the power at the origin of time of the ACF. The ACF has the same information as the power density spectrum of the signal under analysis. From the ACF, however, significant factors may be extracted that are directly related to temporal sensations. The ACF signal representation exists in the auditory pathway, as is discussed in Chapters 4 and 5.

2.1 Analysis of Source Signals

2.1.1 Power Spectrum

As usual, we first discuss signal analysis in the frequency domain, in terms of the power density spectrum of a signal of time domain p(t), which is defined by

$$P_d(\omega) = P(\omega)P^*(\omega), \tag{2.1}$$

where $\omega = 2\pi f$, f is the frequency (Hz) and $P(\omega)$ is the Fourier transform of p(t), given by

$$P(\omega) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} p(t)e^{-j\omega t}dt \tag{2.2}$$

and the asterisk denotes the conjugate.

Y. Ando, P. Cariani (Guest ed.), *Auditory and Visual Sensations*,
DOI 10.1007/b13253_2, © Springer Science+Business Media, LLC 2009

The inverse Fourier transform is the original signal p(t):

$$p(t) = \int_{-\infty}^{+\infty} P(\omega) e^{j\omega t} d\omega \tag{2.3}$$

2.1.2 Autocorrelation Function (ACF)

For our purposes, the most useful signal representation, after the peripheral power spectrum process, is the Autocorrelation Function (ACF) of a source signal, which is defined by

$$\Phi_p(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{+T} p'(t) p'(t+\tau) dt \tag{2.4}$$

where p'(t) = p(t)*s(t), s(t) is the sensitivity of the ear. For practical convenience, s(t) can be chosen as the impulse response of an A-weighted network. It is worth noting that the physical system between the ear entrance and the oval window forms almost the same characteristics as the ear's sensitivity (Ando, 1985, 1998).

The ACF can also be obtained from the power density spectrum, which is defined by Equation (2.4), so that

$$\Phi_p(\tau) = \int_{-\infty}^{+\infty} P_d(\omega) e^{j\omega t} dt \tag{2.5}$$

$$P_d(\omega) = \int_{-\infty}^{+\infty} \Phi_p(\tau) e^{-j\omega\tau} d\tau \tag{2.6}$$

Thus, the ACF and the power density spectrum mathematically contain the same information. The normalized ACF is defined by

$$\phi_p(\tau) = \Phi_p(\tau) / \Phi_p(0) \tag{2.7}$$

There are four significant items that can be extracted from the ACF:

(1) Energy represented at the origin of the delay, $\Phi_p(0)$.
(2) As shown in Fig. 2.1, the width of amplitude $\phi(\tau)$, around the origin of the delay time defined at a value of 0.5, is $W_{\phi(0)}$, according to the fact that $\phi(\tau)$ is en even function.
(3) Fine structure, including peaks and delays: For instance, τ_1 and ϕ_1 are the delay time and the amplitude of the first peak of the ACF, τ_n and ϕ_n being the delay time and the amplitude of the n-th peak. Usually, there are certain correlations between τ_1 and τ_{n+1} and between ϕ_1 and ϕ_{n+1}, so that significant factors are τ_1 and ϕ_1.
(4) The effective duration of the envelope of the normalized ACF, τ_e, which is defined by the tenth-percentile delay and which represents a repetitive feature or reverberation containing the sound source itself (Fig. 2.2).

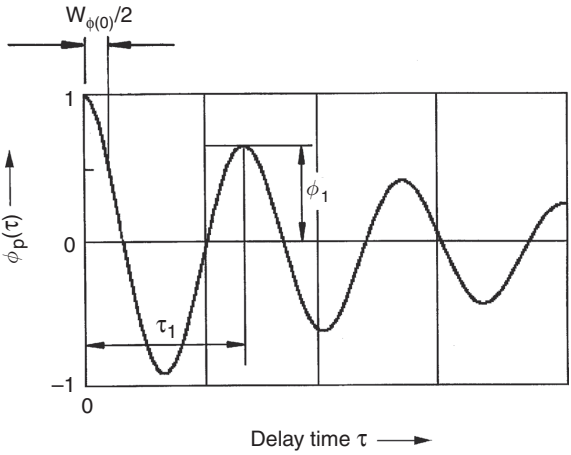

Fig. 2.1 Definition of temporal factors, τ_1 and ϕ_1, as features of the normalized autocorrelation function (NACF). τ_1 is the time delay associated with the first ACF peak. ϕ_1 is the relative amplitude of the first peak

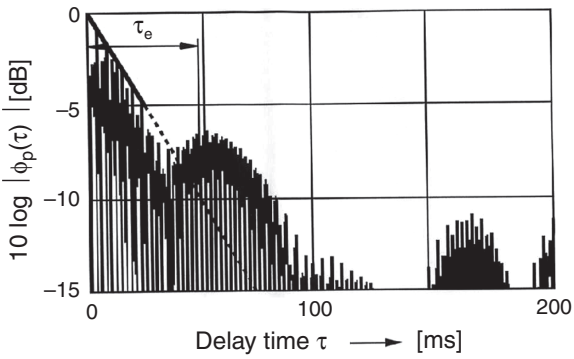

Fig. 2.2 Determination of the effective duration of running ACF, τ_e. Effective duration of the normalized ACF is defined by the delay τ_e at which the envelope of the normalized ACF becomes 0.1

The autocorrelation function (ACF) of any sinusoid (pure tone) having any phase is a zero-phase cosine of the same frequency. Since its waveform and ACF envelope is flat, with a slope of zero, its effective duration τ_e is infinite. The ACF of white noise with infinite bandwidth is the Dirac delta function $\delta(\tau)$ which has an infinite slope. This means that the signal has an effective duration that approaches zero (no temporal coherence). As the bandwidth of the noise decreases, the effective duration (signal coherence) increases.

Table 2.1 lists three music and speech sources that were used extensively in many of our experiments. Motif A is the slow and sombre Royal Pavane by Orlando Gibbons (1583–1625). Motif B is the fast and playful final movement of Sinfonietta by Malcolm Arnold (1921–2006). Speech *S* is a poem by Japanese novelist and poet Doppo Kunikida (1871–1908). Examples of normalized ACFs (2T = 35 s) for the two extremes of slow and fast music are shown in Fig. 2.3.

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, 1977), and the minimum value of running ACF, $(\tau_e)_{min}$ (Ando, 1998)

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; Movement IV	Malcolm Arnold	43 (35)	40
Speech S	Poem read by a female	Doppo Kunikida	10 (12)	

[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 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 analysis $2T = 2$ s, with a running interval of 100 ms (see Section 5.3 for a recommended 2T). Subjective judgments may be made at the most active piece of sound signals.

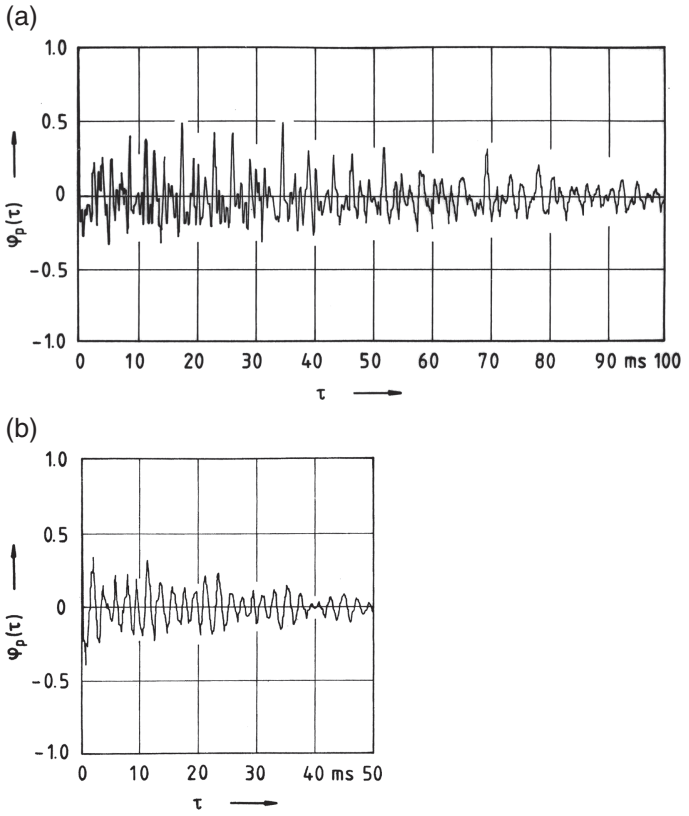

Fig. 2.3 Examples of the long-time ACF ($2T = 35$ s) analyzed in the early stage of systematical investigations (Ando, 1977). (**a**) Music motif A: *Royal Pavane*, $\tau_e = 127$ ms. (**b**) Music motif B: *Sinfonietta*, Opus 48, Movement III, allegro con brio, $\tau_e = 43$ ms. Note that, according to the different characteristics of the loudspeakers used in the subjective judgments, the effective duration of ACF may differ slightly, for example, $\tau_e = 35$ ms (music motif B)

Loundness also has ACF correlates. When p'(t) is measured in reference to the pressure 20 μPa leading to the level L(t) , the sound-pressure level Leq is defined by

$$L_{eq} = 10 \log \frac{1}{T} \int_0^T 10^{\frac{L(t)}{10}} dt \tag{2.8}$$

corresponding to

$$10 \log \Phi_p(0) \tag{2.9}$$

Although SPL is an important factor related to loudness, it is not the whole story. The envelope of the normalized ACF is also related to important subjective attributes, as is detailed later.

2.1.3 Running Autocorrelation

Because a certain degree of coherence exists in the time sequence of the source signal, which may greatly influence subjective attributes of the sound field, use is made here of the short ACF as well as the long-time ACF.

The short-time moving ACF as a function of the time τ is calculated as (Taguti, and Ando, 1997)

$$\begin{aligned} \phi_p(\tau) &= \phi_p(\tau;t,T) \\ &= \frac{\Phi_p(\tau;t,T)}{[\Phi_p(0;t,T)\Phi_p(0;\tau + t,T)]^{1/2}} \end{aligned} \tag{2.10}$$

where

$$\Phi_p(\tau;t,T) = \frac{1}{2T} \int_{t-T}^{t+T} p'(s)p'(s + \tau)ds \tag{2.11}$$

The normalized ACF satisfies the condition that $\phi_p(0) = 1$.

To demonstrate a procedure for obtaining the effective duration of the anaylzed short-time ACF analyzed, Fig. 2.3 shows the absolute value in the logarithmic form as a function of the delay time. The envelope decay of the initial and important part of the ACF may be fitted by a straight line in most cases. The effective duration of the ACF, defined by the delay τ_e at which the envelope of the ACF becomes –10 dB (or 0.1; the tenth-percentile delay), can be easily obtained by the decay rate extrapolated in the range from 0 dB at the origin to –5 dB, for example.

The effective duration of the ACF for various signal durations, 2T, with the moving interval are obtained in such a way. Examples of analyzing the moving ACF of Japanese *Syaku-hachi* music (Kare-Sansui composed by Hozan Yamamoto, which includes extremely dynamic movements with *Ma* and *Fusi*) are shown in Fig. 2.4a–f. The signal duration to be analyzed is discussed in Section 5.3.

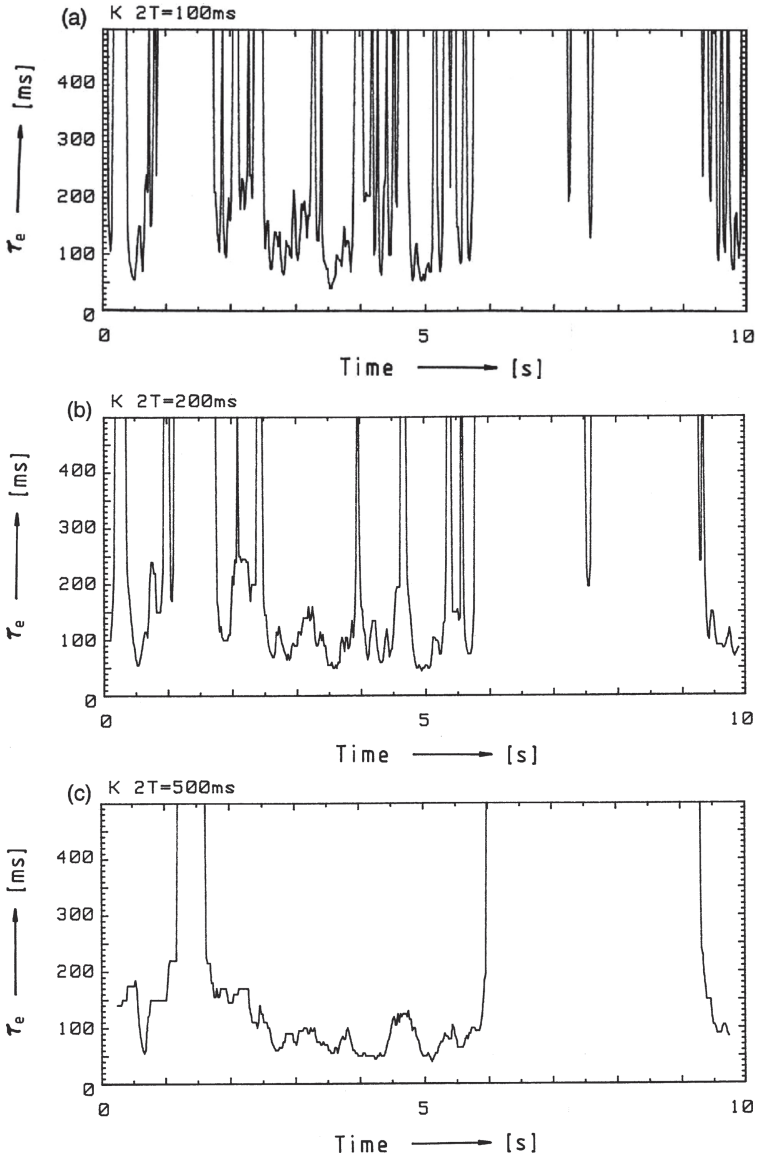

Fig. 2.4 Examples of measured effective durations of a 10 s segment of dynamic, Japanese *Syaku-hachi* music (Music motif K, Kare-Sansui, Yamamoto) computed from its running autocorrelation using different temporal integration windows (2*T*). Temporal stepsize was 100 ms. (**a**) 2*T*=100 ms. (**b**) 2*T*=200 ms. (**c**) 2*T*=500 ms. (**d**) 2*T*=1 s. (**e**) 2*T*=2 s. (**f**) 2*T*=5 s

Figure 2.5a–f show the moving τ_e for music motifs A and B, 2T = 2.0 s and 5.0 s. The recommended signal duration $(2T)_r$ is discussed in Section 5.3. The minimum value of a moving τ_e, the most active part of music, containing important informa-

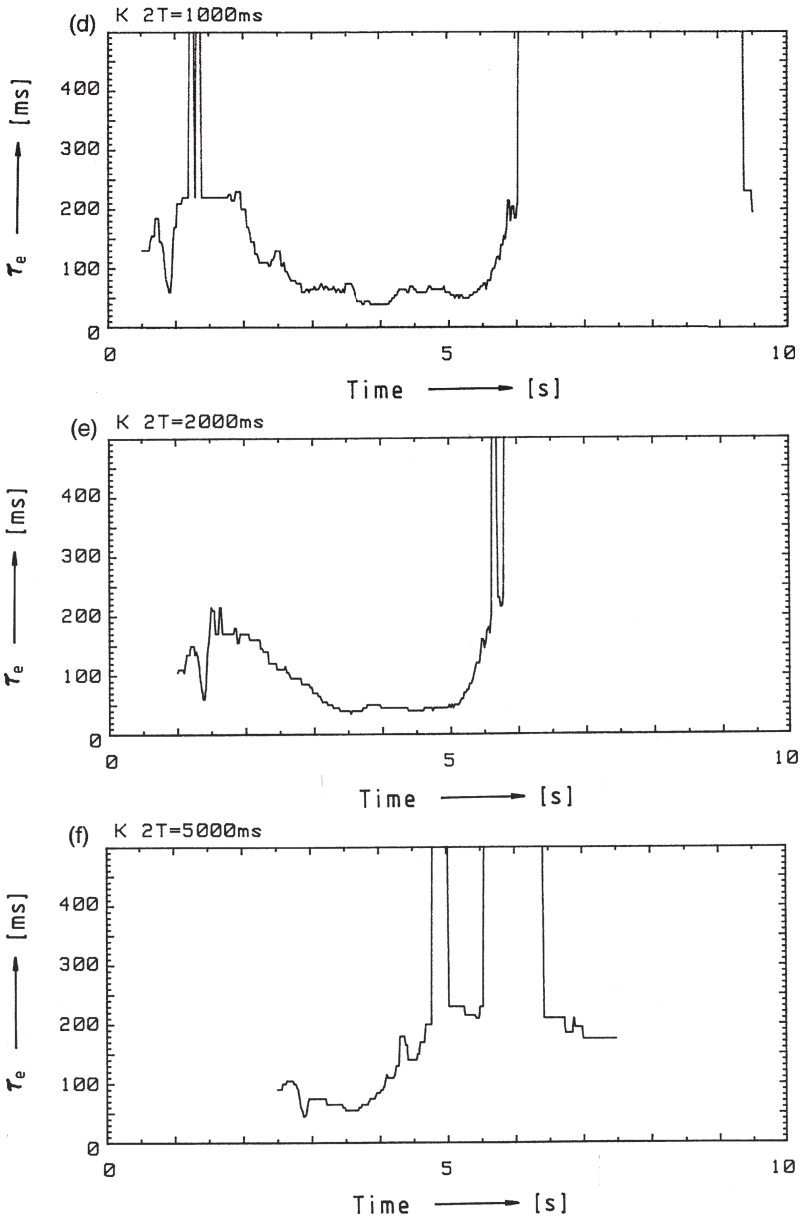

Fig. 2.4 (continued)

tion and influencing subjective preferece, is discussed in Chapter 3. The value of $(\tau_e)_{min}$ is plotted in Fig. 2.6 as a function of 2T. It is worth noting that stable values of $(\tau_e)_{min}$ may be obtained in the range of 2T = 0.5 to 2.0 s for these extreme

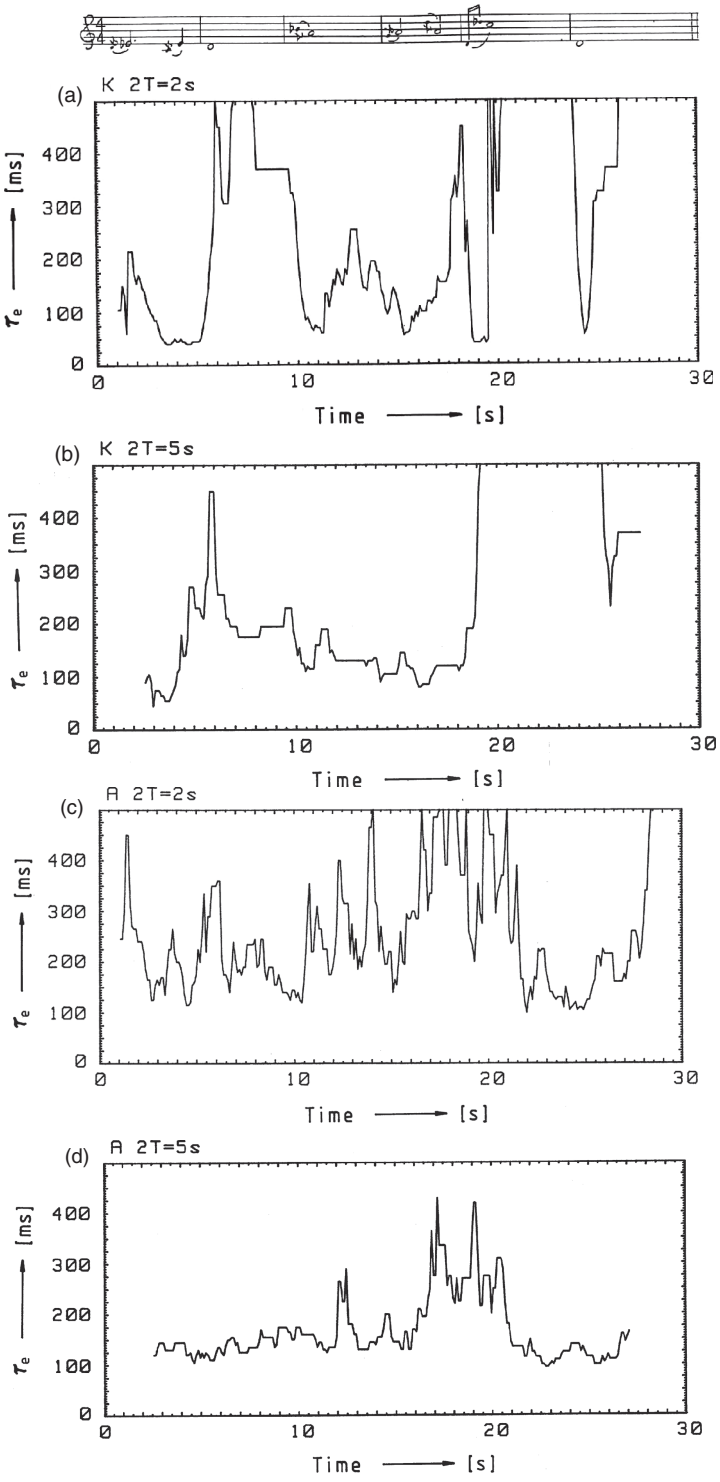

Fig. 2.5 (continued)

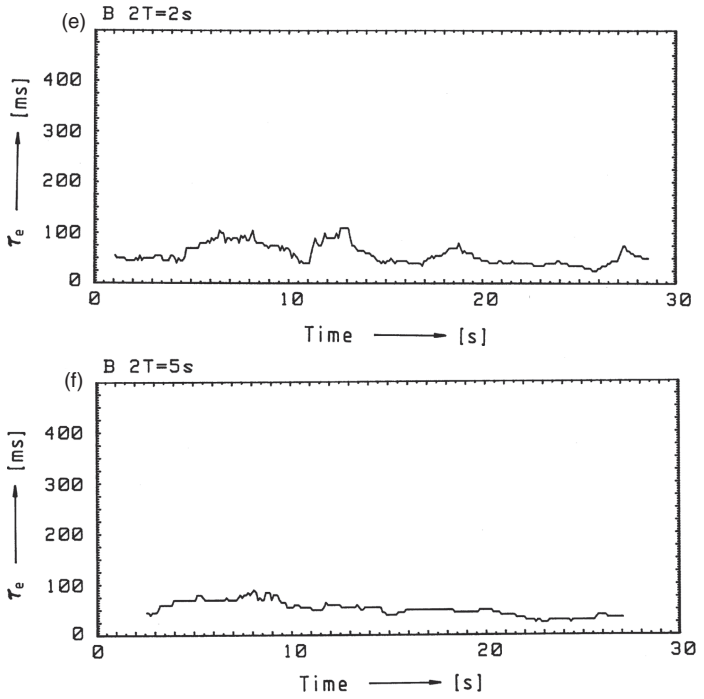

Fig. 2.5 Examples of measured running effective durations for two musical pieces and a spoke poem. Temporal stepsize was 100 ms with temporal integration time $2T = 2$ or 5 s. (**a**) and (**b**) Slow musical piece, motif K, Kare-Sansui, (Yamamoto). (**c**) and (**d**) Slow musical piece, motif A (Gibbons) (**e**) and (**f**) Fast musical piece, motif B (Arnold)

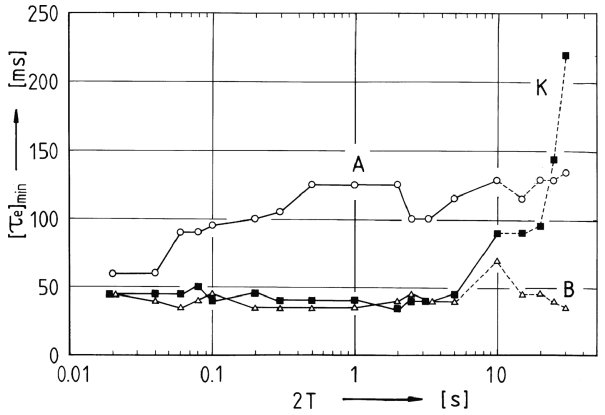

Fig. 2.6 Minimum values of the running effective duration as a function of temporal integration time 2*T*. (*circles*, curve A) Slow musical piece, motif A (Gibbons). (*triangles*, curve B) Fast musical piece, motif B (Arnold). (*filled squares*, curve K) Kare-Sansui, music motif K (Yamamoto)

music motifs. In general, a recommended integration interval of the signal shall be discussed in Section 5.3 as a temporal window.

2.2 Physical Factors of Sound Fields

2.2.1 Sound Transmission from a Point Source through a Room to the Listener

Let us consider the sound transmission from a point source in a free field to the two ear canal entrances. Let p(t) be the source signal as a function of time, t, at its source point, and $g_l(t)$, and $g_r(t)$ be the impulse responses between the source point r_0 and the two ear entrances. Then the sound signals arriving at the entrances are expressed by,

$$\begin{aligned} f_l(t) &= p(t) * g_l(t) \\ f_r(t) &= p(t) * g_r(t) \end{aligned} \tag{2.12}$$

where the asterisk denotes convolution.

The impulse responses $g_{l,r}(t)$ include the direct sound and reflections $w_n(t - \Delta t_n)$ in the room as well as the head-related impulse responses $h_{nl,r}(t)$, so that,

$$g_{l,r}(t) = \sum_{n=0}^{\infty} A_n w_n(t-\Delta t_n) * h_{nl,r}(t) \tag{2.13}$$

where n denotes the number of reflections with a horizontal angle, ξ_n and elevation η_n; $n = 0$ signifies the direct sound ($\xi_0 = 0$, $y_0 = 0$),

$$A_0 W_0(t - \Delta t_0) = \delta(t),\ \Delta t_0 = 0,\ A_0 = 1,$$

where $\delta(t)$ is the Dirac delta function; A_n is the pressure amplitude of the n-th reflection $n > 0$; $w_n(t)$ is the impulse response of the walls for each path of reflection arriving at the listener, Δt_n being the delay time of reflection relative to that of the direct sound; and $h_{nl,r}(t)$ are impulse responses for diffraction of the head and pinnae for the single sound direction of n. Therefore, Equation (2.12) becomes

$$f_{l,r}(t) = \sum_{n=0}^{\infty} p(t)^* A_n w_n(t-\Delta t_n) * h_{nl,r}(t) \tag{2.14}$$

When the source has a certain directivity, the p(t) is replaced by $p_n(t)$.

2.2.2 Temporal-Monaural Factors

As far as the auditory system is concerned, all factors influencing any subjective attribute must be included in the sound pressures at the ear entrances; these are expressed by Equation (2.14). The first important item, which depends on the source program, is the sound signal p(t). This is represented by the ACF defined by Equation (2.4). The ACF is factored into the energy of the sound signal $\Phi_p(0)$ and the normalized ACF as expressed by Equations (2.4)–(2.6).

The second term is the set of impulse responses of the reflecting walls, $A_n w_n(t - \Delta t_n)$. The amplitudes of reflection relative to that of the direct sound, A_1, A_2, ..., are determined by the pressure decay due to the paths d_n, such that

$$A_n = d_0/d_n \tag{2.15}$$

where d_0 is the distance between the source point and the center of the listener's head. The impulse responses of reflections to the listener are $w_n(t - \Delta t_n)$ with the delay times of Δt_1, Δt_2, ... relative to that of the direct sound, which are given by

$$\Delta t_n = (d_n - d_0)/c \tag{2.16}$$

where c is the velocity of sound (m/s). These parameters are not physically independent, in fact, the values of A_n are directly related to Δt_n in the manner of

$$\Delta t_n = d_0(1/A_n - 1)/c \tag{2.17}$$

In addition, the initial time delay gap between the direct sound and the first reflection Δt_1 is statistically related to Δt_2, Δt_3... and depends on the dimensions of the room. In fact the echo density is proportional to the square of the time delay (Kuttruff, 1991). Thus, the initial time delay gap Δt_1 is regarded as a representation of both sets of Δt_n and A_n (n = 1, 2...).

Another item is the set of the impulse responses of the n-th reflection, $w_n(t)$ being expressed by

$$w_n(t) = w_n(t)^{(1)} * w_n(t)^{(2)} * ... * w_n(t)^{(i)} \tag{2.18}$$

where $w_n(t)^{(i)}$ is the impulse response of the i-th wall in the path of the n-th reflection from the source to the listener. Such a set of impulse responses may be represented by a statistical decay rate, namely the subsequent reverberation time, T_{sub}, because $w_n(t)^{(i)}$ includes the absorption coefficient as a function of frequency. This coefficient is given by

$$\alpha_n(\omega)^{(i)} = 1 - |W_n(\omega)^{(i)}|^2 \tag{2.19}$$

According to Sabine's formula (1900), the subsequent reverberation time is approximately calculated by

$$T_{sub} \approx \frac{KV}{\overline{\alpha} S} \tag{2.20}$$

where K is a constant (about 0.162), V is the volume of the room (m^3), S is the total surface (m^2), and $\overline{\alpha}$ is the average absorption coefficient of the walls. The denominator of Equation (2.22) can be calculated more precisely as a function of the frequency by taking into account specific values of absorption coefficient as a function of frequency $\alpha(\omega)_i$ and surface area S_i for each room surface i:

$$\overline{\alpha} S(\omega) = \sum_i \alpha(\omega)_i S_i \tag{2.21}$$

where $\omega = 2\pi f$, f is the frequency.

2.2.3 Spatial-Binaural Factors

Two sets of head-related impulse responses for two ears $h_{nl,r}(t)$ constitute the remaining objective item. These two responses $h_{nl}(t)$ and $h_{nr}(t)$ play an important role in sound localization and spatial impression but are not mutually independent objective factors. For example, $h_{nl}(t) \sim h_{nr}(t)$ for the sound signals in the median plane, and there are certain relations between them for any other directions to a listener. In addition, the interaural time deference (IATD) and the interaural level difference (IALD) are not mutually independent factors of sound fields. In fact, a certain relationship between the IATD and the IALD can be expressed for a single directional sound arriving at a listener for a given source signal and thus for any sound field with multiple reflections. A particular example is that, when the IATD is zero, then the IALD is nearly zero as well.

Therefore, to represent the interdependence between two impulse responses, a single factor may be introduced, that is, the interaural crosscorrelation function (IACF) between the sound signals at both ears $f_l(t)$ and $f_r(t)$, which is defined by

$$\Phi_{lr}(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{+T} f'_l(t) f'_r(t+\tau) dt, \quad |\tau| \leq 1\,\text{ms} \tag{2.22}$$

where $f'_l(t)$ and $f'_r(t)$ are obtained by signals $f_{l,r}(t)$ after passing through the A-weighted network, which corresponds to the ear's sensitivity, s(t). It has been shown that ear sensitivity may be characterized by the physical ear system including the external and the middle ear (Ando, 1985, 1998).

The normalized interaural crosscorrelation function is defined by

$$\phi_{lr}(\tau) = \frac{\Phi_{lr}(\tau)}{\sqrt{\Phi_{ll}(0)\Phi_{rr}(0)}} \tag{2.23}$$

where $\Phi_{ll}(0)$ and $\Phi_{rr}(0)$ are the ACFs at $\tau = 0$ for the left and right ears, respectively, or the sound energies arriving at both ears, and τ the interaural time delay possibly within plus and minus 1 ms. Also, from the denominator of Equation (2.23), we obtain the binaural listening level (LL) such that,

$$\mathrm{LL} = 10\log[\Phi(0)/\Phi(0)_{\mathrm{reference}}] \tag{2.24}$$

where $\Phi(0) = [\Phi_{ll}(0)\ \Phi_{rr}(0)]^{1/2}$, which is the geometrical mean of the sound energies arriving at the two ears, and $\Phi(0)_{\mathrm{reference}}$ is the reference sound energy.

If discrete reflections arrive after the direct sound, then the normalized interaural crosscorrelation is expressed by,

$$\Phi_{1r}^{(N)}(\tau) = \frac{\sum_{n=0}^{N} A^2 \Phi_{1r}^{(n)}(\tau)}{\sqrt{\sum_{n=0}^{N} A^2 \Phi_{11}^{(n)}(0) \sum_{n=0}^{N} A^2 \Phi_{rr}^{(n)}(0)}} \tag{2.25}$$

where we put $w_n(t) = \delta(t)$ for the sake of convenience, and $\Phi_{lr}^{(n)}(\tau)$ is the interaural crosscorrelation of the n-th reflection, $\Phi_{ll}(\tau)^{(n)}$ and $\Phi_{rr}(\tau)^{(n)}$ are the respective sound energies arriving at the two ears from the n-th reflection. The denominator of Equation (2.25) corresponds to the geometric mean of the sound energies at the two ears.

The magnitude of the interaural crosscorrelation is defined by

$$\mathrm{IACC} = |\phi_{1r}(\tau)|_{\max} \tag{2.26}$$

for the possible maximum interaural time delay, say,

$$|\tau| \leq 1\ \mathrm{ms}$$

For several music motifs, the long-time IACF ($2T = 35$ s) was measured for each single reflected sound direction arriving at a dummy head (Table D.1, Ando, 1985). These data may be used for the calculation of the IACF by Equation (2.25).

For example, measured values of the IACF using music motifs A and B are shown in Fig. 2.7.

The interaural delay time, at which the IACC is defined as shown in Fig. 2.8, is the τ_{IACC}. Thus, both the IACC and τ_{IACC} may be obtained at the maximum value of IACF.

For a single source signal arriving from the horizontal angle ξ defined by τ_ξ, the interaural time delay corresponds to τ_{IACC}. When it is observed $\tau_{\mathrm{IACC}} = 0$ in a room, then usually a frontal sound image and a well-balanced sound field are perceived (the preferred condition).

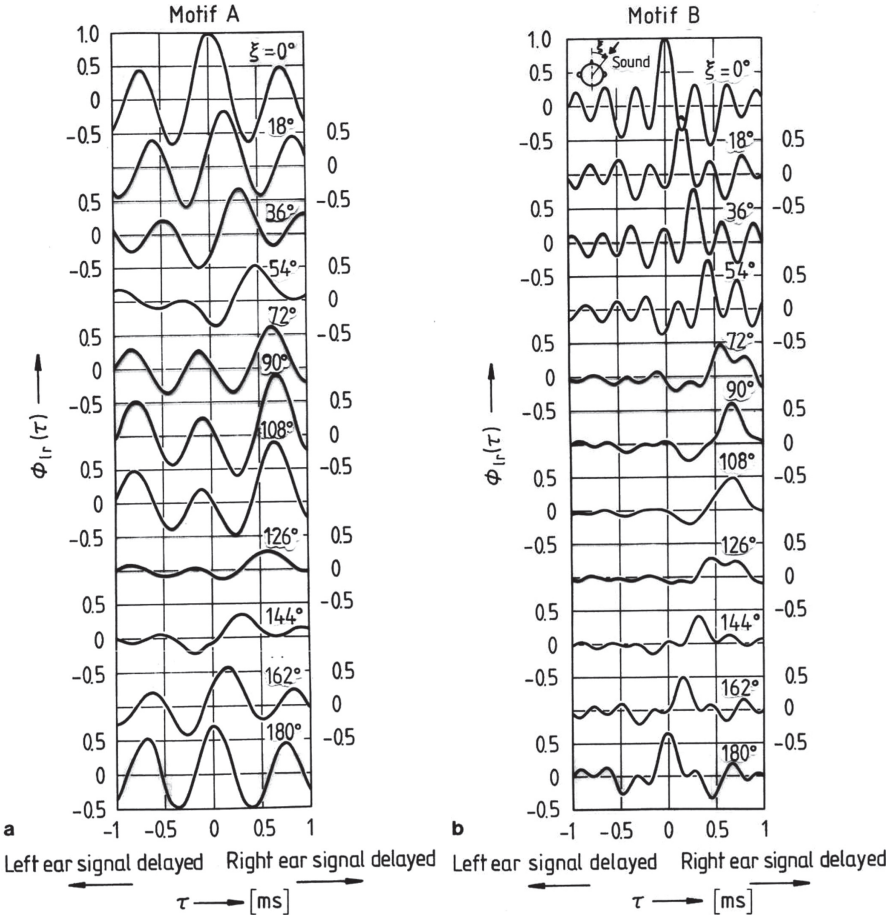

Fig. 2.7 Interaural crosscorrelation function IACF for a single sound as a function of angle of incidence measured at the ear entrances of a dummy head. *Left*: music motif A (Gibbons). *Right*: music motif B (Arnold)

The width of the IACF, defined by the interval of delay time at a value of δ below the IACC, corresponding to the just-noticeable-difference (JND) of the IACC, is given by the W_{IACC} (Fig. 2.8). Thus, the apparent source width (ASW) may be perceived as a directional range corresponding mainly with the W_{IACC}. A well-defined directional impression corresponding to the interaural time delay τ_{IACC} is perceived when listening to a sound with a sharp peak in the IACF with a small value of W_{IACC}. On the other hand, when listening to a sound field with a low value for the IACC <0.15, then a subjectively diffuse sound is perceived (Damaske and Ando, 1972).

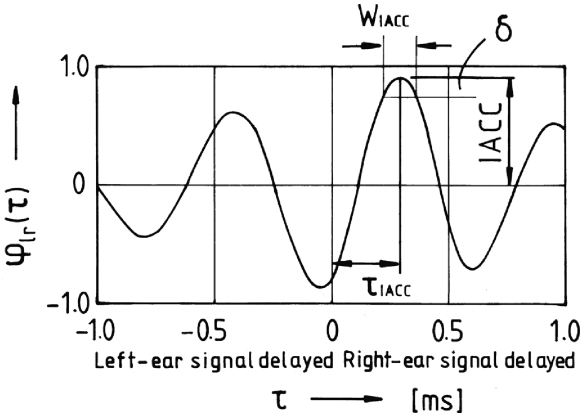

Fig. 2.8 Definition of the three spatial factors extracted from the interaural correlation function (IACF). The interaural correlation magnitude IACC is the maximum value of the IACF. IACC is associated with the subjective diffuseness of the sound. The τ_{IACC} is the delay at which the IACF attains its maximum value of IACC. This interaural delay is associated with sound direction in the horizontal plane. W_{IACC} is the width of the maximal IACF peak, defined by the size of the delay range over which the IACF peak is at least 90% of its maximal value ($\delta = 0.1$*IACC). W_{IACC} is associated with the apparent source width of the sound

These four factors, LL, IACC, τ_{IACC}, and W_{IACC}, are independently related to subjective preference (Chapter 3) and spatial sensations such as subjective diffuseness and the ASW (Chapter 7).

2.3 Simulation of a Sound Field in an Anechoic Enclosure

According to Equation (2.14), one can effectively replicate the sound field in a given enclosure by taking the directional information of the sound source and its reflections into consideration. One can approximately reproduce the perception of a sound field using four signals that are generated by different sound paths: the direct sound, two early reflections, and diffused reverberation.

An example of the block diagram of the simulation system for the is shown in Fig. 2.9 (Ando et al., 1973). The sound source without reverberation is processed to generate the four sound-path signals by adjusting the relative gains of the signals ($A_0 \ldots A_3$), applying pinna-related filters ($w_1(t)$ and ($w_2(t)$) that take into account the directionality of early reflections and adding corresponding delays ($t_0 \ldots \Delta t_3$). The four signals are played back into an anechoic chamber via seven speakers. One speaker directly in front of the listener carries the direct signal, two lateral–frontal speakers carry the right and left early reflections, and the remaining four speakers situated around the listener carry the diffused reverberations. Additional gains and delays regulate the front-rear balance and temporal coherence of the reverberatory signals.

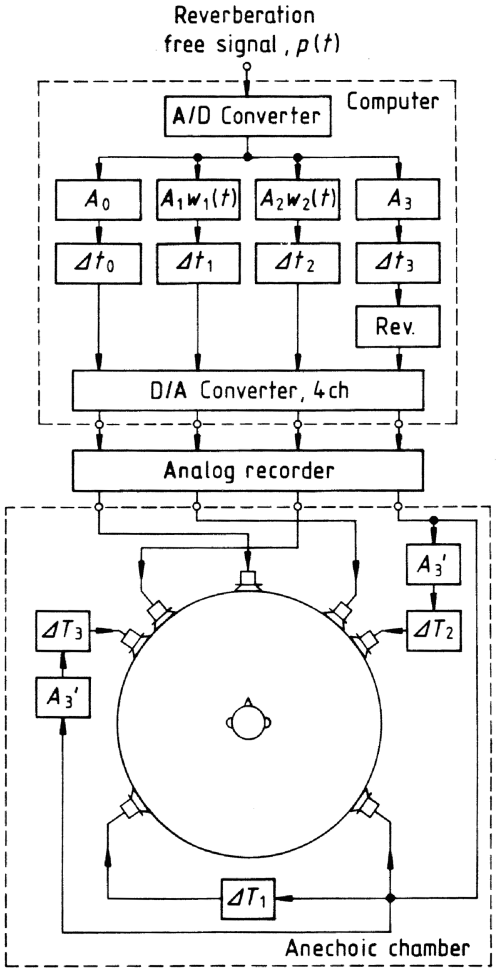

Fig. 2.9 An orthodox system for simulating sound fields using a seven channel playback system and an anechoic chamber. The sound source in the absence of reverberations is recorded and added back to itself after a series of delays and gains are imposed. The frontal speaker simulates the direct path, the two front-lateral speakers simulate early reflections from the stage and walls, and the remaining four rearmost speakers simulate longer reverberations

This kind of setup was used in all subjective judgment experiments and in recording the electrophysiological responses described in this volume. In situations where one seeks to produce more diffuse sound images and correspondingly small IACC values, the directions of the four loudspeakers that convey subsequent reverberations (the Rev. signal path in the figure) are chosen to be well away from the median plane. Here the incoherent reverberation signals supplied to the four rearmost loudspeakers are additionally delayed by only short relative durations, ΔT_j ($j = 1, 2, 3$).

http://www.springer.com/978-1-4419-0171-2

Auditory and Visual Sensations
Ando, Y.
2010, XXV, 344 p., Hardcover
ISBN: 978-1-4419-0171-2