

Chapter 2

Analyses of Temporal Factors of a Source Signal

Sound signals proceed along auditory pathways and are perceived in a time sequence while the brain simultaneously interprets the meaning of signals. Thus, a great deal of attention is paid here to analyzing the signal in the time domain. This chapter mainly treats physical aspects of the running autocorrelation function (ACF) of the signal, which contains the envelope and its finer structures as well as the power at its starting time. Mathematically, the ACF has the same information as the power density spectrum of the signal under analysis. From the ACF, however, significant factors may be easily extracted, which are directly related to temporal percepts (such as four temporal primary sensations, i.e., loudness, pitch, timbre and duration are well described by the temporal factors extracted from the running autocorrelation function of sound signal). The ACF processor exists in the auditory pathway not at the very periphery but close to the brain as discussed in Chap. 5, so that the any psychological responses are affected directly by these factors. And, the running inter-aural crosscorrelation function (IACF) processor exists in the auditory pathway around inferior colliculus. The spatial factors may well describe the spatial percepts (localization or direction of sound signal arriving at a listener position, movement of a sound source on the stage (Sect. 2.1.4), apparent source width (ASW) and subjective diffuseness) associated with the right hemisphere.

2.1 Analyses of a Source Signal

2.1.1 Autocorrelation Function (ACF) of a Sound Source

The most promising signal process, in the auditory system after a rough peripheral power spectrum process, is the ACF, which is defined by

$$\Phi_p(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{+T} p'(t)p'(t + \tau)dt \quad (2.1)$$

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

The normalized ACF is defined by

$$\phi_p(\tau) = \Phi_p(\tau) / \Phi_p(0) \quad (2.2)$$

Thus, $\phi_p(0) = 1$.

2.1.2 Running ACF

The short-time moving ACF or the running ACF as shown in Fig. 2.1 is calculated as (Taguti and Ando 1997).

$$\begin{aligned} \phi_p(\tau) &= \Phi_p(\tau; t, T) \\ &= \frac{\Phi_p(\tau; t, T)}{[\Phi_p(0; t, T)\Phi_p(0; \tau + t, T)]^{1/2}} \end{aligned} \quad (2.3)$$

where

$$\Phi_p(\tau; t, T) = \frac{1}{2T} \int_{t-T}^{t+T} p'(s)p'(s + \tau)ds \quad (2.4)$$

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

2.1.3 Analyses of the Running ACF

In order to avoid confusion in the analyses of the running ACFs, five different signal durations analyzed are illustrated in Fig. 2.2. Resulting ACFs and the power spectra obtained by different signal durations are shown in Fig. 2.3. The direct method is obtained in the time domain. The ACF obtained by FFT also, based on the Wiener–Khinchine theorem, is acquired by a transform in the frequency domain by FFT, followed by performing an inverse FFT calculation. It is important to note that the Wiener–Khinchine theorem is mathematically satisfied only for completely periodic or infinite-length signals, but not mathematically be satisfied for a finite duration of sound signals. A variation in both ACF and power spectrum due to the different signal duration is evident (see Fig. 2.3a–t). It is not possible to find even one matched

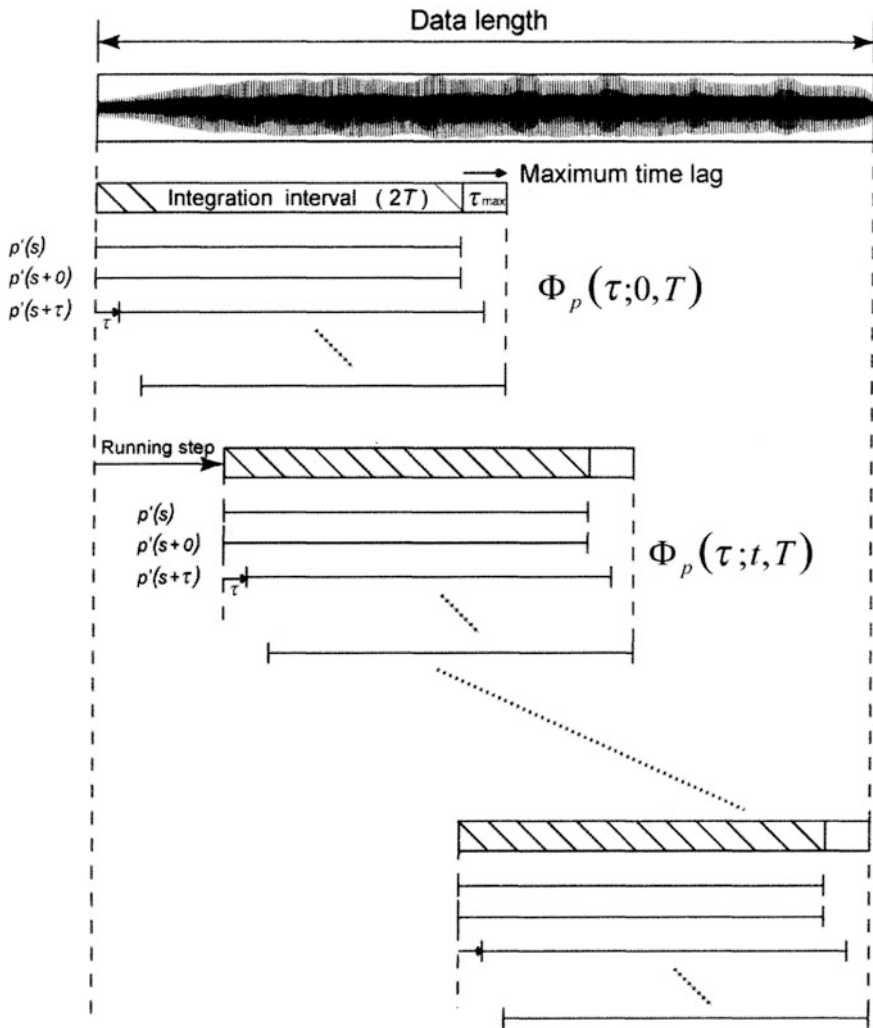


Fig. 2.1 Direct method of analyzing the running autocorrelation function (ACF) in the time domain (Kato et al. 2007)

pair of the running ACF and running power spectrum for quasi-periodic signals. Thus, we reiterate that the transform methods and their precise definitions should be carefully determined before conducting an analysis of signals.

Although “FFT method A” or “FFT method B (method to avoid circular calculation)” is usually used for the purpose of the fast computation and is accompanied by a window function such as Hamming, Hanning, or Blackman, in order to obtain the ACF corresponding to the direct method, “FFT method C” (see Fig. 2.3e) must be used. If “FFT method C” may be chosen instead of the “direct method” for

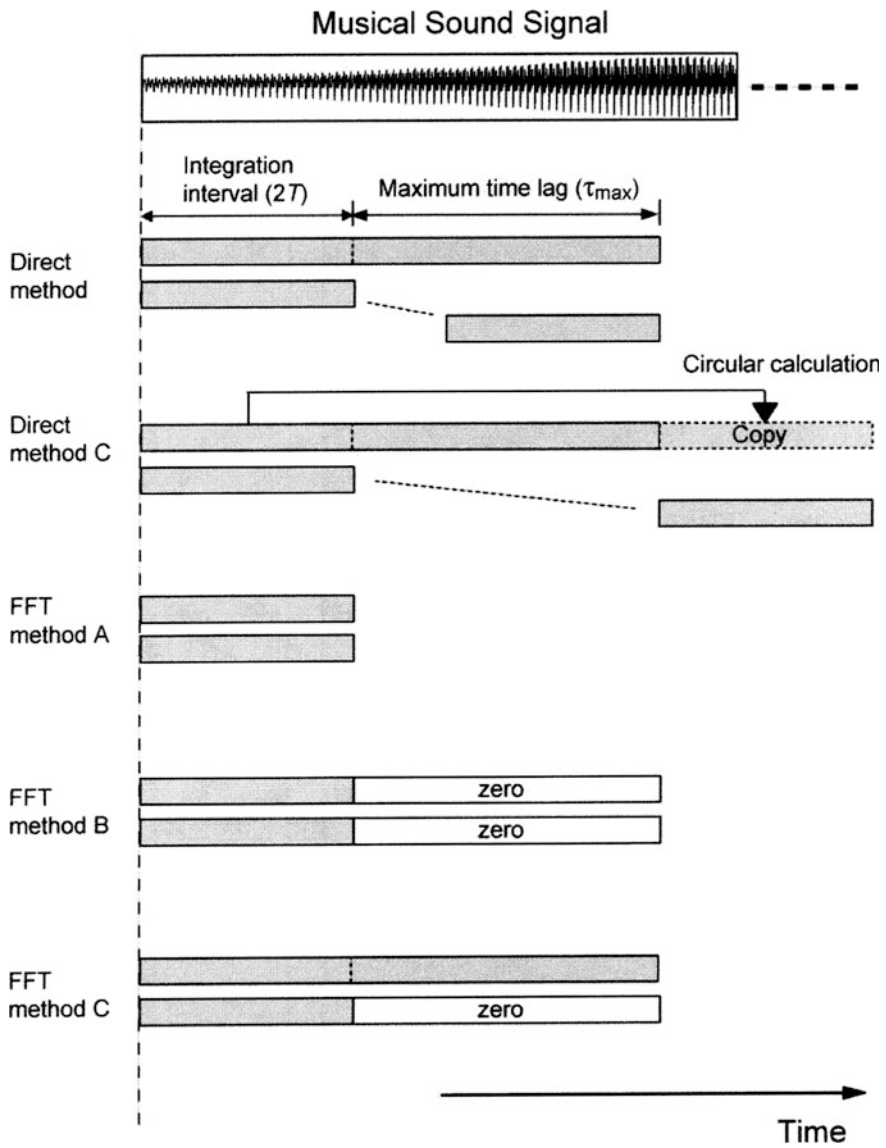


Fig. 2.2 Five different signal durations analyzed for the ACF (Kato et al. 2007)

performing a fast calculation, the segment over the maximum time lag should be deleted because of the circular calculation. Compare the result of direct method with that of FFT method C (Fig. 2.3a, e).

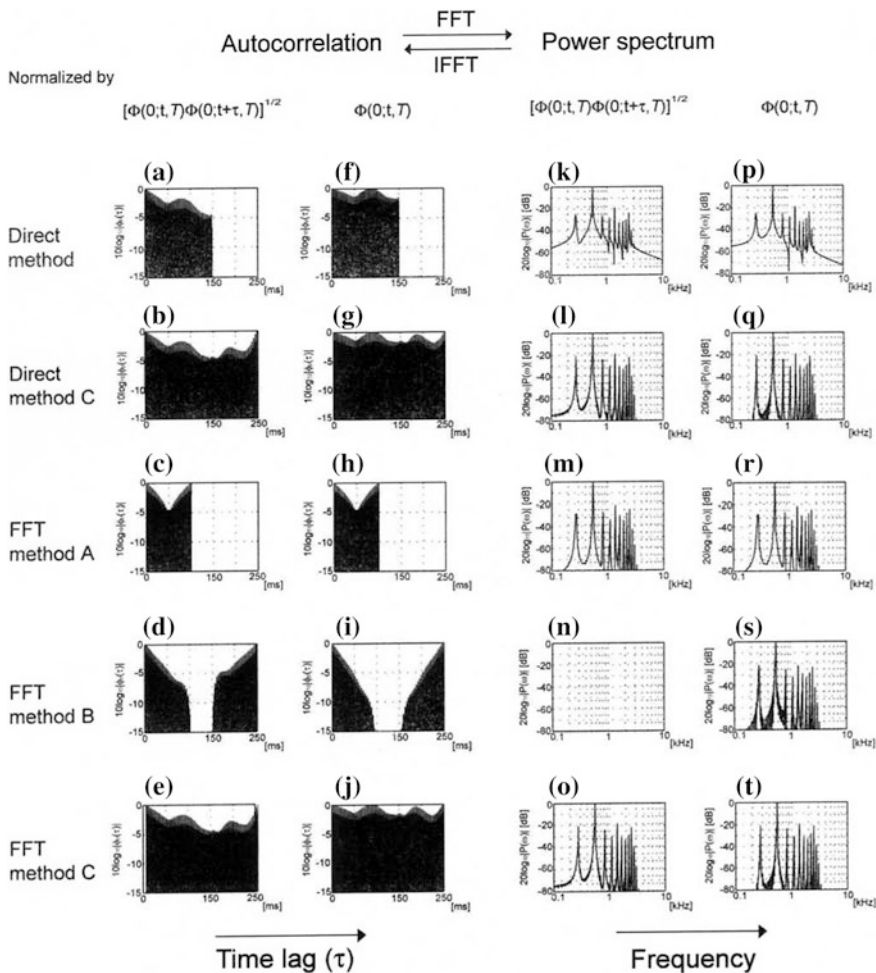


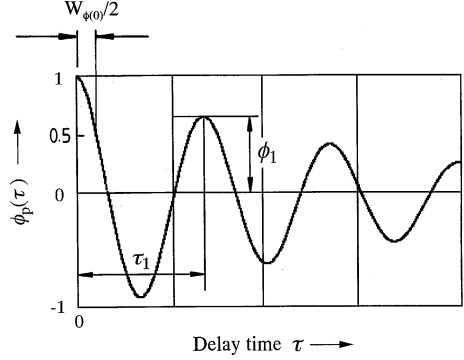
Fig. 2.3 Comparisons of the ACF and its power spectrum obtained by five different signal durations shown in Fig. 2.2. FFT methods *A* and *B* may not obtain the right ACF. FFT method *C* may be recommended for analyzing the ACF up to the maximum delay time, τ_{\max} (Kato et al. 2007)

2.1.4 Temporal Factors Extracted from the Running ACF

There are significant temporal factors influencing subjective responses that can be extracted from the running ACF (Fig. 2.4):

- (1) Energy represented at the origin of the delay, $\Phi_p(0)$;
- (2) 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, respectively, τ_n and

Fig. 2.4 Definition of three temporal factors extracted from the initial part of ACF



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 only τ_1 and ϕ_1 ;

- (3) Widths of the amplitudes of $\phi_p(0)$ defined by $W_{\phi(0)}$.
- (4) The effective duration of the envelope of the ACF, τ_e , which is defined by the 10 % delay and which represents a repetitive feature or reverberation containing the sound source itself.

When $p'(t)$ is measured in reference to the reference pressure leading to the envelope level $L(t)$ in dB, the equivalent sound pressure level L_{eq} , is defined by

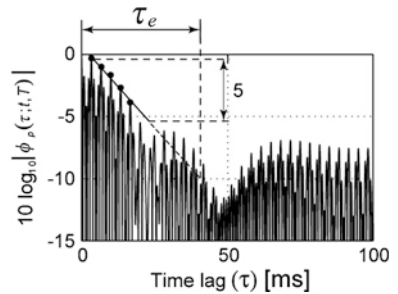
$$L_{eq} = 10 \log \frac{1}{T} \int_0^T 10 \frac{L(t)}{10} dt, \quad (2.5)$$

This corresponds to $10 \log \Phi_p(0)$.

While this is an important factor significantly related to loudness, it is not the whole story. The value of τ_e , which is a repetitive feature of sound signals, for example, is related to loudness and other subjective attributes, as is detailed later (Fig. 5.10).

In order to demonstrate a procedure for obtaining the effective duration of the analyzed short-time ACF analyzed, Fig. 2.5 shows the absolute value in the

Fig. 2.5 Determination of the effective duration extracted from the running ACF



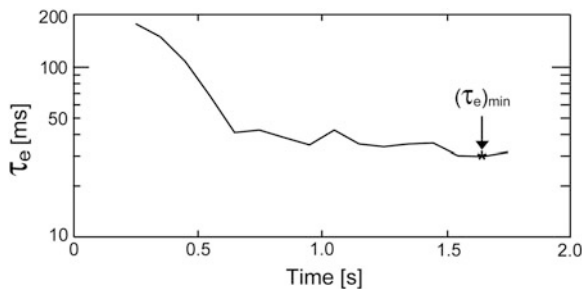


Fig. 2.6 An example of obtaining the minimum value in the effective durations extracted from the ACF

logarithmic form as a function of the delay time. The envelope decay of the initial and early part of the ACF may be fitted usually 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 to -5 dB. When the 5 dB range is available such as for singing voice of vowels, the value of τ_e is obtained by the initial 5 ms-delay interval (Sect. 2.3).

The recommended signal duration $(2T)_r$ to be analyzed is discussed in Sect. 2.2.

2.1.5 Minimum Values of the Effective Duration Extracted from Running ACF

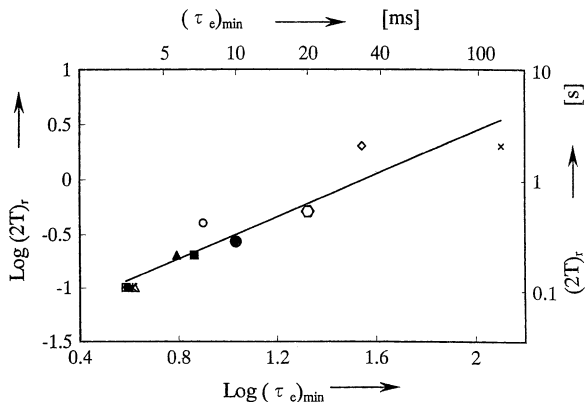
The minimum value of a moving τ_e , the most active part of music and speech including on and off sets of signals, containing important information and influencing subjective responses for the temporal criteria. An example of the value of $(\tau_e)_{\min}$ is illustrated in Fig. 2.6.

2.2 Auditory Temporal Window

In analysis of the running ACF, so-called the “auditory-temporal window” $2T$ in Eqs. (2.3) and (2.4) must be carefully determined. The initial part of 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. Results are shown in Fig. 2.7 and a recommended signal duration $(2T)_r$ to be analyzed is approximately given by

$$(2T)_r \approx 30(\tau_e)_{\min} \quad (2.6)$$

Fig. 2.7 Recommended signal duration to be analyzed in obtaining the ACF



where $(\tau_e)_{\min}$ is the minimum value of τ_e obtained by analyzing the ACF (Mouri et al. 2001). This signifies an adaptive temporal window “depending on the temporal activity” of the sound signal in the auditory system. For example, the temporal recommended windows differ according to music pieces $(2T)_r = 0.5\text{--}5$ s, and to the vowel $(2T)_r = 50\text{--}100$ ms and consonants $(2T)_r = 5\text{--}10$ ms in the continuous speech signal. Thus, brain might be more relaxed when listening to music than listening to speech. In other word, more concentration should be paid in listening speech than listening music.

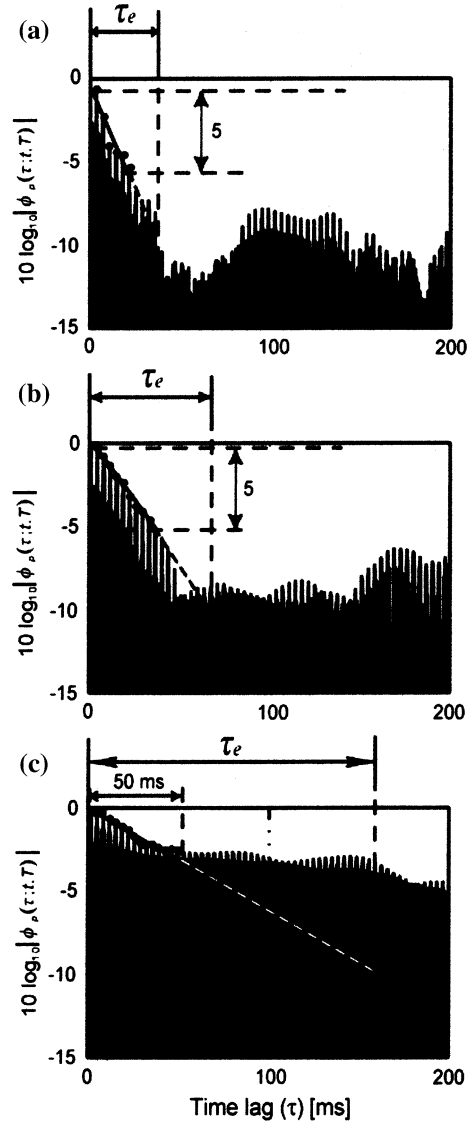
Also, in the noise measurement, for example, the time constant represented by “fast” or “slow” of the sound level meter might be replaced by the temporal window, which is well described by the effective duration of ACF of the source signal. Note that the running step (RS), which signifies a degree of overlap of signal to be analyzed, is not critical. It may be selected as $K_2(2T)_r$, K_2 being chosen, say, in the range of $1/4\text{--}1/2$.

2.3 Vocal Source Signal

In an opera house, vocal music sounds are produced on the stage. In order to demonstrate a procedure of extracting the effective duration from the running ACF analyzed, Fig. 2.5 shows the absolute value in the logarithmic form as a function of the delay time (Kato et al. 2007). The envelope decay of initial and important parts of running ACF may be fitted by a straight line in the range of 5 dB for most cases as shown in Fig. 2.8a, b. But, sometimes such a 5 dB range are not available as shown in Fig. 2.8c, so that the value of τ_e is obtained by the initial 50 ms-delay interval, as far as speech signal is concerted.

Examples of the τ_e values analyzed for vowel signals sung by a soprano are demonstrated in Fig. 2.9 with three different signal durations integrated $(2T)$.

Fig. 2.8 Examples of the ACF envelope in logarithmic form and the τ_e value extracted (Kato et al. 2007)



Although, τ_e values are varied according to $2T$, however, the most important minimum value as well as local minima are independent in certain range of $2T$ for vocal signals.

Further discussion is made in Sects. 9.3 and 9.4 for blending with the sound field for listeners.

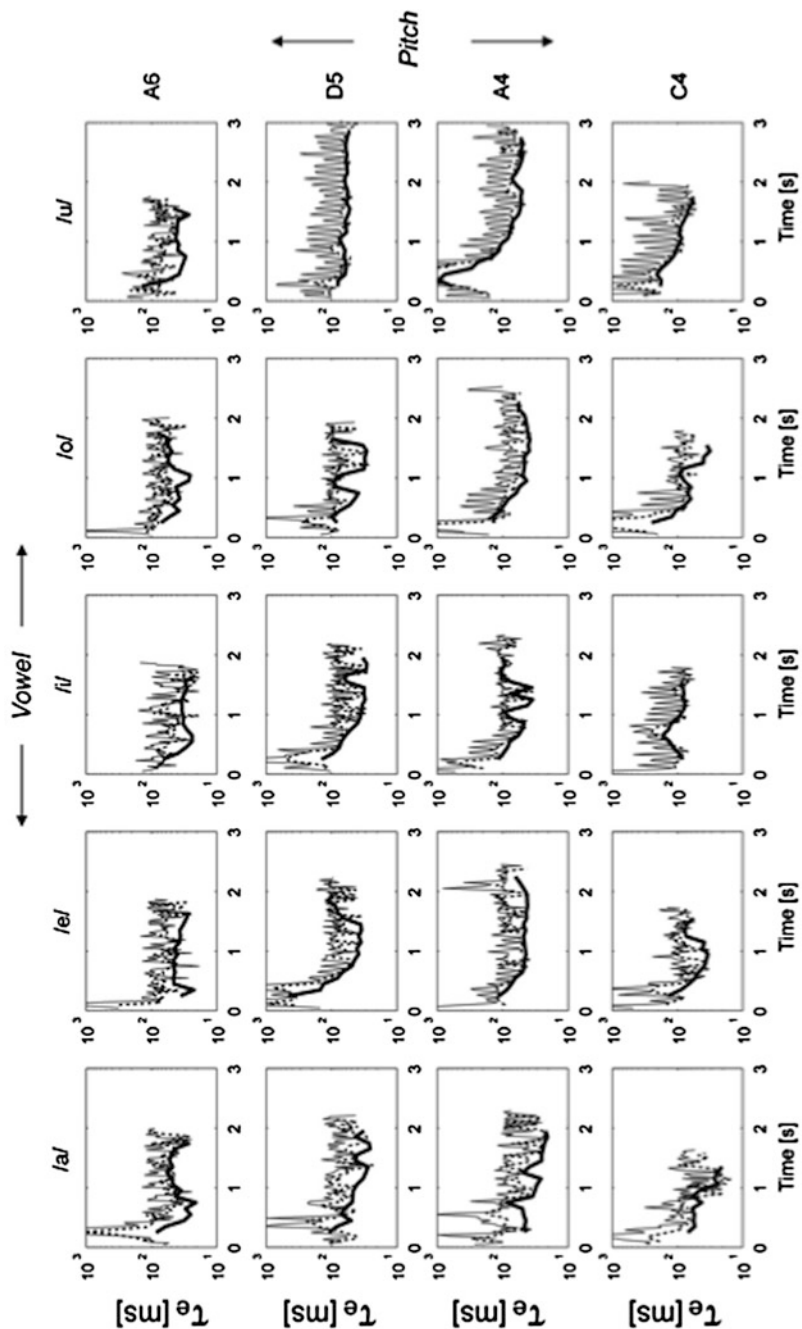


Fig. 2.9 Examples of the measured τ_c value extracted from the ACF of 20 vowels sung by a soprano singer with four different pitches obtained for three different signal durations (Kato et al. 2007). *Tine curve* $2T = 100$ ms, *dotted curve* $2T = 200$ ms, and *thick curve* $2T = 500$ ms

2.4 Running ACF of Piano Signal with Different Performance Style

We shall analyze a piano signal as a sound source in the orchestra pit. In order to examine whether or not we can control the value of τ_e of the running ACF 2 ($= 2$ s) of piano signals of varying performance styles for blending with a given sound field, a piano was controlled by its performing style using a computer.

Signals played by a piano were recorded in an anechoic chamber and analyzed (Taguti and Ando 1997). As is described above, the effective duration of running ACF, τ_e , is the fundamental time unit of the sound field (Eqs. 6.4 and 6.8; Ando 1998, 2009a). The performance style may be controlled blending the temporal factor of the sound field and the most preferred initial time delay gap between the direct sound and the first reflection, and the preferred subsequent reverberation time (Chap. 6). If the effective duration of running ACF is varied by the performing style, then musician may control it to fit the preferred temporal condition of the sound field.

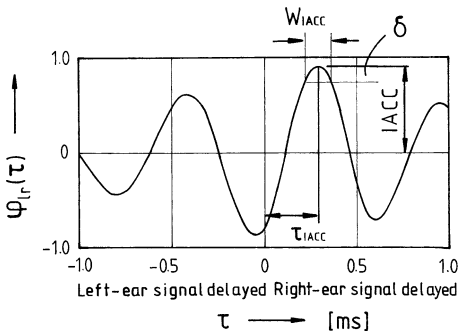
Typical results of the effective duration extracted from the running ACF in changing style of piano performance—staccato and legato—are shown in Table 2.1. As is expected, staccato resulted in a short value of the effective duration, τ_e , and legato and super legato leads to long values. The use of the damper pedal creates long values of the τ_e . The minimum value of τ_e corresponds roughly to values of the note-onset duration (NOD).

Table 2.1 Various styles of piano performance and the effective duration of ACF, τ_e

Style of performance	NOD (ms) ^a	τ_e (ms)
Staccato	50	61–87
Legato	125	106–170
Super legato	160	170–233
Mixed	–	110–155

The music piece used is the opening 8 bars of exercise no. 1, Hanon Tempo mm = 120 under constant dynamics
^a NOD is the note-on duration

Fig. 2.10 Definitions of three spatial factors extracted from the interaural cross-correlation function (IACF)



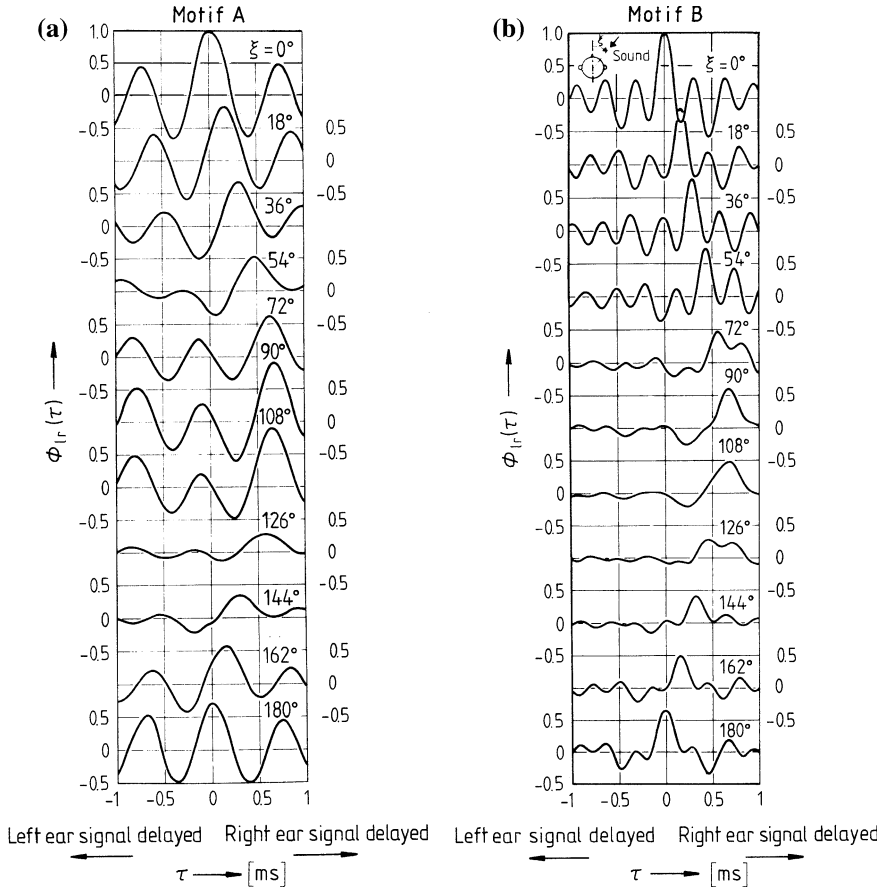


Fig. 2.11 Measured IACF in an anechoic chamber as a function of the interaural delay time and as a parameter of the horizontal angle of sound incidence (Mehrgardt and Mellert 1977). **a** Music motif A. **b** Music motif B

Staccato shortens the value of τ_c as the acuteness increases, but the value becomes no shorter than the minimum value of 60 ms. This lower limit may be caused by a mechanism in producing sound from the piano. So far, the value of τ_c of source signals may be controlled by changing the performing style blending with a given sound field in an opera house (Figs. 2.10 and 2.11).

Opera House Acoustics Based on Subjective Preference
Theory

Ando, Y.

2015, XIV, 179 p. 130 illus., 8 illus. in color., Hardcover

ISBN: 978-4-431-55422-6