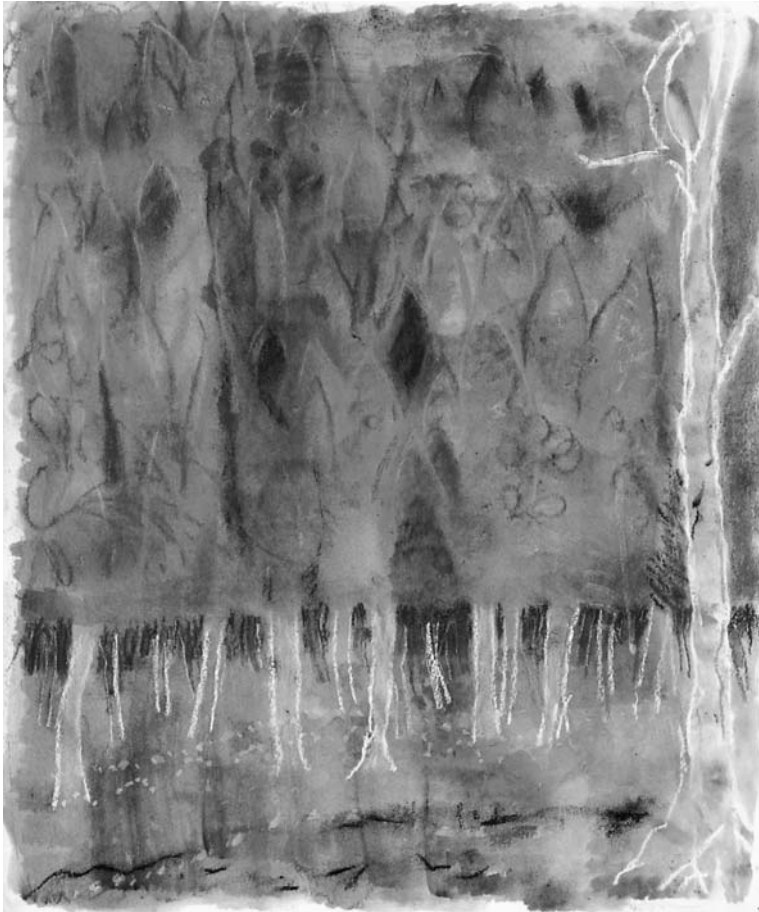


Yoichi Ando • Peter Cariani

# Auditory and Visual Sensations

 Springer

# Auditory and Visual Sensations



"Music in Forest" by Keiko Ando. A forest filled with birdsong, before the emergence of human societies.

Yoichi Ando

*Guest Editor*

Peter Cariani

# Auditory and Visual Sensations

 Springer



Yoichi Ando  
Professor Emeritus, Kobe University  
1-4-132-105 Hiyodoridai  
Kita, Kobe  
657-1123 Japan  
andoy@cameo.plala.or.jp

*Guest Editor*  
Peter Cariani  
Harvard Medical School  
629 Watertown Street, Newton, MA  
USA  
cariani@mac.com

ISBN 978-1-4419-0171-2                      e-ISBN 978-1-4419-0172-9  
DOI 10.1007/b13253  
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2009929375

© Springer Science+Business Media, LLC 2009

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

*Cover illustration:* Cover photo [8.1] was provided by Kirishima International Music Hall, Kagoshima Pref. The back cover photo [8.2] was provided by the Tsuyama Music Cultural Hall, Tsuyama City.

Printed on acid-free paper

Springer is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))

*Dedicated to Manfred R. Schroeder*

# Preface

This book is one of the first scientific attempts to relate aesthetics to neural activity patterns in both auditory and visual areas of the brain. We present a host of correspondences between human subjective preferences and brain activity as observed through electroencephalography (EEG) and magnetoencephalography (MEG).

The multimodal set of investigations presented in this book grew out of the development of a neurally grounded theory of subjective preference for sound fields in concert halls. The theory is based on a model of the human auditory system (Ando, 1985, 1998, 2007). In the theory, subjective preference is shaped by ecological utility to embody patterns of primitive responses that enhance survival. The auditory preference model assumes two kinds of internal representations of sound that are based on the correlation structure of sound as it presents itself to the two ears. These representations are based on autocorrelation and crosscorrelation. The autocorrelation function (ACF) describes the monaural signal at each of the two ears, and the interaural crosscorrelation function (IACF) describes the correlations between the two monaural signals arriving at the entrances of the two ears.

The autocorrelation and crosscorrelation representations have a firm neural basis in the temporal patterning of spike activity in the early stages of auditory processing. These time domain neural representations complement more familiar frequency domain representations that are based on spatial patterns of elevated neural discharge. At the level of the auditory nerve, the spatial, spectral representations embody Helmholtz's cochlear "place principle," while temporal patterns of spike discharge at each cochlear place faithfully follow the time structure of the filtered acoustic waveform. Temporal patterns of spikes can then be analyzed by higher auditory stations to extract information about the acoustic source and its spatial surrounds. Being temporal correlation representations rather than spatial profiles of neuronal excitation, they have a form and neural implementation that is much different from those associated with the power spectrum. In addition to temporal representations of the acoustic form of the stimulus, we have also found correlates of subjective preferences in temporal response properties at several points in the auditory pathway.

Part I of this book discusses central autocorrelation (ACF) and binaural cross-correlation (IACF) representations that we believe respectively subserves perception of tonal quality and of auditory spatial attributes. Many aspects of tonal quality,

including pitch, timbre, loudness, and duration, can be extracted from features of the central autocorrelation representation. On the other hand, spatial sensations such as localization in the horizontal plane, apparent source width (ASW), and subjective diffuseness may be described in terms of spatial factors extracted from the central crosscorrelation representation (IACF). Table 1 gives an overview of the perceptual attributes, their internal representations, likely locations, and neurophysiological observables.

Thus, the various attributes of the “primary sensation” of the whole sound can be divided into two categories: “temporal sensations” and “spatial sensations.” Any subjective responses of the sound field, therefore, may be described in terms of combinations of temporal and spatial factors. Further, these two sets of factors appear to be processed somewhat differently in the two cerebral hemispheres. The temporal factors extracted from the ACF appear to be associated with neuronal responses from the left cerebral hemisphere, whereas the spatial factors extracted from the IACF appear to be associated with those of the right hemisphere. Such hemispheric specialization and the relative independence of the two types of representation may play important roles in shaping the structure of subjective judgments.

Part II of this book discusses similarities between auditory and visual processing. Although the theory of subjective preference was developed with auditory perception in mind, it can plausibly be extended to predict subjective preferences in analogous dimensions of visual perception. Analogies can then be drawn to temporal and spatial sensations of vision, as well as for subjective preferences of visual environments. For example, the most preferred condition of a flickering light is expressed by the temporal factors extracted from the autocorrelation (ACF) of the temporally modulated light stimulus. The preference curve in relation to one of the temporal factors is given by a common formula such that

$$S \approx -\alpha |x|^\beta$$

where  $S$  is relative preference,  $\alpha$  is a scaling coefficient that depends on the particular individual, and  $x$  is the factor normalized by his or her most preferred value. Remarkably, the form of the curve is invariant across subjects, with  $\beta$  always having a fixed value of  $3/2$ .

This relation also holds true for the subjective preference curve in relation to each of four orthogonal factors for the sound field; that is,

1. listening level (LL),
2. initial delay time between the direct sound and the first reflection ( $\Delta t_1$ ),
3. subsequent reverberation time ( $T_{\text{sub}}$ ), and
4. the IACC, which is the maximum magnitude of the IACF.

However, few acousticians or musicians know the value of their own, most-preferred reverberation time, and our tests have shown great individual differences of subjective preference in relation to such temporal factors and listening levels. On the other hand, most subjects prefer sound sources that are spatially more diffuse

**Table 1** Summary of temporal and spatial sensations as well as subjective preference in relation to factors extracted from the ACF and IACF together with four orthogonal factors of the sound field and the physiological locus and observables to be described in Part I of this volume

Quality or attribute	Acoustic correlate	Representation	Main factor	Locus	Observable(s)	Section(s)
<b>Temporal sensations and preference</b>						
Periodicity pitch	Fundamental frequency	ACF	$\tau_1$ and $\phi_1$ extracted from ACF	brainstem (Left hemisphere)	Single units, ABR	5.1, 6.2
Loudness	Source signal	ACF	$\tau_1, \tau_e$ in the condition of constant LL	brainstem (Left hemisphere)	ABR	4.1, 4.2, 5.1, 6.4
Duration sensation	Duration	ACF	Signal duration D and $\tau_1$ extracted from ACF	brainstem (Left hemisphere)	ABR	4.1, 6.5
Timbre	Power spectrum	ACF	$W_{\phi(0)}$ extracted from ACF	Left hemisphere	Single unit	5.1
Preferred first reflection, $[\Delta\tau_1]_p$	Distance of nearest reflecting surface	ACF	$\tau_e$ extracted from ACF	Left hemisphere	SVR, EEG, MEG	3.1, 4.3, 4.4
Preferred reverberation time, $[T_{sub}]_p$	Volume of a room and absorption	ACF	$\tau_e$ extracted from ACF	Left hemisphere	EEG	3.2, 4.3
<b>Spatial sensations and preference</b>						
Localization (azimuth)	Source position	IACF	$\tau_{IACC}$ and IACC extracted from IACF	brainstem	ABR	4.1, 7.1
ASW	Location of reflectors and frequency component	IACF	IACC, $W_{IACC}$ , LL	brainstem Right hemisphere	ABR, SVR	4.1, 4.2, 7.2
Subjective diffuseness	Location of reflectors	IACF	IACC, LL	brainstem	ABR, SVR	4.1, 4.2, 7.3
Preferred listening level, $[LL]_p$	Distance from the source to receiving position, volume of a room and absorption	IACF	LL	brainstem Right hemisphere	ABR, SVR	3.2, 4.1, 4.2
Preferred IACC: a small value	Location of reflectors	IACF	IACC	brainstem Right hemisphere	ABR, SVR	3.2, 4.1, 4.2

and enveloping, i.e. they produce low IACC values with correspondingly high subjective diffuseness. Because subjects may differ as to which aspects of sounds they care about most, we have developed a model of individual preference that adjusts the relative weights of the associated spatial and temporal factors.

It is remarkable that neural activities include sufficient information to reconstruct the ACF of the stimulus. For example, at the level of the auditory nerve, Cariani and Delgutte (1996) found that pooled interspike interval distributions resemble the short time, or running ACF for lower-frequency components that are perceptually resolved. For sets of stimulus components with higher harmonic numbers that cannot be perceptually resolved, the pooled interval distributions that are produced reflect time-locking to waveform envelopes such that their form therefore resembles the envelope of the running ACF.

At the cortical level, a feature of the autocorrelation of alpha waves of both EEG and MEG corresponded to subjective preference of the reverberatory characteristics of sound fields. The repetitive feature of the EEG alpha wave, the “effective duration” of its ACF, was always observed at the preferred condition. This means that the basic theory of subjective preference can be applied to each individual’s preferences for this parameter. We reconfirmed by the alpha wave of MEG that the left cerebral hemisphere response is associated with preferred first reflection time  $\Delta t_1$  and found that the effective duration of the ACF of the MEG alpha wave directly corresponds to the magnitude of an individual’s subjective preference. The right cerebral hemisphere was activated by the typical spatial factors, that is, the magnitude of interaural crosscorrelation (IACC), which reflects the spatial subjective diffuseness of the sound field.

This book largely serves as a record of the research performed at the Ando Laboratory, Graduate School of Science and Technology, Kobe University, between 1969 and 2009, even after the author’s retirement. The first part of this volume recapitulates previous experiments whose results were reported by the author in 1985 and 1998. To describe clearly the important facts that were discovered, details of each investigation are recounted to avoid confusion. In the interest of space, we have left several issues out of this volume that were discussed in the 1998 book. Among them are the descriptions of the physical system from the free field to the oval window, which accounts for the sensitivity of the ear, and the details of our method for obtaining individual preference profiles.

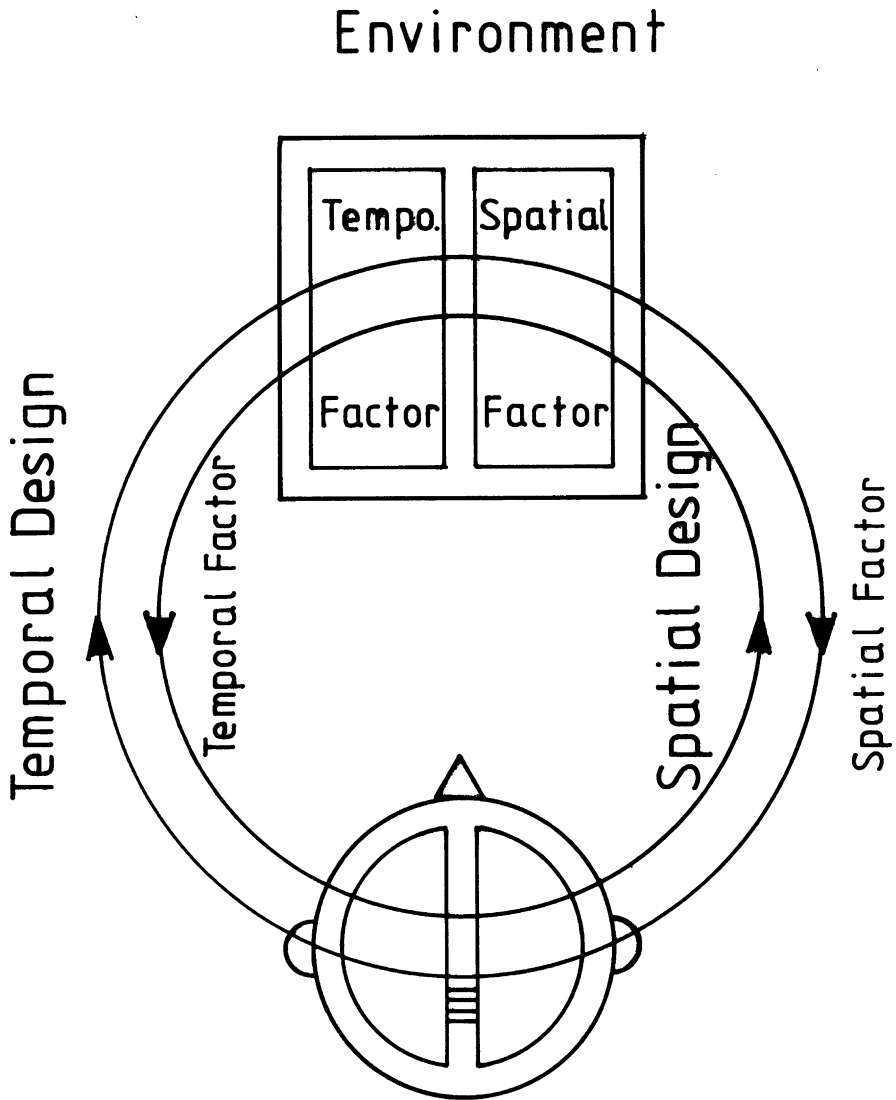
It is hoped that this volume will make a useful and lasting contribution to research that relates acoustics and the brain, from the architectural acoustics of concert halls and opera houses, to the effects of noise on humans, to the psychological and physiologic acoustics of speech and music perception.

The minimum unit of society is the individual, and an individual’s subjective preference may be a reflection of a unique personality or perspective. The individual personality generates creative ideas, which persist long after the individual has passed on. If the life of the body is the “first life,” and that of the mind is the second, then the propagation of one’s ideas constitutes yet another, “third life.” A healthy creation can contribute to human society and the environment in this third life for a long time even after the end of the first and second ones. Ideas can create better environments for thinking. A self-conscious design process that incorporates spa-

tial and temporal factors can facilitate more productive interactions between brains and environments that in turn induces still further discoveries and creations. In this way, the design of human thinking environments amplifies itself in an open-ended, creative way that persists long after the human beings who initiated it have left the scene (Fig. 1).

Kobe, Japan

Yoichi Ando



**Fig. 1** Interaction between brain and environment created by incorporating with temporal and spatial factors, which are associated with the left and right cerebral hemispheres, respectively. Such an environment realized by each individual may induce further discoveries and creations (Ando, 2009)

## **Previous Publications Related to This Topic**

- Ando, Y. (1985). *Concert Hall Acoustics*. Springer-Verlag, Heidelberg.
- Ando, Y. (1998). *Architectural Acoustics, Blending Sound Sources, Sound Fields, and Listeners*. AIP Press/Springer-Verlag, New York.
- Ando, Y. (2007). Concert hall acoustics based on subjective preference theory. In: Rossing, T. (ed.), *Handbook of Acoustics*. Springer-Verlag, New York, Chapter 10.
- Ando, Y. (2009). Theory of temporal and spatial design of environment. In: Blumel, D. (ed.), *McGraw-Hill 2009 Yearbook of Science and Technology*. McGraw-Hill, New York.



# Guest Editor's Preface

Yoichi Ando's theory of architectural acoustics can be summarized by the sequence sound–space–representation–percept–preference–design (Ando, 1985, 1998). The theory begins with the transmission of sound through an architectural enclosure to the ears of listeners situated at different locations. An auditory signal processing model forms idealized internal neural representations that subserves the auditory percepts that a listener experiences. These auditory percepts include sound qualities such as pitch, timbre, loudness, and duration that are mainly attributed to the sound source; spatial attributes of direction, apparent size, and subjective diffuseness that are attributed to the spatial relation of source and listener; and reverberatory characteristics that are attributed to the enclosure (Ando, 2008). All other factors being equal, most listeners prefer particular ranges of these attributes. Once subjective preferences of listeners are understood, then it becomes possible to relate them to the acoustic properties of sources and spaces and the perceptual dimensions that are generated by auditory representations. One can then rationally optimize designs for acoustic performance spaces on the basis of the acoustics, the auditory neurophysiology and psychoacoustics, and the psychology of subjective preferences. The ultimate task is to find the optimal form of an enclosure that maximally enhances the perceptual experiences of its listeners, subject to architectural constraints and the acoustic characteristics of the program material. Ando's theory takes us all the way from the physical world of acoustics to the mental world of auditory percepts and preferences to the pragmatics of architectural acoustic design (Ando, 2007b).

Architectural acoustics necessarily involves both physical properties of constructed spaces and psychological properties of the human beings within. Each of these complementary aspects requires a different set of observables. The physical models that are used to design architectural spaces are grounded in acoustic measurements in enclosures. Ando and his colleagues have also carried out extensive acoustic, neurophysiological, and psychophysical measurements in order to develop the psychological side of the theory. The model of auditory signal processing is grounded in auditory physiology and in the neurophysiology of early neural representations in the cochlea and auditory brainstem. I will discuss below how Ando's central autocorrelation function (ACF) and interaural crosscorrelation function (IACF) are grounded in observed temporal patterns of neural discharge. Higher in the auditory pathway, using evoked averaged electrical and magnetic responses,

they observe systematic changes in the response latencies of neural populations that correspond to perceptual attributes. In many cases, changes in the magnitudes and latencies of response differ across the two hemispheres of the cerebral cortex. Their behavioral psychoacoustics experiments probe the discriminability of auditory attributes of interest, and their subjective preference tests assess which values of those attributes are most preferred by listeners. Finally, they observe pervasive neural correlates of subjective preference in the persistence of neural oscillations in the alpha band.

Not content with auditory experience, Ando and colleagues have extended the theory to cover temporal aspects of visual sensations. Comparing the structure of auditory and visual percepts and preferences, they find deep and intriguing parallels between the two systems. Not content with perception and preference, they have formulated a general theory and philosophy of environmental design that also incorporates physics and physiology to harmonize the relationship of human beings, the built environment, and the natural surrounds (Ando, 2009).

I met Yoichi Ando through a convergence between his theory of architectural acoustics and my neurophysiological investigations of temporal coding in the auditory system. Following on correspondences, we met in person at the annual meeting of the Acoustical Society of America in New York in 2004 and then again in 2005 in Kirishima, Japan, at the Second Meeting on Temporal Design (<http://www.jtdweb.org/>). Yoichi had developed a systematic, perceptual theory of architectural acoustics based on temporal autocorrelation and crosscorrelation analysis of the acoustic signal. Through my experimental work with Bertrand Delgutte at the Eaton Peabody Laboratory of Auditory Physiology on temporal coding of pitch and timbre (Cariani and Delgutte, 1996a,b; Cariani, 1999), it had become apparent to me that the perception of most auditory qualities is based on information embedded in the timings of spikes.

What follows is a brief account of how temporal autocorrelation and crosscorrelation representations arise from spike timing patterns in the auditory nerve. We will also discuss the possibility that temporal codes and computations constitute general information processes strategies that have applicability well beyond the auditory system.

In a temporal neural code, information is represented and conveyed through temporal patterns of spikes. This can be contrasted with “rate-place” codes in which information is encoded in firing rate profiles across neurons in a population. In addition, “place” means cochlear place, which, due to cochlear tuning, functionally forms a one-dimensional array of band-pass filters in which center frequency corresponds to cochlear place. Spike timing patterns convey information about the stimulus in each cochlear frequency region. The timings of spikes in auditory nerve fibers innervating inner hair cells at each cochlear “place” are highly correlated with the filtered vibrations of the basilar membrane at that point. The character of this stimulus-driven spiking activity is called “phase-locking” because when acoustic stimuli are presented at sound-pressure levels sufficient to produce elevated firing rates in these auditory nerve fibers, spikes are almost exclusively produced in response to only one phase of the filtered stimulus waveform. Locking of spikes to

one phase is the consequence of half-wave rectification caused by molecular biophysical mechanisms that transduce mechanical vibrations to electrical, ionic currents in inner hair cells. A further chain of events connects ionic currents in hair cells with time-locked neurotransmitter releases and the consequent postsynaptic currents in auditory nerve fibers that influence the fine timing of action potential initiation. The overall result is that auditory nerve fibers innervating each cochlear place produce trains of spikes whose temporal patterns reflect the time structure of the stimulus as it has been modified by cochlear filtering at that location.

A direct consequence of phase-locking is at that time intervals between spikes, called interspike intervals, convey information about stimulus periodicities, all the way from an upper limit of several kilohertz down to periodicities associated with infrapitch, room reverberations, rhythms, and longer patterns of acoustic events. If intervals between both consecutive and nonconsecutive spikes are taken into account, distributions of these “all-order” interspike intervals are equivalent to the temporal autocorrelation function of the spike trains. Because the spikes themselves are correlated with the cochlear-filtered stimulus waveform, the all-order interspike interval distributions of auditory nerve fibers at any given cochlear place in effect form an autocorrelation-like representation of the stimulus waveform at that cochlear place. This representation is statistical, based on many time intervals between pulsatile action potentials. The sum of the autocorrelations of the outputs from a dense set overlapping band-pass filters is equal to the autocorrelation of the unfiltered waveform. As a consequence, if one pools together all-order interspike intervals from all of the spike trains of the 30,000 auditory nerve fibers that make up the auditory nerve, then the “population-interval” distribution that is formed strongly resembles the positive portion of the stimulus autocorrelation. Thus an analog representation of the autocorrelation of the acoustic signal can be formed from the mass statistics of discrete spiking events in a neural population.

The notion of periodicity, the repetition rate of a series of events, arguably predates more modern concepts of frequency that are grounded explicitly in sinusoidal decompositions. As de Cheveigné has pointed out, the respective archetypal exemplars of periodicity and frequency are respectively the string with its multiple harmonic resonance modes and the Helmholtz resonator with its single resonance frequency (de Cheveigné, 2004). The idea that the primary analysis carried out by the auditory system might be periodicity-based rather than frequency-based has a long history that surfaced with the Seebeck–Helmholtz–Ohm debate concerning the low, fundamental pitch of harmonic complex tones. It runs through Rutherford's “telephone theory” of neural coding (Boring, 1942), Troland's temporal modulation theory of pitch (Troland, 1929), and Wever's “volley principle” (Wever, 1949). Correlation analysis of signals became widely used in electrical engineering in the 1940s, after Wiener and others had demonstrated the formal relations between autocorrelation functions and Fourier spectra. In 1948, Jeffress proposed his neural temporal crosscorrelation model of binaural localization. In 1951, J.C.R. Licklider proposed that the auditory system carries out both a frequency analysis by virtue of cochlear filtering and an autocorrelation analysis by virtue of phase locking (Licklider, 1951). Licklider also proposed a “duplex” time-delay neural network that could estimate

power spectra of acoustic stimuli from neural firing rates and their autocorrelations from interspike intervals. Later that decade, both Licklider and Cherry each proposed auditory signal processing architectures that combined monaural autocorrelation analysis with binaural crosscorrelation (Licklider, 1959; Cherry and Sayers, 1956; Cherry, 1961). I think of these two models as the ancestors of Ando's dual autocorrelation and crosscorrelation representations. Unfortunately, the neural autocorrelation models of the 1950s were ignored for several decades, but the basic idea was revived in the early 1980s by models of pitch (Moore, 1982; van Noorden, 1982) that summed together all first-order interspike intervals (interspike intervals between consecutive spikes) from across the auditory nerve. These population-based models, which were based on first-order intervals, were superseded less than a decade later by models based on all-order intervals, the "summary autocorrelation" models of Meddis and Hewitt (1991a) and Slaney and Lyon (1993).

In our neurophysiological studies of the neural coding of pitch in the auditory nerve (Cariani and Delgutte, 1996a,b; Cariani, 1999), we investigated the properties of population-wide all-order interspike interval distributions *in vivo*. These "population-interval" distributions are functionally equivalent to computed "summary autocorrelations." We looked for correspondences between features of these global interval distributions and the perception of the low pitches of harmonic complexes. We found (1) that the low pitch that is heard at the fundamental of harmonic complexes almost invariably corresponds to the predominant all-order interspike interval present in the auditory nerve array; (2) that the perceived strength of this pitch "pitch salience" corresponds to the fraction of interspike intervals in the auditory nerve array that are related to that pitch; and (3) that the form of the population-interval distribution strongly resembles the positive portion of the stimulus autocorrelation function. We also found that synthetic vowels produce characteristic population-interval distributions by which they can be readily distinguished.

The main import of these findings in the present context is that an autocorrelation-like representation of the stimulus is directly present in the mass statistics of all-order interspike intervals in the auditory nerve. Such a representation holds for periodicities up to the frequency limits of significant phase-locking, roughly 4–5 kHz. Not only can such a representation handle the pitch of pure and complex tones up to this limit but also those aspects of timbre that are associated with the shape of the power spectrum.

Such evidence provides neural grounding for Ando's correlation-based signal processing model. In Ando's theory, those auditory qualities that can be represented by features in the monaural temporal ACF are called "temporal sensations." These include pitch, timbre, duration, loudness, and also percepts associated with the longer delays that are created through patterns of sound reflection. These latter perceptual qualities are attributed to the room rather than the source, and they correspond to acoustic parameters such as first reflection time, patterns of later reverberations, and the interaction of direct and reflected sound (as captured in the analysis of "effective duration"). The adoption of extended autocorrelation representations that include both short time lags associated with pitch, timbre, and consonance and longer ones associated with room reflections is naturally appealing

to architectural acousticians who desire integrative, systematic, and parsimonious theories of hearing.

In Ando's theory, those qualities of sound that involve perception of the spatial relations of sound vis-à-vis the listener are called "spatial sensations." The main spatial qualities are sound direction, apparent size, and the apparent diffuseness of the sound source. The quality of subjective diffuseness is related to the listener's sense of envelopment by the sound. For the architectural acoustics of performance spaces, sound localization in the horizontal plane is by far most important. In this plane, spatial qualities are computed from features in the interaural temporal crosscorrelation function (IACF), mostly from the mean interaural time delay ( $\tau_{IACC}$ ) associated with a given source and the dispersion of its values (apparent source width).

The role of temporal coding in binaural hearing is generally much more widely appreciated than in monaural hearing. It is relatively well known that human listeners are able to distinguish sound directions as little as  $1^\circ$  to  $2^\circ$  in azimuth, which corresponds to interaural time differences of roughly  $20 \mu\text{s}$ . Less widely appreciated is that precisions of frequency discriminations in monaural hearing, e.g. Weber fractions in the vicinity of 0.2% for 1 kHz tones, yield just-noticeable time differences for the periods of sounds that are also on the same order ( $20 \mu\text{s}$ ). In effect, a neural crosscorrelation operation in the auditory brainstem is implemented via phase-locking of spikes with submillisecond precisions in the left and right auditory nerves, secure synapses in the cochlear nucleus, systematic arrangement of axonal delays, precisely timed neural inhibition, and an array of bipolar neural spike coincidence detection elements in the nucleus of the medial superior olive (MSO). The details of exactly how interaural temporal disparities are represented at still higher auditory stations (e.g., via firing rates, temporal patterns, or spike latencies) are still being worked out. In my opinion, the crosscorrelation analysis (IACF) in Ando's auditory signal processing model best coincides with the neural representations produced at the output of this brainstem stage of binaural auditory processing.

The autocorrelation and crosscorrelation analyses of Ando's theory can be extended to other sensory modalities (Cariani, 2001). To generalize, if one has a sensory system with receptors and primary sensory neurons that are capable of producing spikes that are temporally correlated with the adequate stimulus, then correlation operations can perform two kinds of analyses. First, an autocorrelation analysis of the temporal pattern of spikes produced at any one sensory receptor surface can provide information about the temporal form of the stimulus that is driving the receptors. Salient examples include rhythm, pitch, and timbre in audition, estimation of range using echo delays, flutter-vibration in the somatosensory system, and visual perception of temporal modulation patterns such as the rates of flashing lights. Second, a crosscorrelation analysis of spike trains generated in sensory receptors at different locations can provide information about the relative time-of-arrival of the stimulus at those respective locations. The relative time-of-arrival at different body locations in turn serves as an indicator of the direction of waves that propagate from a common source. Perspicuous examples include binaural localization, electroreception, somatosensory localization of mechanical and electrical pulse pairs, olfactory localization, and depth illusions (Pulfrich effect) that arise from binocular

temporal disparities. Georg von Békésy famously demonstrated these kinds of spatial orientation mechanisms in almost every imaginable sensory modality (Békésy, 1967). Alternately, the relative time-of-arrival at different receptors in an array can serve as an indicator of motion, be it across whiskers or retinal elements or acoustic frequencies.

This book is an important integrative step toward a theory of the brain in which information is based on temporal correlations of spikes. What remains to be seen is whether the temporal representations that are seen in early stages of sensory processing in various modalities are also used, albeit with transformations, in some covert and possibly distributed form at more central stations. I do believe that we shall live to see the day when the precise nature of cortical codes and computations comes to be elucidated, such that we finally understand how brains work as informational systems.

In the editing of Yoichi's manuscript, I have attempted to remain faithful to its structure, formal details, and ideas while rendering the text more readable and intelligible for those not fully familiar with architectural and physiological acoustics. It has been my pleasure to meet and work with Yoichi Ando and my privilege to contribute to the presentation of his life's work in the form of this book.

Newton, Massachusetts, USA

Peter Cariani

# Acknowledgments

I would like to thank Manfred Schroeder, University of Goettingen, who has been instrumental in encouraging this work for more than three decades. Since the mid-1970s, he has invited me several times to his physics institute in Goettingen as an Alexander-von-Humboldt Fellow to work on the research and to prepare manuscripts of three books, including this volume. I would also like to thank Takeshi Itoh, Waseda University, who acted as my doctoral adviser during the years 1973–1975. I owe a lifelong debt to Takaharu Sato, who spent many hours, after work in 1956, teaching me elementary mathematics once a week without compensation.



**Photo 1** A wild rabbit in the garden of the Schroeder’s home in Nikolausberg who jumped and turned around with us. This made us aware that such an animal has mind, primitive and pure, but not creative power

I discussed with Werner Lauterborn the drawings exhibited on every wall of his institute, the Drittes Physikalisches Institut (DPI), University of Goettingen, during my stay for the period July–October 2003. We concluded that a kind of “vibrato” is one of the expressions of beauty for drawings, and he prepared a photograph (Photo 14.1). During this period, Keiko and I stayed at the house of Manfred and Anny Schroeder in Nikolausberg, and I worked on the manuscript of this book at Manfred’s office at DPI as well. In the large garden of their home, Keiko and I enjoyed communicating with a wild rabbit named “Hasefeld” who visited us almost every day, dancing and jumping and turning around with us (Photo 1). This good experience made us more aware of the importance of human creations, because even animals have their own minds, primitive and pure.

Shin-ichi Sato, Hiroyuki Sakai, Yoshiharu Soeta, and Kenji Fujii, who received their Ph.D. degrees at the graduate school, have written a number of excellent works, and their publications are cited here. They and other colleagues of Ando Lab provided illustrations and comments on this manuscript. Peter Cariani, Harvard Medical School, kindly provided his original plots of auditory nerve responses (Fig. 5.2) in addition to providing useful comments and suggestions for improving the manuscript as a guest editor. I would also like to express my appreciation to the authors and publishers who have granted permission for use of their works and to Robert Beyer for his encouragement to publish this volume.



# Contents

<b>Preface</b> . . . . .	vii
<b>Guest Editor's Preface by Peter Cariani</b> . . . . .	xiii
<b>Acknowledgments</b> . . . . .	xix
<b>Part I Temporal and Spatial Sensations in the Human Auditory System</b>	
<b>1 Introduction</b> . . . . .	3
1.1 Auditory Temporal and Spatial Factors . . . . .	3
1.2 Auditory System Model for Temporal and Spatial Information Processing . . . . .	4
<b>2 Temporal and Spatial Aspects of Sounds and Sound Fields</b> . . . . .	9
2.1 Analysis of Source Signals . . . . .	9
2.1.1 Power Spectrum . . . . .	9
2.1.2 Autocorrelation Function (ACF) . . . . .	10
2.1.3 Running Autocorrelation . . . . .	13
2.2 Physical Factors of Sound Fields . . . . .	18
2.2.1 Sound Transmission from a Point Source through a Room to the Listener . . . . .	18
2.2.2 Temporal-Monaural Factors . . . . .	19
2.2.3 Spatial-Binaural Factors . . . . .	20
2.3 Simulation of a Sound Field in an Anechoic Enclosure . . . . .	23
<b>3 Subjective Preferences for Sound Fields</b> . . . . .	25
3.1 Preferred Properties for Sound Fields with Multiple Reflections . . . . .	26
3.1.1 Preferred Delay Time of a Single Reflection . . . . .	26
3.1.2 Preferred Horizontal Direction of a Single Reflection . . . . .	29
3.2 Preferred Conditions for Sound Fields with Multiple Reflections . . . . .	30

- 3.2.1 Optimal Listening Level (LL) . . . . . 30
- 3.2.2 Optimal First Reflection Time ( $\Delta t_1$ ) . . . . . 31
- 3.2.3 Optimal Subsequent Reverberation Times ( $T_{sub}$ ) . . . . . 31
- 3.2.4 Optimal Magnitude of Interaural  
Crosscorrelation (IACC) . . . . . 33
- 3.3 Theory of Subjective Preferences for Sound Fields . . . . . 34
- 3.4 Evaluation of Boston Symphony Hall Based on  
Temporal and Spatial Factors . . . . . 37
- 4 Electrical and Magnetic Responses in the Central Auditory  
System . . . . . 39**
- 4.1 Auditory Brainstem Responses (ABRs) . . . . . 40
  - 4.1.1 Brainstem Response Correlates of Sound  
Direction in the Horizontal Plane . . . . . 40
  - 4.1.2 Brainstem Response Correlates of Listening  
Level (LL) and Interaural Crosscorrelation  
Magnitude (IACC) . . . . . 44
  - 4.1.3 Remarks . . . . . 46
- 4.2 Slow Vertex Responses (SVRs) . . . . . 48
  - 4.2.1 SVR Correlates of First Reflection Time  $\Delta t_1$   
Contrast . . . . . 48
  - 4.2.2 Hemispheric Lateralization Related to Spatial  
Aspects of Sound . . . . . 50
  - 4.2.3 Response Latency Correlates of Subjective  
Preference . . . . . 53
- 4.3 Electroencephalographic (EEG) Correlates of  
Subjective Preference . . . . . 55
  - 4.3.1 EEG Correlates of First Reflection Time  $\Delta t_1$   
Changes . . . . . 55
  - 4.3.2 EEG Correlates of Reverberation Time  $T_{sub}$   
Changes . . . . . 58
  - 4.3.3 EEG Correlates of Interaural Correlation  
Magnitude (IACC) Changes . . . . . 60
- 4.4 Magnetoencephalographic (MEG) Correlates of  
Preference and Annoyance . . . . . 63
  - 4.4.1 Preferences and the Persistence of Alpha Rhythms . . . . . 63
  - 4.4.2 Preferences and the Spatial Extent of Alpha  
Rhythms . . . . . 68
  - 4.4.3 Alpha Rhythm Correlates of Annoyance . . . . . 68
- 5 Model of Temporal and Spatial Factors in the Central  
Auditory System . . . . . 73**
- 5.1 Signal Processing Model of the Human Auditory System . . . . . 73
  - 5.1.1 Summary of Neural Evidence . . . . . 73
  - 5.1.2 Auditory Signal Processing Model . . . . . 75

5.2	Temporal Factors Extracted from Autocorrelations of Sound Signals . . . . .	83
5.3	Auditory Temporal Window for Autocorrelation Processing . . . . .	84
5.4	Spatial Factors and Interaural Crosscorrelation . . . . .	86
5.5	Auditory Temporal Window for Binaural Processing . . . . .	87
5.6	Hemispheric Specialization for Spatial Attributes of Sound Fields . . . . .	87
<b>6</b>	<b>Temporal Sensations of the Sound Signal . . . . .</b>	<b>91</b>
6.1	Combinations of Temporal and Spatial Sensations . . . . .	91
6.2	Pitch of Complex Tones and Multiband Noise . . . . .	93
6.2.1	Perception of the Low Pitch of Complex Tones . . . . .	93
6.2.2	Pitch of Multiband “Complex Noise” . . . . .	100
6.2.3	Frequency Limits of Missing Fundamentals . . . . .	101
6.3	Beats Induced by Dual Missing Fundamentals . . . . .	105
6.4	Loudness . . . . .	108
6.4.1	Loudness of Sharply Filtered Noise . . . . .	108
6.4.2	Loudness of Complex Noise . . . . .	114
6.5	Duration Sensation . . . . .	119
6.6	Timbre of an Electric Guitar Sound with Distortion . . . . .	120
6.6.1	Experiment 1 – Peak Clipping . . . . .	122
6.6.2	Experiment 2 – Commercial Effects Box . . . . .	124
6.6.3	Concluding Remarks . . . . .	124
<b>7</b>	<b>Spatial Sensations of Binaural Signals . . . . .</b>	<b>125</b>
7.1	Sound Localization . . . . .	125
7.1.1	Cues of Localization in the Horizontal Plane . . . . .	125
7.1.2	Cues of Localization in the Median Plane . . . . .	126
7.2	Apparent Source Width (ASW) . . . . .	127
7.2.1	Apparent Width of Bandpass Noise . . . . .	130
7.2.2	Apparent Width of Multiband Noise . . . . .	131
7.3	Subjective Diffuseness . . . . .	136
<b>8</b>	<b>Applications (I) – Music and Concert Hall Acoustics . . . . .</b>	<b>143</b>
8.1	Pitches of Piano Notes . . . . .	143
8.2	Design Studies of Concert Halls as Public Spaces . . . . .	148
8.2.1	Genetic Algorithms (GAs) for Shape Optimization . . . . .	148
8.2.2	Two Actual Designs: Kirishima and Tsuyama . . . . .	153
8.3	Individualized Seat Selection Systems for Enhancing Aural Experience . . . . .	158
8.3.1	A Seat Selection System . . . . .	158
8.3.2	Individual Subjective Preference . . . . .	158
8.3.3	Distributions of Listener Preferences . . . . .	161
8.4	Subjective Preferences of Cello Soloists for First Reflection Time, $\Delta t_1$ . . . . .	165
8.5	Concert Hall as Musical Instrument . . . . .	172

- 8.5.1 Composing with the Hall in Mind: Matching Music and Reverberation . . . . . 172
- 8.5.2 Expanding the Musical Image: Spatial Expression and Apparent Source Width . . . . . 174
- 8.5.3 Enveloping Music: Spatial Expression and Musical Dynamics . . . . . 175
- 8.6 Performing in a Hall: Blending Musical Performances with Sound Fields . . . . . 175
  - 8.6.1 Choosing a Performing Position on the Stage . . . . . 175
  - 8.6.2 Performance Adjustments that Optimize Temporal Factors . . . . . 176
  - 8.6.3 Towards Future Integration of Composition, Performance and Hall Acoustics . . . . . 177
- 9 Applications (II) – Speech Reception in Sound Fields . . . . . 179**
  - 9.1 Effects of Temporal Factors on Speech Reception . . . . . 179
  - 9.2 Effects of Spatial Factors on Speech Reception . . . . . 185
  - 9.3 Effects of Sound Fields on Perceptual Dissimilarity . . . . . 189
    - 9.3.1 Perceptual Distance due to Temporal Factors . . . . . 194
    - 9.3.2 Perceptual Distance due to Spatial Factors . . . . . 195
- 10 Applications (III) – Noise Measurement . . . . . 199**
  - 10.1 Method of Noise Measurement . . . . . 199
  - 10.2 Aircraft Noise . . . . . 200
  - 10.3 Flushing Toilet Noise . . . . . 207
- 11 Applications (IV) – Noise Annoyance . . . . . 213**
  - 11.1 Noise Annoyance in Relation to Temporal Factors . . . . . 213
    - 11.1.1 Annoyance of Band-Pass Noise . . . . . 213
    - 11.1.2 Annoyance of Traffic Noise . . . . . 218
  - 11.2 Noise Annoyance in Relation to Spatial Factors . . . . . 223
    - 11.2.1 Experiment 1: Effects of SPL and IACC Fluctuations . . . . . 223
    - 11.2.2 Experiment 2: Effects of Sound Movement . . . . . 225
  - 11.3 Effects of Noise and Music on Children . . . . . 228
- Part II Temporal and Spatial Sensations in the Human Visual System**
- 12 Introduction to Visual Sensations . . . . . 235**
- 13 Temporal and Spatial Sensations in Vision . . . . . 237**
  - 13.1 Temporal Sensations of Flickering Light . . . . . 237
    - 13.1.1 Conclusions . . . . . 243
  - 13.2 Spatial Sensations . . . . . 243
- 14 Subjective Preferences in Vision . . . . . 253**
  - 14.1 Subjective Preferences for Flickering Lights . . . . . 253
  - 14.2 Subjective Preferences for Oscillatory Movements . . . . . 259

- 14.3 Subjective Preferences for Texture . . . . . 263
  - 14.3.1 Preferred Regularity of Texture . . . . . 263
  - 14.3.2 Application: Spatial “Vibrato” in a Drawing . . . . . 264
- 15 EEG and MEG Correlates of Visual Subjective Preferences . . . . . 267**
  - 15.1 EEG Correlates of Preferences for Flickering Lights . . . . . 267
    - 15.1.1 Persistence of Alpha Rhythms . . . . . 267
    - 15.1.2 Spatial Extent of Alpha Rhythms . . . . . 275
  - 15.2 MEG Correlates of Preferences for Flickering Lights . . . . . 282
    - 15.2.1 MEG Correlates of Sinusoidal Flicker . . . . . 282
    - 15.2.2 MEG Correlates of Fluctuating Flicker Rates . . . . . 288
  - 15.3 EEG Correlates of Preferences for Oscillatory Movements . . . . . 289
  - 15.4 Hemispheric Specializations in Vision . . . . . 295
- 16 Summary of Auditory and Visual Sensations . . . . . 297**
  - 16.1 Auditory Sensations . . . . . 298
    - 16.1.1 Auditory Temporal Sensations . . . . . 298
    - 16.1.2 Auditory Spatial Sensations . . . . . 299
    - 16.1.3 Auditory Subjective Preferences . . . . . 300
    - 16.1.4 Effects of Noise on Tasks and Annoyance . . . . . 301
  - 16.2 Visual Sensations . . . . . 304
    - 16.2.1 Temporal and Spatial Sensations in Vision . . . . . 304
    - 16.2.2 Visual Subjective Preferences . . . . . 305
- References . . . . . 307**
- Glossary of Symbols . . . . . 323**
- Abbreviations . . . . . 329**
- Author Index . . . . . 333**
- Subject Index . . . . . 337**

# Chapter 1

## Introduction

### 1.1 Auditory Temporal and Spatial Factors

For the past four decades, we have pursued a theory of architectural acoustics based on acoustics and auditory theory. In the summer of 1971, I visited the III Physics Institute at the University of Goettingen and its director, Professor Manfred R. Schroeder, with the intent of investigating aspects of spatial hearing most relevant to design of concert halls. Peter Damaske and I were interested in explaining the subjective diffuseness of sound fields using the interaural crosscorrelation function (IACF). The magnitude (maximum value) of the IACF defined by IACC is an indication of how subjectively diffuse a given sound will be perceived. We played sounds into a room using a multiple loudspeaker-reproduction system and recorded them at the two ears of a dummy head (Damaske and Ando, 1972). Because the IACC was known to be an important determinant in the horizontal localization of sounds, we thought it likely then that it would also play an important role for subjective diffuseness. Two years later, in 1974, a comparative study of European concert halls performed by Schroeder, Gottlob, and Siebrasse showed that the IACC was an important factor for subjective preferences and established a consensus across individuals.

In early 1975 at Kobe University, we had a special aural experience. By optimizing the horizontal direction and the delay time of a single reflection of a speech signal through successive adjustment, we were able to achieve a superior sound field. A loudspeaker in front of a single listener reproduced the direct sound. In the optimal setup, we found that the angle of the single reflection measured from the front was 30 degrees in the horizontal plane, the single reflection delay time was about 20 ms, and the amplitude of the reflection was equal to that of the direct sound. For our continuous speech signal, the effective duration was also close to this same value of 20 ms. The effective duration ( $\tau_e$ ) of a signal is a measure of its temporal coherence, the duration over which the signal is self-similar and less degraded by room reverberations. Subjective preference therefore seemed to us to be characterized in terms of both temporal and spatial factors. These working hypotheses were reconfirmed in the fall of 1975 while the author was an Alexander-von-Humboldt Fellow in Goettingen (Ando, 1977; Ando and Gottlob, 1979). During that period (1975–1977), we

also were able to explain sound colorations produced by single reflections in terms of the envelope of the autocorrelation function (ACF) of the source signal (Ando and Alrutz, 1982).

In 1983, a method of calculating subjective preference at each seat in a concert hall was published (Ando, 1983). Soon after, a theory of designing concert halls was formulated based on a model of the auditory system. We assumed that some aspects depended on the auditory periphery (the ear), whereas others depended on processing in the auditory central nervous system (Ando, 2003). The model takes into account both temporal factors and spatial factors that determine the subjective preference of sound fields (Ando, 1985). The model consists of a monaural autocorrelation (ACF) mechanism and an interaural crosscorrelation (IACF) mechanism for binaural processing. These two representations are used to describe spatial hearing operations that we presume to be taking place in several stations in the auditory pathway, from the auditory brainstem to the hemispheres of the cerebral cortex.

Once one has an explicit model of subjective preference alongside a model of auditory perception, then rational design of spaces for listening becomes possible. Special attention is paid in this volume to computing optimal individual preferences by adjusting the weighting coefficient of each temporal and spatial factor. “Subjective preference” is important to us for philosophical and aesthetic as well as practical, architectural acoustics reasons. We consider preference as the primitive response of a living creature that influences its direction and judgment in the pursuit of maintaining life: of body, of mind, and of personality (Ando, 2004).

## **1.2 Auditory System Model for Temporal and Spatial Information Processing**

We have attempted to learn how sounds are represented and processed in the cochlea, auditory nerve, and in the two cerebral hemispheres in order to develop a theory of spatial hearing for architectural acoustics that is grounded in the human auditory system. Once effective models of auditory processing are developed, designs for concert halls can proceed in a straightforward fashion, according to the guidelines derived from the model. In addition, understanding the basic operations of the auditory system may lead to development of automatic systems for recognizing speech and analyzing music, as well as identifying environmental noise and its subjective effects.

In our attempts to understand how the auditory system works, we have sought the neural response correlates of auditory percepts that are most relevant to architectural acoustics. Table 1.1 summarizes the different perceptual attributes associated with sound sources and sound fields as well as several dimensions of listening preferences. The table lists the perceptual attribute, its acoustic correlate, the nature of the acoustic and neuronal representations thought to be involved, the features in these representations that subserve the attribute, as well as related signs of neuronal response and the presumed locations of their generation.

**Table 1.1** Summary of temporal and spatial sensations most relevant to architectural acoustics with their acoustical and neurophysiological correlates. For each sensation, the table lists the internal correlational representation thought to be involved, the features of the representation that encode that specific perceptual attribute, the presumptive location of the neural generators, and their corresponding neurophysiological observables. Loci listed in parentheses are hypothetical processing locations for these attributes

Quality or attribute	Acoustic correlate	Representation	Main factor	Locus	Observable(s)	Section(s) of volume
<b>Temporal sensations and preference</b>						
Periodicity pitch	Fundamental frequency	ACF	$\tau_1$ and $\phi_1$ extracted from ACF	brainstem (left hemisphere)	Single units, ABR	5.1, 6.2
Loudness	Sound pressure level	ACF	$\tau_1, \tau_c$ in the condition of constant LL	brainstem (left hemisphere)	ABR	4.1, 4.2, 5.1, 6.4
Duration sensation	Physical duration	ACF	Signal duration D and $\tau_1$ extracted from ACF	brainstem (left hemisphere)	ABR	4.1, 6.5
Timbre	Spectral composition	ACF	$W_{\phi(0)}$ extracted from ACF	Left hemisphere	Single unit	5.1
Preferred first reflection, $[\Delta t_1]_p$	Distance of nearest reflecting surface	ACF	$\tau_c$ extracted from ACF	Left hemisphere	SVR, EEG, MEG	3.1, 4.3, 4.4
Preferred reverberation time, $[T_{sub}]_p$	Volume of a room and absorption	ACF	$\tau_c$ extracted from ACF	Left hemisphere	EEG	3.2, 4.3
<b>Spatial sensations and preference</b>						
Localization (azimuth)	Source position	IACF	$\tau_{IACC}$ and IACC extracted from IACF	brainstem (right hemisphere)	ABR	4.1, 7.1
ASW	Location of reflectors and frequency component	IACF	IACC, $W_{IACC}$ , LL	brainstem (right hemisphere)	ABR, SVR	4.1, 4.2, 7.2
Subjective diffuseness	Location of reflectors	IACF	IACC, LL	brainstem (right hemisphere)	ABR, SVR	4.1, 4.2, 7.3
Preferred listening level, $[LL]_p$	Distance from the source to receiving position, volume of a room and absorption	IACF	LL	brainstem (right hemisphere)	ABR, SVR	3.2, 4.1, 4.2
Preferred IACC: a small value	Location of reflectors	IACF	IACC	brainstem (right hemisphere)	ABR, SVR	3.2, 4.1, 4.2



Temporal sensations are those auditory qualities that are most directly related to temporal patterns and time delays, while spatial sensations are auditory qualities most directly related to spatial aspects of sound and sound fields. Temporal sensations naturally lend themselves to autocorrelation-like acoustical descriptions and putative neuronal representations, while spatial sensations lend themselves to interaural crosscorrelation descriptions of sound fields and internal representations. It is striking how much of what we hear can be accounted for in terms of features of these two kinds of representations, the monaural autocorrelation function (ACF) and the binaural interaural crosscorrelation function (IACF). It is perhaps equally surprising that observable neuronal correlates for all these perceptual attributes can be found in the summed activity of large numbers of auditory neurons.

It is remarkable that the temporal discharge patterns of neurons at the level of the auditory nerve and brainstem include sufficient information to effectively represent the ACF of an acoustic stimulus. Mechanisms for the neural analysis of interaural time differences through neural temporal crosscorrelation operations and for analysis of stimulus periodicities through neural temporal autocorrelations were proposed more than 50 years ago (Jeffress, 1948; Licklider, 1951; Cherry and Sayers, 1956). Since then, a host of electrophysiological studies recording from single neurons and neural populations has more clearly elucidated the neuronal basis for these operations. Binaural crosscorrelations are computed by axonal tapped delay transmission lines that feed into neurons in the medial superior nucleus of the auditory brainstem that act as coincidence detectors (Colburn, 1996). If one examines the temporal patterning of discharges in the auditory nerve (Secker-Walker and Searle, 1990), one immediately sees the basis for a robust time-domain representation of the acoustic stimulus. Here the stimulus autocorrelation is represented directly in the interspike interval distribution of the entire population of auditory nerve fibers (Cariani and Delgutte, 1996a,b). This autocorrelation-like neural representation subserves the perception of pitch and tonal quality (aspects of timbre based on spectral contour) (Meddis and O'Mard, 1997; Cariani, 1999).

Higher in the auditory pathway, at brainstem, midbrain, and cortical levels, neural correlates of both temporal and spatial sensations can also be found. In our own laboratory in Japan we found these correlates at several levels of the auditory pathway – in auditory brainstem responses (ABRs) and in several measures of cortical response (SVRs, EEG, MEG).

We have observed neural response correlates of temporal aspects of sensation, such as pitch or the missing fundamental (Inoue et al., 2001), loudness (Sato et al., 2001), and in addition duration sensation (Saifuddin et al., 2002). These are well described by the temporal factors extracted from the ACF (Ando et al., 2000; Ando, 2002). We found that the temporal factors of sound fields such as  $\Delta t_1$  and  $T_{\text{sub}}$  are mainly associated with left hemisphere responses (Ando et al, 1987; Ando, 1992; Ando and Chen, 1996; Chen and Ando, 1996; Soeta et al., 2002). In contrast, aspects of sound fields involving spatial sensations such as subjective diffuseness (Ando and Kurihara, 1986) and apparent source width (ASW) as well as subjective preferences (Sato and Ando, 1996; Ando et al., 2000) are mainly associated with right hemisphere responses (Ando and Hosaka, 1983; Ando, 1985, 1992, 1998; Ando et al.,

1987; Sato et al., 2003). Some qualities involve both spatial and temporal sensations. For example, the timbre of the sound field in an existing hall was investigated by a dissimilarity judgment in relation to both temporal and spatial factors (Hotehama et al., 2002).

Observable correlates of spatial sensations can also be found at many levels of auditory processing, in evoked auditory brainstem responses (ABRs) and slow vertex responses (SVRs). ABRs are short latency (0–10 ms) averaged evoked electrical potentials that reflect the successive excitation of neuronal populations in the ascending auditory pathway (Section 4.1). Waves I–V in ABRs are produced by near-synchronous firing of large numbers of neurons in the cochlea, brainstem, and midbrain. The later waves in this response reflect the contributions of binaural neural processing stations in the auditory brainstem. We found that the magnitudes of Wave IV on either side ( $IV_{1,r}$ ) varied with the sound energies present at the entrances of the two ears. We also observed that the magnitude of wave V in the ABR, which reflects the amount of synchronized neural activity at the level of the midbrain, covaries with the magnitude (IACC) of the interaural crosscorrelation function (IACF) (Ando et al., 1991). Slow vertex responses (SVRs) are longer latency (10–500 ms) averaged auditory-evoked responses that are computed from scalp electroencephalogram (EEG) signals. These potentials reflect the electrical contributions of neuronal populations mainly in auditory cortical regions. We found SVR correlates of subjective diffuseness and apparent source width that are related to the interaural correlation magnitude (IACC) in the latencies of the negative right hemisphere cortical response peak  $N_2$  (260–300 ms latency, Fig. 4.9). Systematic shifts in the latencies and/or amplitudes of SVR peaks were also observed for first reflection time ( $\Delta t_1$ ), sensation level (SL), and interaural correlation magnitude (IACC) (Section 4.2).

We also carried out a long series of experiments directed at finding observable neural correlates of subjective preferences of sound fields. Overall subjective preference has been well described by four orthogonal acoustic factors, two of them temporal and two spatial. The two temporal factors are (1) the initial time delay gap between the direct sound and the first reflection ( $\Delta t_1$ ) and (2) the subsequent reverberation time ( $T_{sub}$ ). The two spatial factors are (1) listening level (LL) and (2) the maximum magnitude of IACF (IACC), the latter being an index of the sound's apparent spatial compactness in azimuth. The SVR- and EEG-based neural correlates of the two temporal factors were found to be associated with the left hemisphere, while the two spatial factors were associated with the right hemisphere. For example, we reconfirmed by magnetoencephalography (MEG) that the left cerebral hemisphere is associated with first reflection time  $\Delta t_1$ . The right cerebral hemisphere was activated by the typical spatial factors, i.e., the magnitude of interaural crosscorrelation (IACC) (Sato et al., 2003). Not surprisingly, the neural correlates of listener preferences for these temporal and spatial factors were also found in the corresponding hemispheres. Neural correlates for preferences typically involved prolongation of alpha wave activity under preferred listening conditions as seen in both EEG and MEG. More specifically, we found that the duration of coherent, repetitive alpha wave activity (effective duration of the ACF of the response signal)

directly corresponds to how much a given stimulus is preferred, i.e., the scale value of individual subjective preference (Soeta et al., 2002; 2003). This evidence ensures that the basic theory of subjective preference can be applied to each individual preference as well (Ando, 1998).

Because individual differences of subjective preference for spatial factors are small (nearly everyone has the same basic preferences), given the type of program material involved, we can a priori determine the optimal architectural space form of the room by taking common preferences into account. The temporal factors that involve successive delays produced by sets of reflective surfaces are closely related to the dimensions of the specific concert hall under design. These dimensions can potentially be altered to optimize the space for specific types of music, such as organ music, chamber music or choral works (Ando, 1998).

# 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.

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

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

### 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

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

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

$$\Phi_p(\tau) = \int_{-\infty}^{+\infty} P_d(\omega)e^{j\omega\tau} d\omega \quad (2.5)$$

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

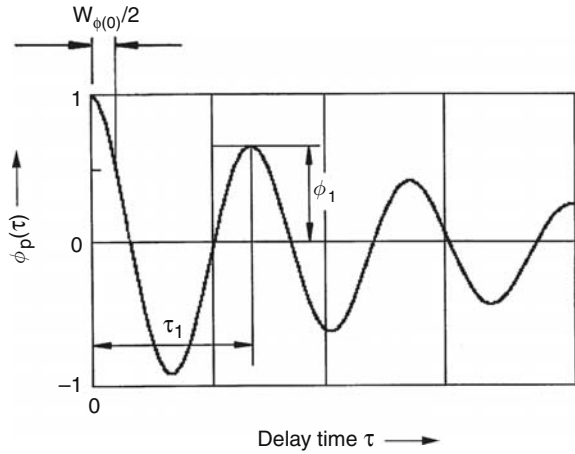
Thus, the ACF and the power density spectrum mathematically contain the same information. The normalized ACF is defined by

$$\phi_p(\tau) = \Phi_p(\tau)/\Phi_p(0) \quad (2.7)$$

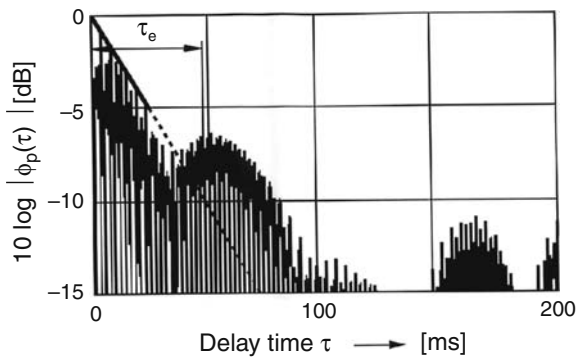
There are four significant items that can be extracted from the ACF:

- (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 an even function.
- (3) Fine structure, including peaks and delays: For instance,  $\tau_1$  and  $\phi_1$  are the delay time and the amplitude of the first peak of the ACF,  $\tau_n$  and  $\phi_n$  being the delay time and the amplitude of the n-th peak. Usually, there are certain correlations between  $\tau_1$  and  $\tau_{n+1}$  and between  $\phi_1$  and  $\phi_{n+1}$ , so that significant factors are  $\tau_1$  and  $\phi_1$ .
- (4) The effective duration of the envelope of the normalized ACF,  $\tau_e$ , 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,  $\tau_1$  and  $\phi_1$ , as features of the normalized autocorrelation function (NACF).  $\tau_1$  is the time delay associated with the first ACF peak.  $\phi_1$  is the relative amplitude of the first peak



**Fig. 2.2** Determination of the effective duration of running ACF,  $\tau_e$ . Effective duration of the normalized ACF is defined by the delay  $\tau_e$  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  $\tau_e$  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.

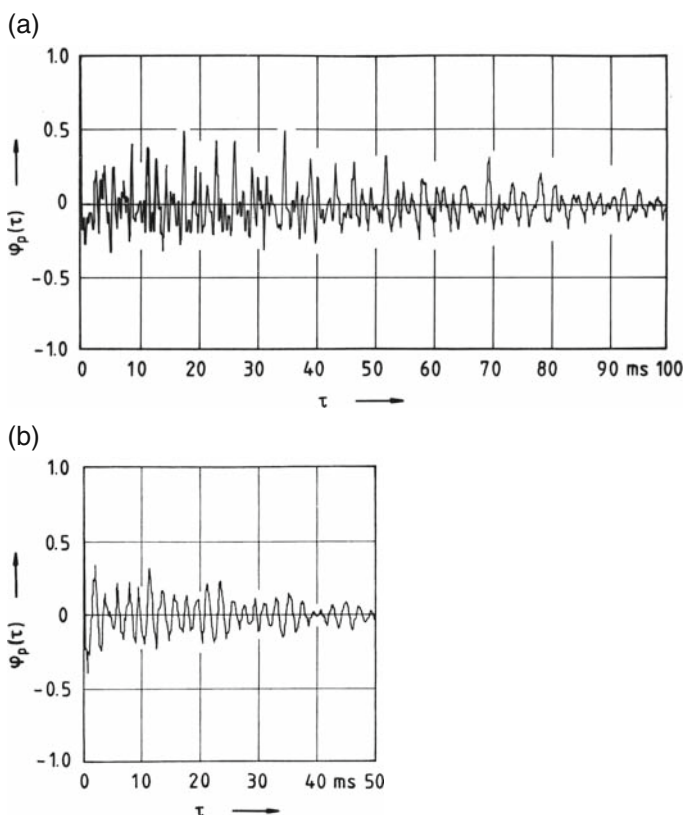
**Table 2.1** Music and speech source signals used and their effective duration of the long-term ACF,  $\tau_e$  measured in the early investigations (Ando, 1977; Ando and Kageyama, 1977), and the minimum value of running ACF,  $(\tau_e)_{\min}$  (Ando, 1998)

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

<sup>1</sup>The left channel signals of original recorded signals (Burd, 1969) were used.

<sup>2</sup>Values of  $\tau_e$  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.

<sup>3</sup>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)

Loudness also has ACF correlates. When  $p'(t)$  is measured in reference to the pressure  $20 \mu\text{Pa}$  leading to the level  $L(t)$ , the sound-pressure level  $L_{\text{eq}}$  is defined by

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

corresponding to

$$10 \log \Phi_p(0) \quad (2.9)$$

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  $\tau$  is calculated as (Taguti, and Ando, 1997)

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

where

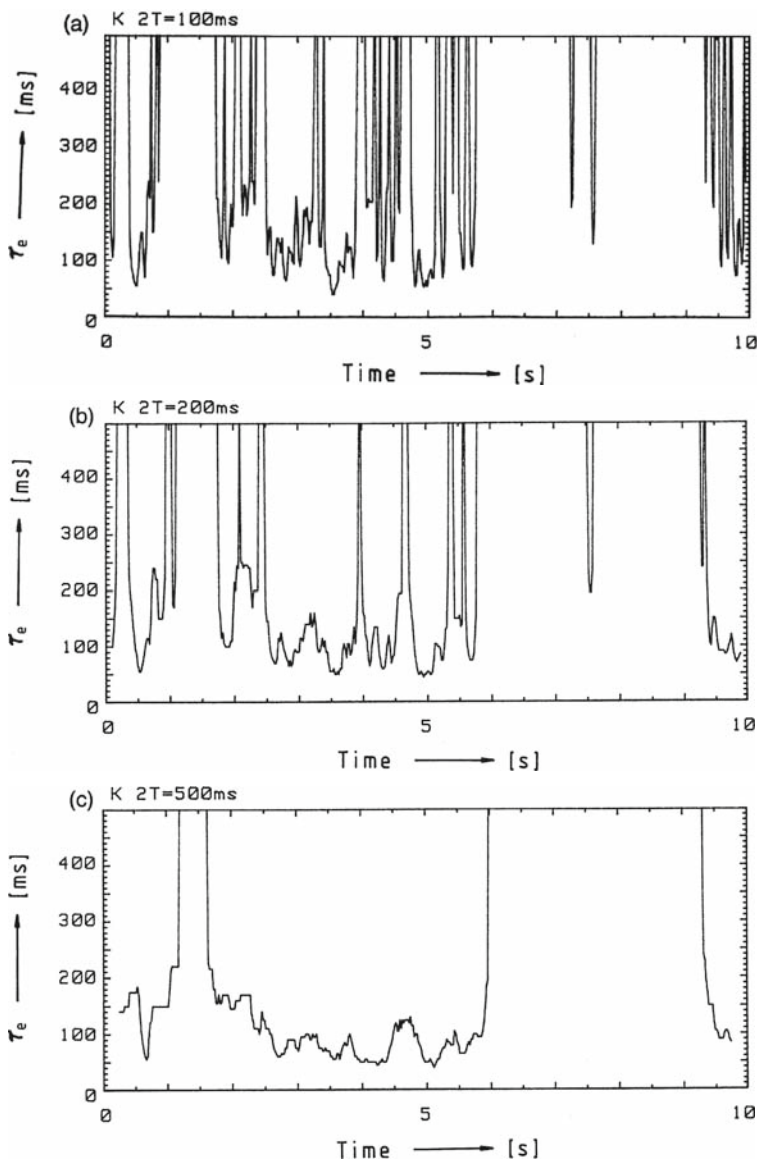
$$\Phi_p(\tau; t, T) = \frac{1}{2T} \int_{t-T}^{t+T} p'(s)p'(s + \tau) ds \quad (2.11)$$

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

To demonstrate a procedure for obtaining the effective duration of the analyzed short-time ACF, 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  $\tau_e$  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 auto-correlation using different temporal integration windows ( $2T$ ). Temporal stepsize was 100 ms. (a)  $2T=100$  ms. (b)  $2T=200$  ms. (c)  $2T=500$  ms. (d)  $2T=1$  s. (e)  $2T=2$  s. (f)  $2T=5$  s

Figure 2.5a–f show the moving  $\tau_e$  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  $\tau_e$ , the most active part of music, containing important informa-

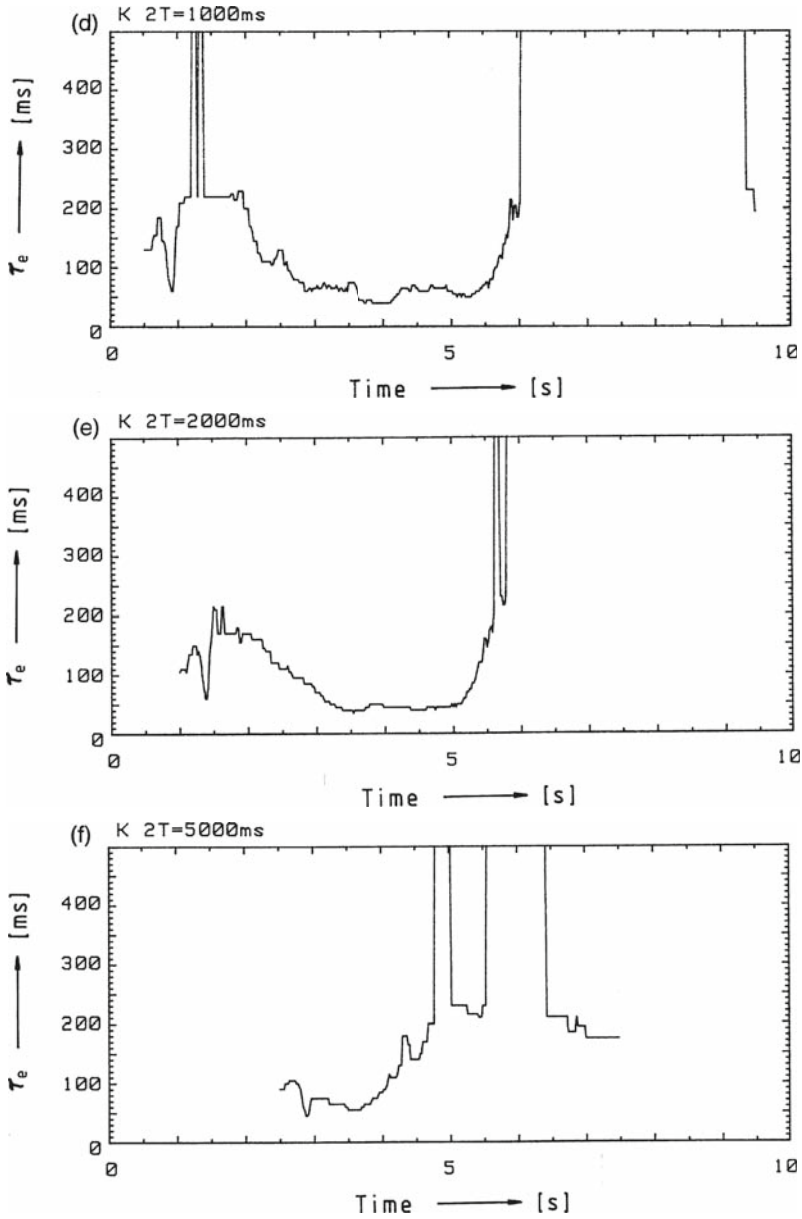


Fig. 2.4 (continued)

tion and influencing subjective preference, 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

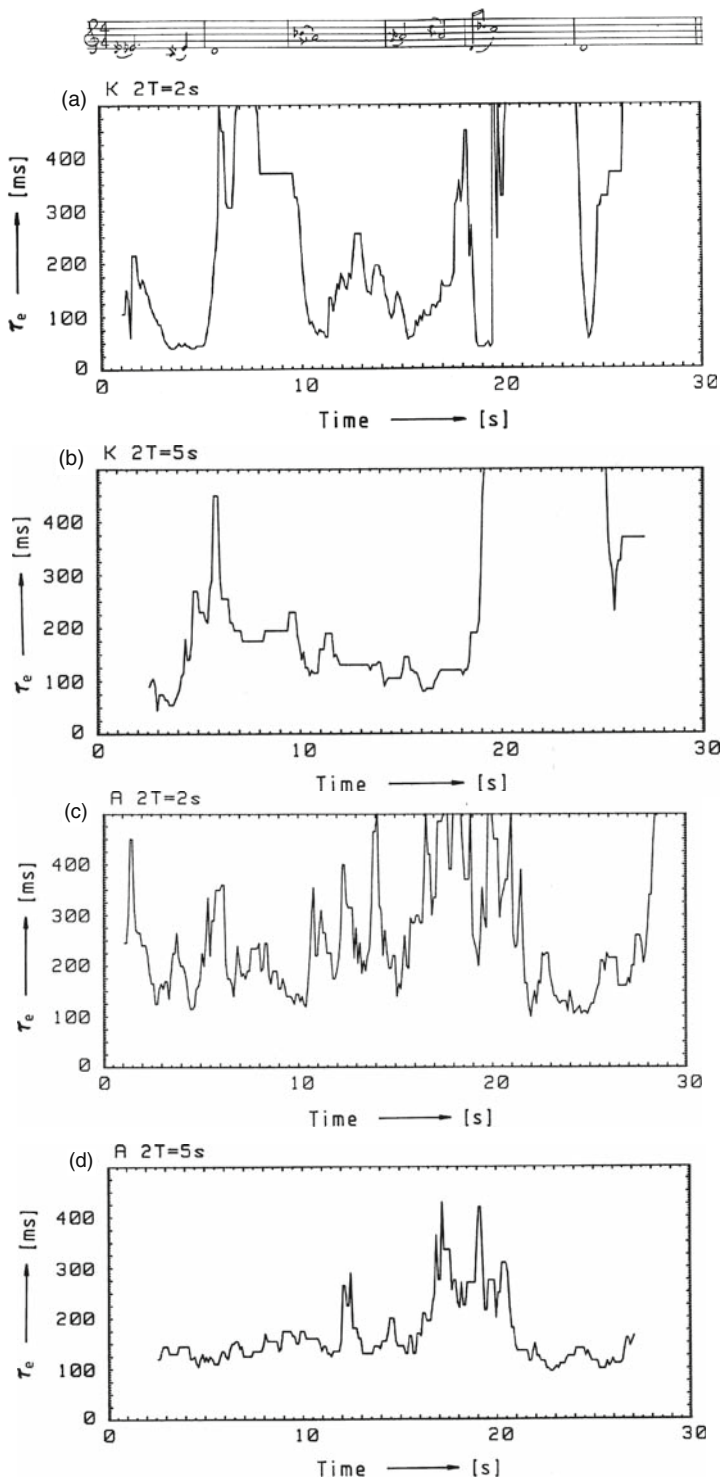
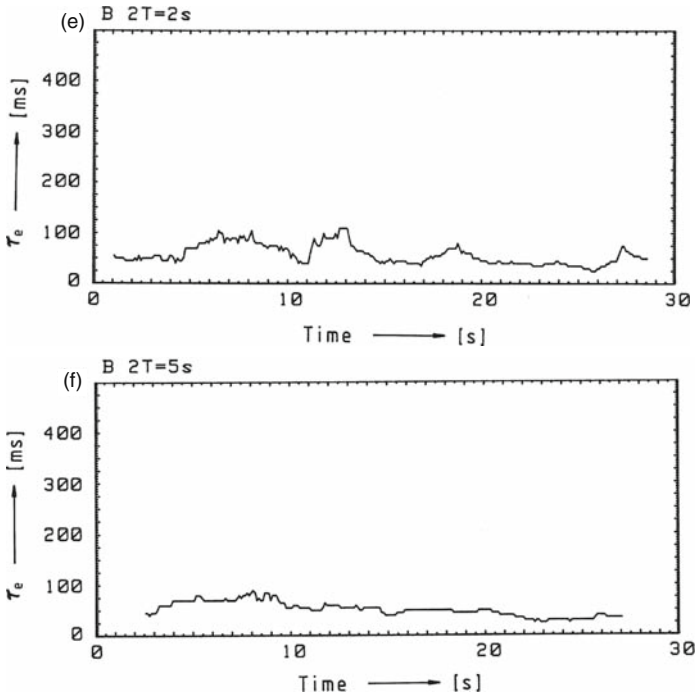
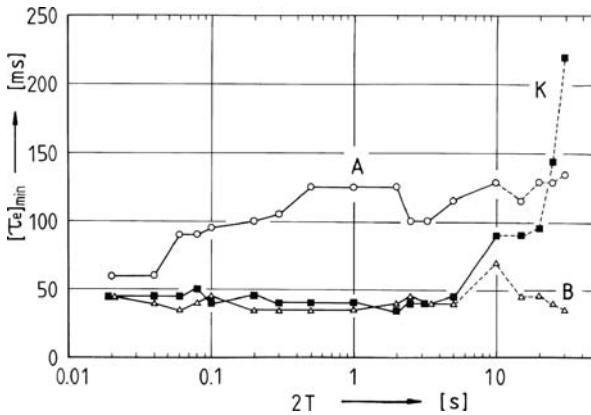


Fig. 2.5 (continued)



**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  $2T$ . (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,

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

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,

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

where  $n$  denotes the number of reflections with a horizontal angle,  $\xi_n$  and elevation  $\eta_n$ ;  $n = 0$  signifies the direct sound ( $\xi_0 = 0$ ,  $\eta_0 = 0$ ),

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

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,  $\Delta t_n$  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

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

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,  $A_1, A_2, \dots$ , are determined by the pressure decay due to the paths  $d_n$ , such that

$$A_n = d_0/d_n \quad (2.15)$$

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  $\Delta t_1, \Delta t_2, \dots$  relative to that of the direct sound, which are given by

$$\Delta t_n = (d_n - d_0)/c \quad (2.16)$$

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  $\Delta t_n$  in the manner of

$$\Delta t_n = d_0(1/A_n - 1)/c \quad (2.17)$$

In addition, the initial time delay gap between the direct sound and the first reflection  $\Delta t_1$  is statistically related to  $\Delta t_2, \Delta t_3, \dots$  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  $\Delta t_1$  is regarded as a representation of both sets of  $\Delta t_n$  and  $A_n$  ( $n = 1, 2, \dots$ ).

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

$$w_n(t) = w_n(t)^{(1)} * w_n(t)^{(2)} * \dots * w_n(t)^{(i)} \quad (2.18)$$

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,  $T_{\text{sub}}$ , because  $w_n(t)^{(i)}$  includes the absorption coefficient as a function of frequency. This coefficient is given by

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

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

$$T_{\text{sub}} \approx \frac{KV}{\bar{\alpha}S} \quad (2.20)$$

where  $K$  is a constant (about 0.162),  $V$  is the volume of the room ( $\text{m}^3$ ),  $S$  is the total surface ( $\text{m}^2$ ), and  $\bar{\alpha}$  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  $i$ :

$$\bar{\alpha}S(\omega) = \sum_i \alpha(\omega)_i S_i \quad (2.21)$$

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

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

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

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

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  $\tau$  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,

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

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  $\Phi(0)_{\text{reference}}$  is the reference sound energy.

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

$$\Phi_{lr}^{(N)}(\tau) = \frac{\sum_{n=0}^N A^2 \Phi_{lr}^{(n)}(\tau)}{\sqrt{\sum_{n=0}^N A^2 \Phi_{ll}^{(n)}(0) \sum_{n=0}^N A^2 \Phi_{rr}^{(n)}(0)}} \quad (2.25)$$

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

$$IACC = |\phi_{lr}(\tau)|_{\text{max}} \quad (2.26)$$

for the possible maximum interaural time delay, say,

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

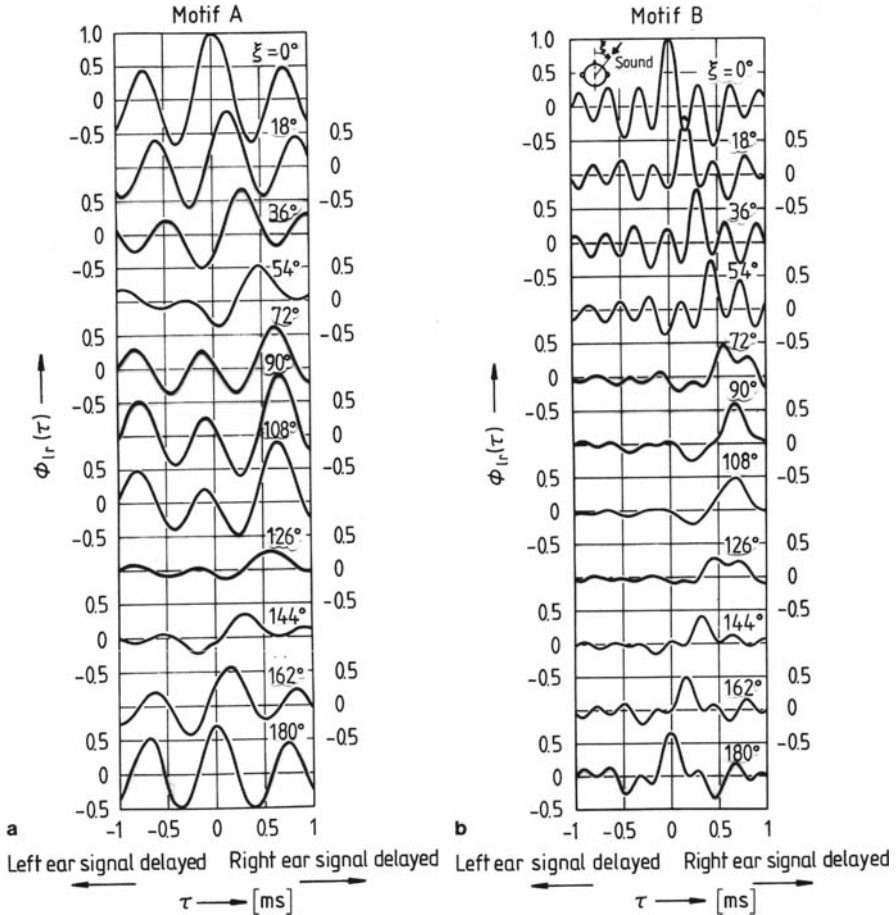
For several music motifs, the long-time IACF ( $2T = 35$  s) was measured for each single reflected sound direction arriving at a dummy head (Table D.1, Ando, 1985). These data may be used for the calculation of the IACF by Equation (2.25).

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  $\tau_{IACC}$ . Thus, both the IACC and  $\tau_{IACC}$  may be obtained at the maximum value of IACF.

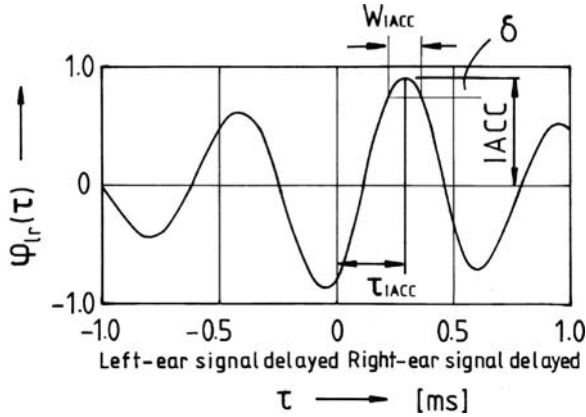
For a single source signal arriving from the horizontal angle  $\xi$  defined by  $\tau_{\xi}$ , the interaural time delay corresponds to  $\tau_{IACC}$ . When it is observed  $\tau_{IACC} = 0$  in a room, then usually a frontal sound image and a well-balanced sound field are perceived (the preferred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



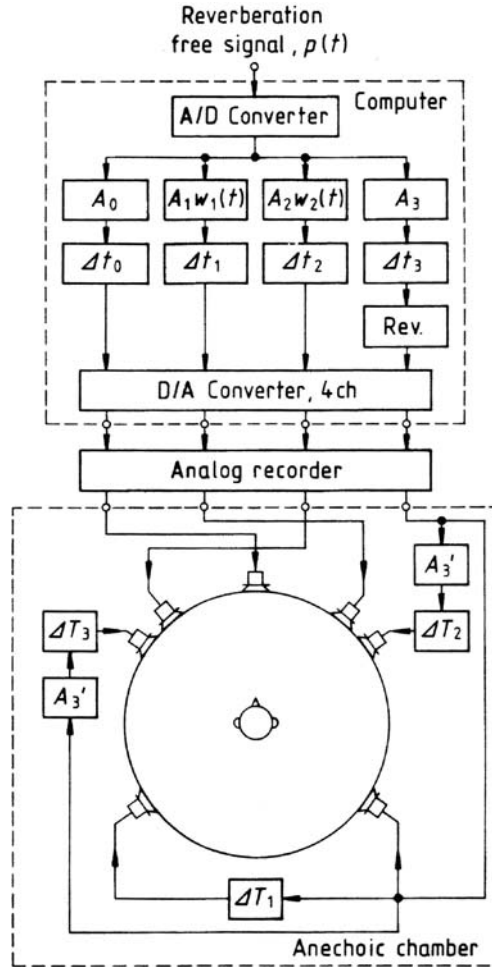
**Fig. 2.8** Definition of the three spatial factors extracted from the interaural correlation function (IACF). The interaural correlation magnitude IACC is the maximum value of the IACF. IACC is associated with the subjective diffuseness of the sound. The  $\tau_{IACC}$  is the delay at which the IACF 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,  $\tau_{IACC}$ , 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 ( $A_0 \dots A_3$ ), applying pinna-related filters ( $w_1(t)$  and  $w_2(t)$ ) that take into account the directionality of early reflections and adding corresponding delays ( $t_0 \dots \Delta t_3$ ). 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,  $\Delta T_j$  ( $j = 1, 2, 3$ ).

## Chapter 3

# Subjective Preferences for Sound Fields

Preferences are the most primitive responses of subjective attributes, because preferences are the evaluative judgments that actually steer an organism's behavior so that it can survive and propagate. Preferences guide the organism in the direction of maintaining life. In humans, these preferences are deeply related to aesthetic issues. The experiments described throughout this book have mainly used the paired-comparison test (PCT) to assess subjective preferences for sound fields (Thurstone, 1927; Mosteller, 1951; Gullikson, 1956; Torgerson, 1958; for assessment of individual preferences, see Ando, 1998). The paired comparison is the simplest and most accurate method. It permits both inexperienced and experienced listeners to participate, making the method appropriate for a wide range of applications.

Our experimental observations of listener-preferred conditions for temporal and spatial features of the sound field are first presented, and then a theory of subjective preference is derived based on four orthogonal factors [Table 3.1, (Ando, 1983, 1985, 1998)]. These factors are (1) binaural listening level (LL), (2) delay time of the first reflection ( $\Delta t_1$ ), (3) subsequent reverberation time ( $T_{\text{sub}}$ ), and (4) magnitude of the interaural crosscorrelation function (IACC). The calculation of subjective preference at each listener's position in Symphony Hall in Boston, Massachusetts is

**Table 3.1** Four orthogonal factors of the sound field identified (Ando, 1983, 1985, 1998)

Orthogonal factor	Explanation
(1) LL ( <i>Binaural and spatial factor</i> )	Binaural listening level (dB). When the sound pressure level (SPL) at two ear entrances is identical, then this may be replaced by the monaural listening level.
(2) $\Delta t_1$ ( <i>Monaural and temporal factor</i> )	Delay time of the first reflection after the direct sound (s). Note that delay times of the second, third, . . . are mutually dependent.
(3) $T_{\text{sub}}$ ( <i>Monaural and temporal factor</i> )	Subsequent reverberation time after the first reflection (s).
(4) IACC ( <i>Binaural and spatial factor</i> )	Magnitude of the interaural crosscorrelation, or the maximum absolute value of the IACF.

shown as an illustrative example (Section 3.4). In Chapter 6, we discuss temporal sensations of sounds, and in Chapter 7 we discuss spatial sensations of sound fields.

### 3.1 Preferred Properties for Sound Fields with Multiple Reflections

We first discuss the temporal and spatial percepts associated with the simplest sound field, which has a single reflection. The next Section 3.2 extends this analysis to more complex sound fields with multiple reflections. Preferences for particular sound field parameters were obtained by means of paired-comparison tests (PCT).

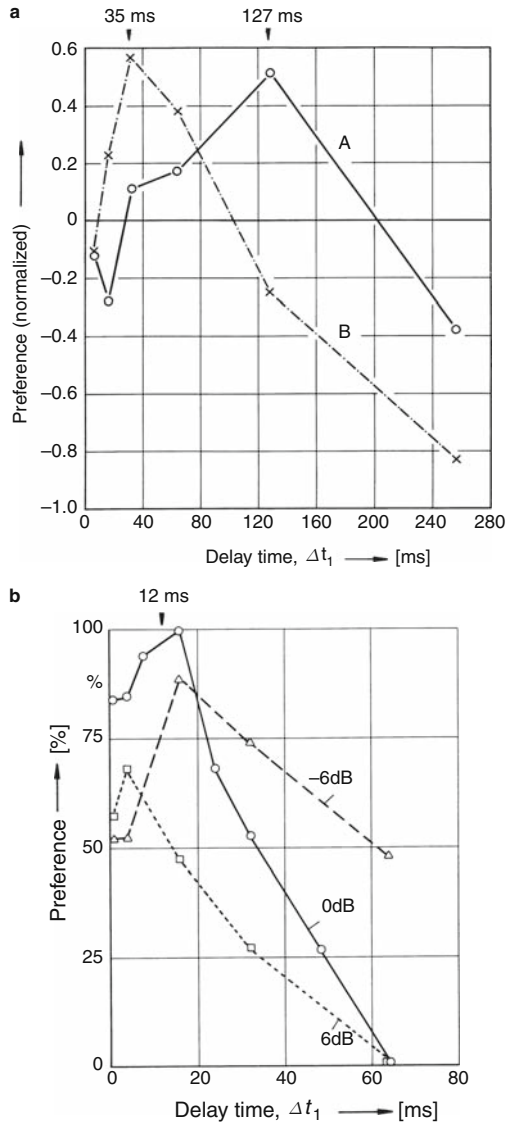
#### 3.1.1 Preferred Delay Time of a Single Reflection

In these experiments the sound field consisted of the frontal direct sound and a single reflection from a fixed direction. The reflection's horizontal angle from the front is  $\xi = 36^\circ$  and its elevation is  $\eta = 9^\circ$ . The delay time  $\Delta t_1$  was adjusted within the range of 6–256 ms. Using two different music motifs, A and B (Table 2.1), as stimuli, the paired-comparison test was performed in an anechoic chamber in Goettingen by subjects of normal hearing ability using the sound field simulation system outlined in Section 2.3 using the three front speakers. All pairs of stimulus sound field conditions were tested.

In order to ascertain the most preferred listening conditions, the total score was simply obtained by accumulating scores, giving +1 or -1 for positive and negative preference judgments, respectively. In order to obtain a normalized score, the total score was then divided by the number of stimulus sound field pairs  $S(F-1)$ ,  $S$  being the number of subjects and  $F$  the number of sound fields. Normalized scores and percentage of preference of the sound field as a function of first reflection delay time ( $\Delta t_1$ ) are shown in Fig. 3.1 (Ando, 1977; Ando and Morioka, 1981; Kang and Ando, 1985).

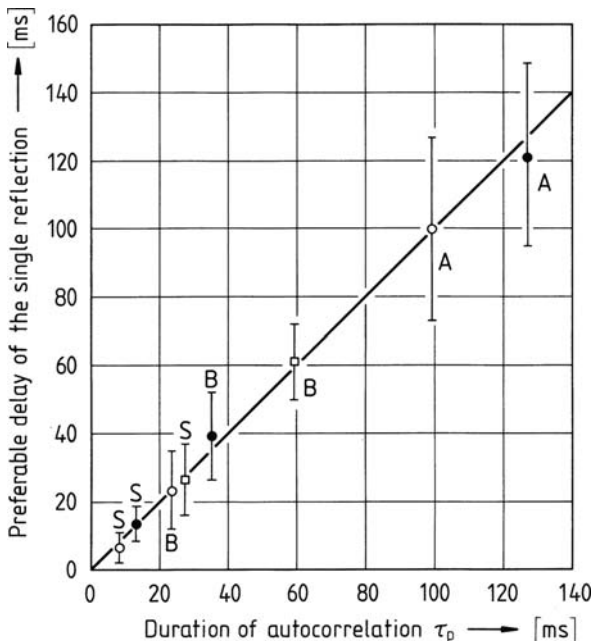
Obviously, the most preferred first reflection time, i.e., that delay with the maximum score, differs greatly between the two music motifs and continuous speech. With an amplitude of reflection,  $A_1 = 1$ , the most preferred delays are around 130 ms for music motif A, 35 ms for music motif B (Fig. 3.1a), and 16 ms for speech (Fig. 3.1b). It was found that these values correspond well to the effective duration of the ACF of the respective source signals: 127 ms (motif A), 35 ms (motif B), and 12 ms (speech), as indicated in Table 2.1. These music and speech programs sound best when the first reflection delays of the sound field are matched to the effective durations of the program material. The experimental data for preferred values of single reflections collected as a function of the duration  $\tau_p$  are shown in Fig. 3.2, where data from continuous speech signal of  $\tau_e = 12$  ms are also plotted.

**Fig. 3.1** Listener preference ratings for two musical pieces (a) and a speech segment (b) as a function of first reflection delay time of the sound field (a, b) and amplitude of the reflection (b). (a) Preference scores for two musical pieces (motifs A and B) as a function of first reflection time  $\Delta t_1$  with unity gain amplitude  $A_1=1$  for the reflected (delayed) signal. Data for 6 sound fields and 13 subjects (Ando, 1977). (A) Music motif A, *Royal Pavane* by Gibbons,  $\tau_e = 127$  ms. (B) Music motif B, *Sinfonietta*, Opus 48, III movement, by Malcolm Arnold,  $\tau_e = 35$  ms. Arrows above the plots indicate effective durations of the sound sources. (b) Percent preference for the sound field for a speech segment (speech S,  $\tau_e = 12$  ms) as a function of single reflection delay time  $\Delta t_1$  and as a parameter of amplitude  $A_1$  of the reflection in comparison with the direct sound. The reflection arrived from direction ( $\xi = 30^\circ, \eta = 0^\circ$ ). Data for 19 subjects (Ando and Kageyama, 1977)

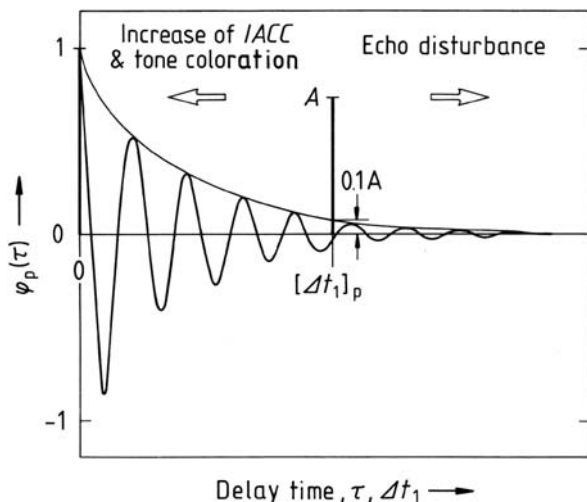


Preference decreases both when the first reflection time is either substantially longer or shorter than the effective duration of the program material (Fig. 3.3). The effective duration of the source signal’s autocorrelation function (ACF) is the delay time for which the envelope of the function has decayed to 10% of its maximal, zero-lag magnitude. This is a measure of the time window over which the signal is temporally coherent or self-similar. Reflection delays that are substantially longer than this time window create echo disturbances that degrade its clarity. Sound segments that

**Fig. 3.2** Preferred first reflection time as a function of the effective duration of the program material. Relationship between the preferred delay of a single reflection and the duration of ACF such that its envelope becomes  $0.1 A_1$  [Equation (3.1); Equation (3.3) with  $k = 0.1$  and  $c = 1$ ]. Ranges of the preferred delays are graphically obtained at 0.1 below the maximum score. A, B, and S refer to motif A, motif B, and speech, respectively. Different symbols indicate the center values obtained at the reflection amplitudes of +6 dB (○), 0 dB (●), and -6 dB (□), respectively (13–19 subjects)



are substantially different from each other interact and interfere via the convergence of direct and reflected sound paths. Thus, to preserve clarity, the first reflection time should not be substantially longer than the effective duration of the sound source. On the other hand, when the first reflection time is short, the interference of direct



**Fig. 3.3** Subjective attributes before and after the preferred delay time of reflection [ $\Delta t_1]_p (= \tau_p)$

and reflected sound in the coherent time region can produce tone coloration effects that are the result of periodicities in the pitch range that have been introduced. If the reflection times are very short, less than 1 ms, the delay can produce binaural effects that increase the interaural crosscorrelation magnitude (IACC), and thus decrease listening satisfaction.

The preferred first reflection time also changes as a function of the relative amplitude of the reflected sound vis-à-vis the direct path (Fig. 3.1b). The value of the ACF at zero delay is  $A_1$ , so its value at  $\tau_e$  becomes  $0.1A_1$ . Thus,  $\tau_p = \tau_e$  only when  $A_1 = 1$  (0 dB re:direct level). When the envelope of the ACF is exponential, then the preferred delay is expressed approximately by (Ando, 1985)

$$\tau_p = [\Delta t_1]_p \approx (1 - \log_{10} A_1) \tau_e \quad (3.1a)$$

It is worth noting that the amplitude of reflection  $A_1$  relative to that of the direct sound in Equation (3.1) should be measured by the most accurate method, i.e., the square root of the ACF at the zero-delay origin:  $[\Phi(0)]^{1/2}$  of the reflection relative to that of the direct sound.

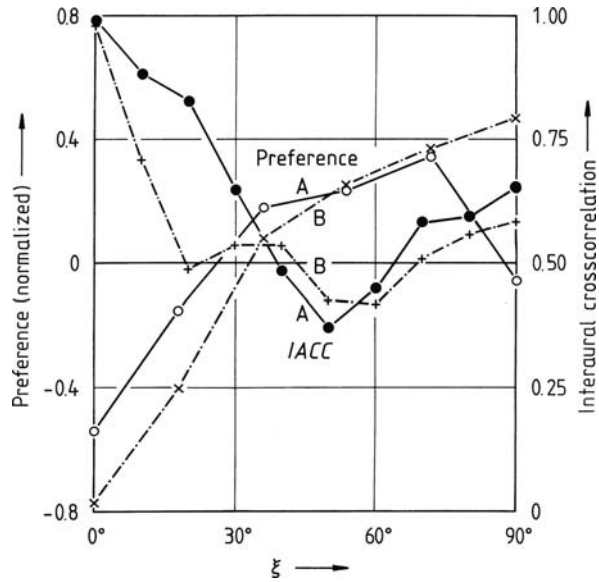
### 3.1.2 Preferred Horizontal Direction of a Single Reflection

In the experiment that estimated the preferred direction of a single reflection, the delay time of the single reflection was fixed at 32 ms. The direct sound was presented by a loudspeaker positioned at  $\xi = 0^\circ$  ( $\eta = 27^\circ$ ) and the second, reflected sound was presented by the loudspeaker located at five equally spaced angular positions that ranged from  $\xi = 18^\circ$ – $90^\circ$ , with ( $\eta = 9^\circ$ ).

Results of the preference test are shown in Fig. 3.4. No fundamental differences were observed between the curves for two music motifs with respect to changes in horizontal direction in spite of the great difference between their values of  $\tau_e$ . The preferred score increases roughly with a decreasing IACC. The correlation coefficient between the preference score and the IACC is 0.8 ( $p < 0.01$ ). The score with motif A at  $\xi = 90^\circ$  drops to a negative value, indicating that the lateral reflections, coming only from around  $90^\circ$  are not always preferred. The figure shows that there is a preference for angles greater than  $\xi = 30^\circ$ , and on an average there may be an optimum range centered on about  $\xi = 55^\circ$  (see also, Fig. 7.14 for the frequency range around 1 kHz). Similar results can be seen in the data from speech signals (Ando and Kageyama, 1977). The general conclusion is that dissimilarity between the signals arriving at the two ears is the preferred condition (Damaske and Ando, 1972; Schroeder et al., 1974).



**Fig. 3.4** Normalized preference score and interaural correlation magnitude IACC as a function of the horizontal angle of a single reflection for two extreme music motifs A and B.  $A_1=1$ . Data for 6 sound fields and 13 subjects:  $\circ$ : Preference scores for music motif A,  $\times$ : preference scores for music motif B.  $\bullet$ : Values of IACC with music motif A,  $\times$ : values of IACC with music motif B



## 3.2 Preferred Conditions for Sound Fields with Multiple Reflections

We will now discuss the more general case of sound fields with multiple reflections. Subjective preference obtained by analysis of paired comparisons can be described effectively in terms of four orthogonal properties of the sound field: two monaural temporal factors and two binaural spatial factors [Table 3.1, (Ando, 1983, 1985, 1998)]. These factors are binaural listening level (LL), timing of early reflections ( $\Delta t_i$ ), timing of subsequent reflections ( $T_{\text{sub}}$ ), and the dissimilarity of the sounds presented to the two ears.

### 3.2.1 Optimal Listening Level (LL)

The binaural listening level (LL) is the average sound pressure level at the listener's ears. This is the primary factor that influences listening preferences for sound fields in concert halls. The preferred listening level depends upon the music and the particular passage being performed.

Table 2.1 in Section 2.1.2 lists the two music sources that were used in these listening level experiments. Motif A is the slow and sombre Royal Pavane and Motif B is the fast and playful Sinfonietta. For these two musical extremes, the gross preferred levels obtained by 16 subjects were in the peak ranges of 77–80 dBA, 77–79 dBA for motif A and 79–80 dBA for motif B.

### 3.2.2 Optimal First Reflection Time ( $\Delta t_1$ )

The timing of early reflections relative to the direct sound is the second major factor that influences listening preference. 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). This approximation holds when the envelope of the ACF has an exponential decay:

$$[\Delta t_1]_p = (\log_{10} \frac{1}{k} - c \log_{10} A) \tau_e \quad (3.1b)$$

where  $k$  and  $c$  are constants that depend on the subjective attributes as listed in Table 3.2.

Here the total pressure amplitude of reflection is given by

$$A = [A_1^2 + A_2^2 + A_3^2 + \dots]^{1/2} \quad (3.2)$$

The relationship of Equation (3.1a) for a single reflection, where  $A = A_1$ ,  $k = 0.1$ , and  $c = 1$ , becomes

$$\tau_p = [\Delta t_1]_p \approx (1 - \log_{10} A_1) \tau_e \quad (3.3)$$

Later, we found that the value of  $\tau_e$  in Equations (3.1b) and (3.3) can be replaced by  $(\tau_e)_{\min}$  of the running ACF (Ando et al., 1989; Mouri et al., 2000). The minimum value of  $\tau_e$  in a music piece is generally observed in its most active part, the part with the least redundancy, the sharpest musical contrasts, and the one that usually containing the most “artistic” expressive timing information such as vibrato or accelerando in the musical flow. Echo disturbances, therefore, are easily perceived in musical segments where  $(\tau_e)_{\min}$  occurs. Even for a long music composition, musical flow can be divided into short segments, so that the minimal values of  $(\tau_e)_{\min}$  of the ACF of the whole musical piece can be taken into consideration. This value is useful for matching musical programs to concert halls, for choosing those music programs that will sound best in a given concert hall. The same is true for preferred reverberation times, as in Equation (3.4).

Methods for controlling the  $(\tau_e)_{\min}$  for vocal music performances have been discussed (Taguti and Ando, 1997; Kato and Ando, 2002; Kato et al., 2004). If vibrato is included during singing, for example, we can effectively decrease the  $(\tau_e)_{\min}$  of the music to better match the acoustics of the hall.

### 3.2.3 Optimal Subsequent Reverberation Times ( $T_{sub}$ )

Later reverberations play a significant role in sound field preferences for concert halls. In our experiments, the total amplitude  $A$  of late reflections was in the range of 1.1 and 4.1, which covers the usual conditions of a sound field in a room. For flat frequency characteristics of reverberation, the times of later reflections ( $T_{sub}$ )

**Table 3.2** Constants in Equation (3.1) related to the ACF envelope of source signals for calculating various subjective responses to the sound field with a single reflection

Subjective attributes	In Equation (3.1)		Delay time to be obtained	Amplitude examined	Source signals	Authors investigated
	k	c				
Preference of listeners	0.1	1	Preferred delay time	$7.5 \geq A_1 \geq -7.5$	Speech and music	Ando (1977)
Threshold of perception of reflection	2	1	Critical delay time	$-10.0 \geq A_1 \geq -50.0$	Speech	Seraphim (1961)
50%-echo disturbance	0.01	4	Disturbed delay time	$0 \geq A_1 \geq -6.0$	Speech	Haas (1951); Ando et al. (1973)
Coloration	$10^{-5/2}$	2	Critical delay time	$-7.0 \geq A_1 \geq -27.0$	Gaussian noise	Ando and Alrutz (1982)
Preference of alto-recorder	$2/3$	$1/4$	Preferred delay time	$-34.0 \geq A_1 \geq -10.0$	Music	Nakayama (1984)
Preference of cello	$1/2$	1	Preferred delay time	$-21.0 \geq A_1 \geq -15.0$	Music	Sato et al. (2000)

constitute one of the most important preferred conditions (Ando et al., 1989). The preferred subsequent reverberation time is expressed approximately by a constant multiple of the effective duration of the program material (Ando et al., 1982, 1983):

$$[T_{\text{sub}}]_p \approx 23\tau_e \quad (3.4)$$

Considering the fact that the effective duration value of  $\tau_e$  is obtained at the tenth percentile (or  $-10$  dB) delay of the ACF envelope of a source signal, the  $-60$  dB decay 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 a sound field [Equation (3.4)] is about four times the “reverberation time” contained in the source signal itself.

The optimal design of a building must take into account its acoustical role. A lecture and conference room should be designed for speech; an opera house should be designed primarily for vocal music but also orchestral music. For orchestral music, it is recommended that a concert hall be selected from one of two or three types of concert halls according to the effective duration of the music programs that will be performed there. For example, Symphony No. 41 by Mozart, “Le Sacre du Printemps” by Stravinsky and Arnold’s Sinfonietta have short values of  $(\tau_e)_{\text{min}}$ . On the other hand, Symphony No. 4 by Brahms and Symphony No. 7 by Buckner are more generally typical of orchestral music. Much longer values of  $(\tau_e)_{\text{min}}$  are common for pipe organ music, for example, by Bach. Thus, the most preferred reverberation time for each sound source [Equation (3.4)] can potentially play an important role for the selection of the music program to be performed.

### 3.2.4 Optimal Magnitude of Interaural Crosscorrelation (IACC)

Binaural similarity or dissimilarity of the two signals arriving at the two ears influences subjective preference. All available data with listeners of normal hearing ability indicate a negative relationship between interaural crosscorrelation magnitude (IACC) and subjective preference (i.e., it has been reconfirmed that listeners prefer somewhat dissimilar signals arriving at their two ears). This relation holds under the condition that the maximum value of the binaural, interaural crosscorrelation function (IACF) is near zero interaural delay, with the sound image directly in front and an equal balance between the sound fields for the 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  $\xi = \pm 55^\circ$ . It is obvious that the sound arriving from the median plane makes the IACC greater. Sound arriving from  $\xi = \pm 90^\circ$  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 and 2 kHz are roughly centered on  $\xi = \pm 55^\circ$  and  $\xi = \pm 36^\circ$ , respectively (see

Fig. 7.14). To realize these conditions simultaneously, geometrically uneven surfaces for the side walls of the concert hall have been proposed (Ando and Sakamoto, 1988).

### 3.3 Theory of Subjective Preferences for Sound Fields

We will now put these results into theory. Because preference involves a limited number of orthogonal acoustic factors that are implicit in the sound signals arriving at both ears (Table 3.1), the scale value of any one-dimensional subjective response may be expressed as a function of these factors

$$S=f(x_1, x_2, \dots x_l). \tag{3.5}$$

A linear scale value for subjective preference can be derived using the law of comparative judgment or paired-comparisons tests (Thurstone, 1927; Mosteller, 1951; Gullikson, 1956; Torgerson, 1958) for both groups of subjects as well as for individuals (Ando, 1998). Through a series of experiments, it has been verified that four objective acoustic factors act independently of the scale value when two of the four factors are varied simultaneously, as indicated in Table 3.3. Results obtained in a series of experiments that used different source signals indicate that the four units of scale appear to be almost constant, so we can add individual scale values to obtain the total scale value (Ando, 1983),

$$\begin{aligned} S &= f(x_1) + f(x_2) + f(x_3) + f(x_4) \\ &= S_1 + S_2 + S_3 + S_4 \end{aligned} \tag{3.6}$$

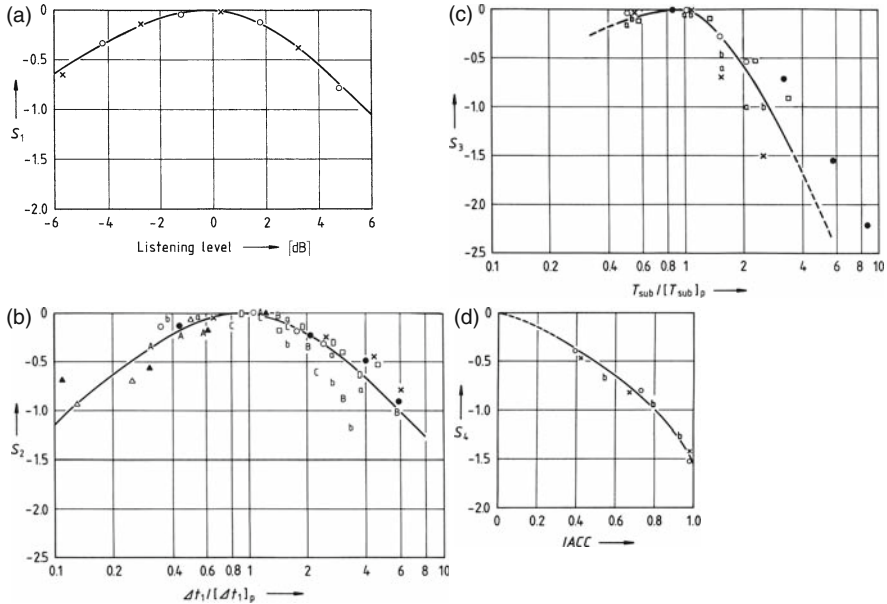
where  $S_i$ ,  $i = 1, 2, 3, 4$  is the scale value obtained relative to each objective factor. Equation (3.6) represents a four-dimensional continuum.

The dependence of the scale values on each objective factor is shown graphically in Fig. 3.5. From the nature of the scale value, it is convenient to set its value to zero at the most preferred conditions, as shown in this figure. Scale values of subjective preference obtained from other experimental series that used different music programs, yield similar results when each factor is normalized by its most preferred value. The following common formula is given:

**Table 3.3** Subjective preference tests examining independent effects changing two of four orthogonal factors of the sound field

Factors	LL	$\Delta t_1$ (SD)	$T_{sub}$	IACC
LL	–	Ando and Okada <sup>1</sup>	Not examined	<b>Test B:</b> Ando and Morioka (1981)
$\Delta t_1$ (SD)		–	<b>Test A:</b> Ando et al. (1982)	Ando and Imamura (1979); Ando and Gottlob (1979)
$T_{sub}$			–	<b>Test C:</b> Ando et al. (1983)

<sup>1</sup>Unpublished (see Ando 1998).



**Fig. 3.5** Scale values of subjective preference obtained for simulated sound fields in an anechoic chamber, as a function of four normalized orthogonal factors of the sound field. Different symbols indicate scale values obtained from different source signals (Ando, 1985). (a) As a function of listening level, LL. The most preferred listening level,  $[LL]_p = 0$  dB. (b) As a function of first reflection time  $\Delta t_1 / [\Delta t_1]_p$ . (c) As a function of later reverberation time  $T_{sub} / [T_{sub}]_p$ . (d) As a function of interaural correlation magnitude IACC. The most preferred values  $[\Delta t_1]_p$  and  $[T_{sub}]_p$  are calculated by Equations (3.3) and (3.4), respectively. Even if different signals are used, consistency of scale values as a function of the normalized factor is maintained, fitting a single curve

$$S_i \approx -\alpha_i |x_i|^{3/2}, \quad i = 1, 2, 3, 4 \tag{3.7}$$

where  $x_i$  is the normalized factor and the values of  $\alpha_i$  are weighting coefficients as listed in Table 3.4. If  $\alpha_i$  is close to zero, then a lesser contribution of the factor  $x_i$  on subjective preference is signified.

**Table 3.4** Four orthogonal factors of the sound field, and its weighting coefficients  $\alpha_i$  in Equation (3.7), which was obtained with a number of subjects

i	$x_i$	$\alpha_i$	
		$x_i > 0$	$x_i < 0$
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	–

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

$$x_1 = 20 \log P - 20 \log [P]_p \tag{3.8}$$

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

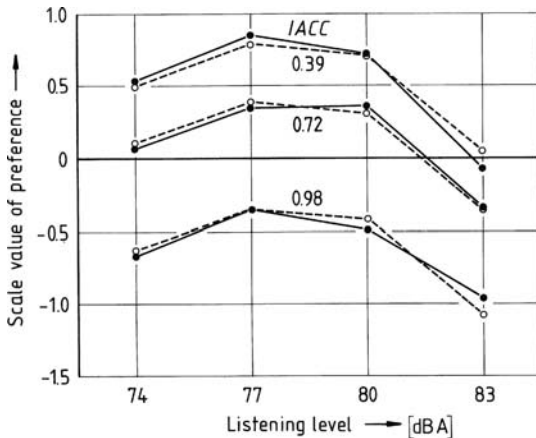
$$x_2 = \log (\Delta t_1 / [\Delta t_1]_p) \tag{3.9}$$

$$x_3 = \log (T_{\text{sub}} / [T_{\text{sub}}]_p) \tag{3.10}$$

$$x_4 = \text{IACC} \tag{3.11}$$

The values of  $[\Delta t_1]_p$  and  $[T_{\text{sub}}]_p$  may be calculated using Equations (3.1) and (3.4), respectively.

The scale value of preference has been formulated approximately in terms of the  $3/2$  power of the normalized factor, expressed in terms of the logarithm of the normalized factors,  $x_1$ ,  $x_2$ , and  $x_3$ . The remarkable fact is that the spatial binaural factor  $x_4 = \text{IACC}$  is expressed in terms of the  $3/2$  power 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 this range. Since the experiments were conducted to find the optimal conditions, this theory remains valid in the range of preferred conditions tested for the four orthogonal factors. When  $\Delta t_1$  and  $T_{\text{sub}}$  are fixed near their preferred conditions, for example, the scale value of subjective preference calculated by Equation (3.6) for the LL and the IACC is demonstrated in Fig. 3.6. Agreement between cal-



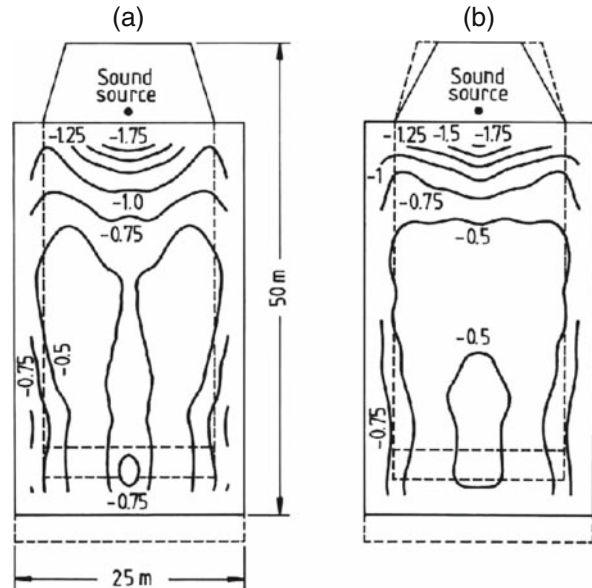
**Fig. 3.6** Scale values of subjective preference for the sound field with music motif A as a function of the listening level (LL) and as a parameter of the IACC (Ando and Morioka, 1981). (---): Calculated values based on Equation (3.6) taking the two factors into consideration; (—): Measured values

culated values and observed ones are satisfactory, so that the independence of the listening level 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).

### 3.4 Evaluation of Boston Symphony Hall Based on Temporal and Spatial Factors

As a typical example, we will consider the quality of the sound field at each seating position in an existing concert hall using the Symphony Hall in Boston, Massachusetts, as a model. Suppose that a single source is located at center stage, 1.2 m above the stage floor. Ear positions are receiving points situated at a height of 1.1 m above floor level. The 30 earliest reflections with their amplitudes, delay times, and directions of arrival at the listeners are taken into account using the image method.

Contour lines of the total scale value of preference, calculated for music motif B are shown in Fig. 3.7. The left plot (a) demonstrates the effects of the reflections from the sides on the stage in their original shape configuration. Adding angled reflecting sidewalls on the sides of the stage, as in right plot (b), may produce decreasing values of the IACC for substantial portions of the audience area, which increases the total preference value at each seat (compare left and right plots). In this calculation, reverberation time is assumed to be 1.8 s throughout the hall and the most preferred listening level,  $[LL]_p = 20 \log[P]_p$  in Equation (3.10), is set for a point on the center line 20 m from the source position.



**Fig. 3.7** Predicted effects of modified room dimensions on listener satisfaction. An example of calculating scale values with the four orthogonal factors using Equations (3.6) through (3.11). (a) Contour lines of the total scale value for Boston Symphony Hall, with original side walls on the stage. (b) Contour lines of the total scale values for the side walls optimized



In order to further test the subjective preference theory, subjective preference judgments in another existing hall (Uhara Hall in Kobe, Japan) were performed using paired-comparison tests (PCT) at each set of seats. Here source locations on the stage were varied. The theory of subjective preference with orthogonal factors was reconfirmed (Sato et al., 1997; Ando, 1998). It was also shown that the theory holds at the condition of  $\tau_{IACC} = 0$ , i.e., a source at its optimal location at center stage. For other, off-center positions, preference values decrease.

## Chapter 4

# Electrical and Magnetic Responses in the Central Auditory System

Four significant, orthogonal physical factors that describe temporal and spatial characteristics of sound fields in concert halls were discussed in the preceding chapter. The fields of physical, physiological, and psychological acoustics are deeply related to each other. If enough were known about how the brain analyzes nerve impulses from cochlea to cortex, the design of concert halls and other acoustic environments could proceed rationally, according to guidelines derived from the knowledge of these processes. This motivated us to make a sustained effort to describe important qualities of sound in terms of neural information processing in the auditory pathway and the rest of the brain.

Formulation of such a neurally grounded strategy for acoustic design has been initiated through a study of auditory-evoked electrical potentials, short-latency auditory brainstem responses (ABR) that are generated in the cochlea, brainstem, and midbrain, and longer-latency slow vertex responses (SVR) that are generated in the cerebral cortex. The ultimate goal of these experiments was to identify potential neuronal response correlates of subjective preference for the orthogonal acoustic parameters most important for the perception of sound fields. Using paired-comparison methods (Ando, 1977, 1983, 1985, 1998), we had found that particular ranges of the four factors were preferred by most listeners, such that reliable predictions of subjective preferences could be made.

In order to formulate a comprehensive model of signal processing in the central auditory system, early auditory brainstem responses (ABRs, 0–10 ms latency) were first examined to characterize signal flows in ascending auditory pathways and their possible functions at each level of processing. Then, cortical, longer-latency slow vertex responses (SVR) corresponding to subjective preferences were examined in relation to temporal and spatial factors. In the third stage, we investigated aspects of human electroencephalography (EEG) and the magnetoencephalography (MEG) responses that correspond to subjective preference. EEG and MEG signals in the alpha frequency band were analyzed using autocorrelations and crosscorrelations. Table 4.1 summarizes the acoustic factors, their corresponding percepts and preferences, and the neuronal correlates that were found.

**Table 4.1** Summary of overall argument in this chapter

Acoustic factor	Subjective response	Neuronal correlate and locus
$\Phi_1(0)$ , $\Phi_r(0)$ , $\tau_{IACC}$ , IACC	Localization in the horizontal plane	Auditory brainstem responses (ABRs)
$\Delta t_1$ , LL	Subjective preference	N <sub>2</sub> -latency in slow vertex responses (SVRs), Right hemispheric amplitude response
IACC	Subjective preference Subjective diffuseness	N <sub>2</sub> -latency in SVR Right hemispheric amplitude response
$\Delta t_1$ , T <sub>sub</sub> , IACC	Subjective preference	Alpha wave in electroencephalography (EEG) Left ( $\Delta t_1$ , T <sub>sub</sub> ) and right (IACC) hemispheric responses
$\Delta t_1$	Subjective preference	Alpha wave in magnetoencephalography (MEG) Left hemispheric response
$\tau_\varepsilon$ ( $\phi_1$ ) in the band-pass noise	Annoyance	Alpha wave in magnetoencephalography (MEG) Right hemispheric response

## 4.1 Auditory Brainstem Responses (ABRs)

The internal binaural crosscorrelation function may provide a neural representation for spatial sensations and their subjective preferences. In a series of experiments, we recorded and analyzed left and right auditory brainstem responses (ABRs) in order to investigate the nature of neural representations and signal processing in the auditory pathway.

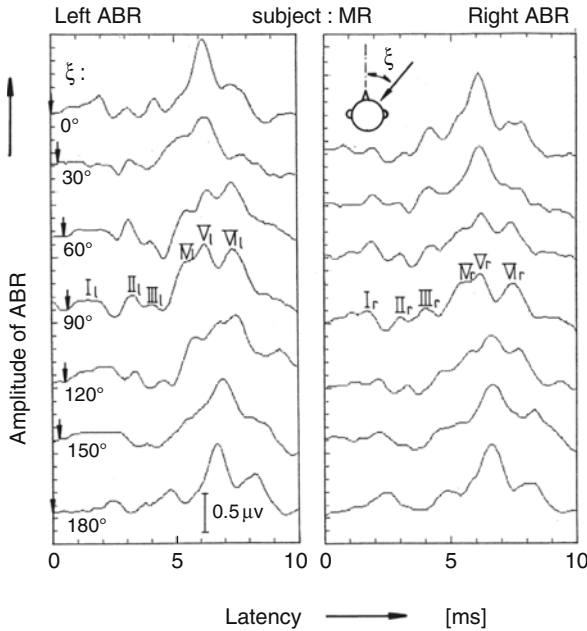
Auditory brainstem responses (ABRs) are short-latency (0–10 ms) auditory-evoked electrical potentials that are stimulus-triggered averages of the summed electrical responses of many thousands of neurons in the cochlea, brainstem, and midbrain. In effect, the ABR measures the early impulse response of the auditory system, which primarily reflects electrical potentials generated by the first stages of auditory processing. The slow vertex response (SVR) is a longer latency (10–500 ms) averaged auditory-evoked potential that, due to electrode placement on the scalp, reflects activity in later, cortical stages of auditory processing. Peaks in ABRs and SVRs reflect the synchronized component of electrical activity in dendrites, cell bodies, and axons of large populations of neurons. Most of this electrical activity is thought to be generated by synaptic currents associated with dendritic inputs. The methods used in our studies for recording human electrical potentials (ABRs, SVRs, and EEGs) are safe, non-invasive, and relatively inexpensive.

### 4.1.1 Brainstem Response Correlates of Sound Direction in the Horizontal Plane

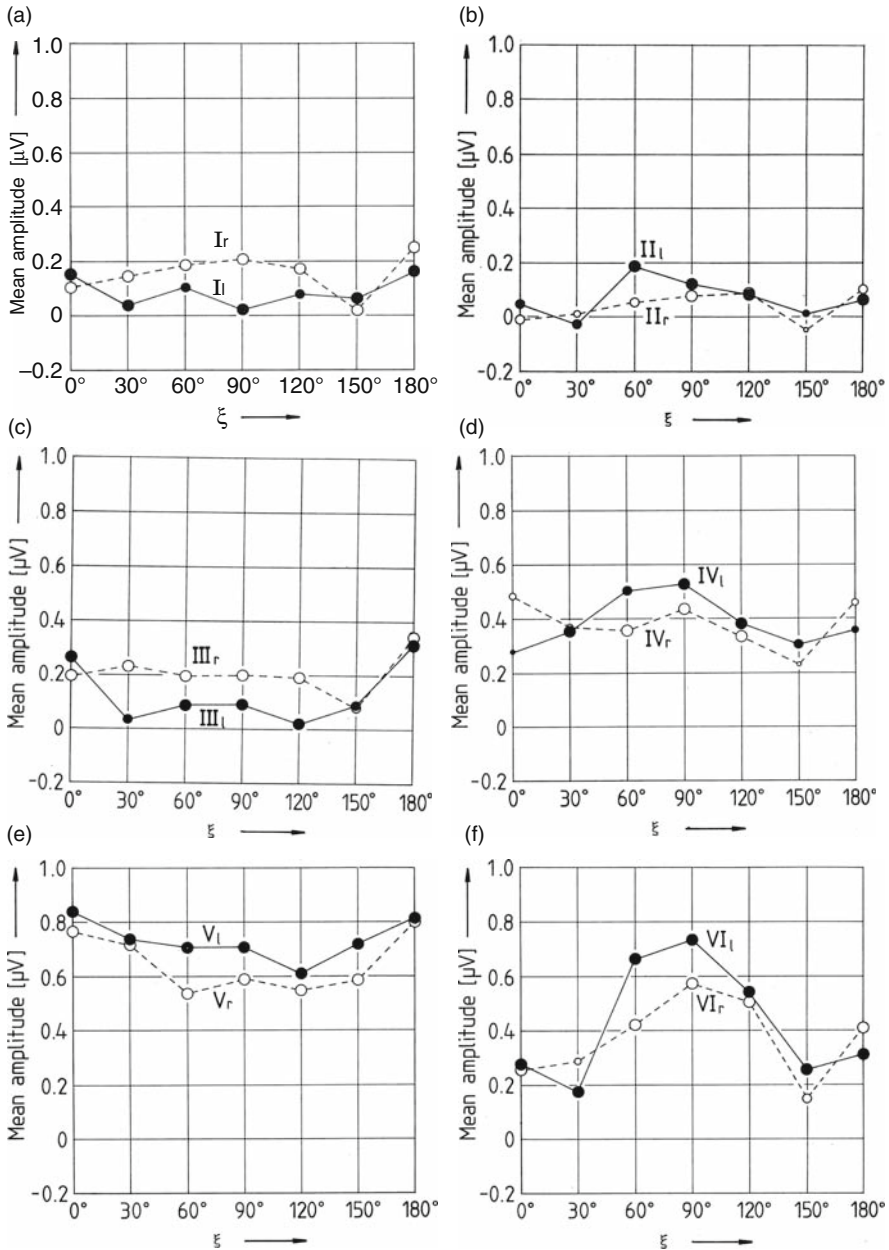
To probe the neural correlates of horizontal sound direction (azimuth), source signals  $p(t)$  of trains of clicks (50  $\mu$ s pulses) were presented every 100 ms for 200 s

(2,000 times). Signals were supplied to loudspeakers positioned at various horizontal angles (0–180°) with respect to the front of the subject, all on the subject’s right-hand side. The distance between each loudspeaker and the center of the head was kept at  $68 \pm 1$  cm. The speakers had a frequency response of  $\pm 3$  dB for 100 Hz–10 kHz.

Left and right ABRs were recorded through electrodes placed on the vertex, and the left and right mastoids (Ando and Hosaka, 1983; Ando, 1985). Typical examples of recorded ABR waveforms as a function of the horizontal angle of sound incidence are shown in Fig. 4.1 (Ando et al., 1991). It can be readily appreciated that waves I–VI differ in peak amplitude and latency as the sound location changes its angle of incidence relative to the subject’s head. Similar ABR waveforms were obtained from each of four participating subjects (males,  $23 \pm 2$  years of age). Their ABRs were averaged together and the mean amplitude of the ABR waveform peaks (waves I–VI) was computed as a function of the horizontal angle (Fig. 4.2a–f).



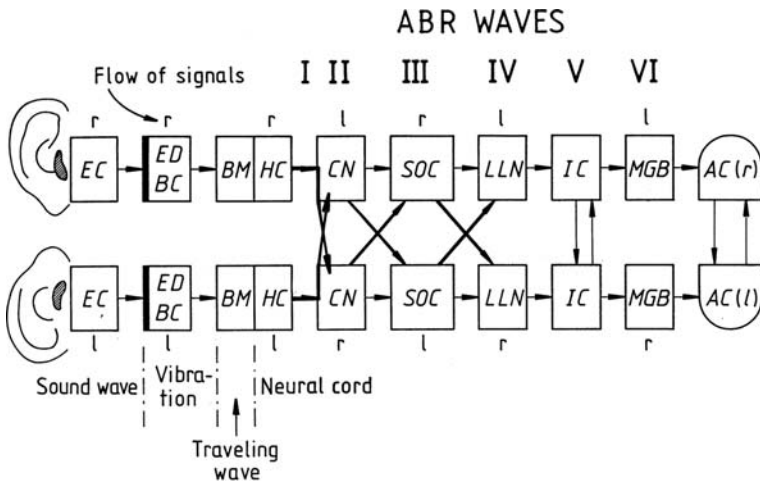
**Fig. 4.1** Examples of auditory brainstem response (ABR), as a function of response latency (0–10 ms) and horizontal angle of sound incidence. The abscissa indicates the latency of neuronal response in right and left auditory pathways relative to the time when the single pulse arrives at the right ear entrance. *Arrows* indicate the arrival time of the free-field sound at the cochlea, which depends upon the sound source location of the right hand side of the subject, and the baseline amplitude of the ABR. Roman numerals I–VI indicate successive peaks in the ABR that reflect synchronized activity at successive stations in the ascending auditory pathway. The suffix signifies the response from the left and right electrodes, which preferentially record neural responses from that side. Signals were obtained between electrodes at the vertex and left and right mastoids (Ando et al., 1991)



**Fig. 4.2** Averaged amplitudes of ABR for each wave I-VI. The four different sizes of circles indicated the number of available data from four subjects. *Filled circles*: Left ABRs; *Empty circles*: Right ABRs. (a) Wave I. (b) Wave II. (c) Wave III. (d) Wave IV. (e) Wave V. (f) Wave VI. The source location of each wave (ABR) was previously investigated for both animal and human subjects (Jewett, 1970; Lev and Sohmer, 1972; Buchwald and Huang, 1975)

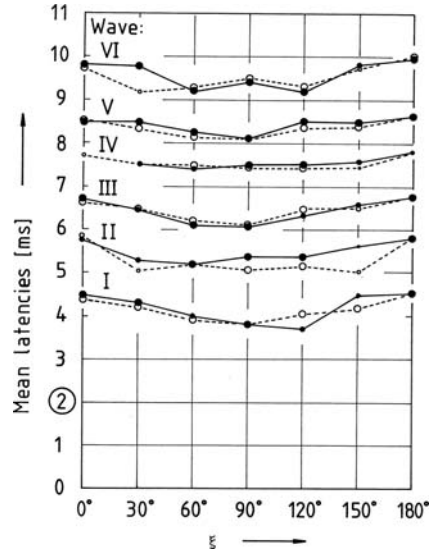
As one might expect, the average peak I amplitudes from the right electrode are greater than those from the left ( $r > l$  for angles  $\xi = 30\text{--}120^\circ$ ,  $p < 0.01$ ). This asymmetry may reflect interaural differences in sound pressure (head shadowing) produced by the source location on the right-hand side. However, this tendency is reversed for wave II for two angles  $\xi = 60\text{--}90^\circ$  ( $l > r$ ,  $p < 0.05$ , Fig. 4.2b). Yet another reversal is seen in the behavior of wave III (Fig. 4.2c), which is similar to that of wave I ( $r > l$ ,  $p < 0.01$ ). This tendency once again reverses for wave IV (Fig. 4.2d,  $l > r$ ,  $p < 0.05$ ) and is maintained further in wave VI (Fig. 4.2f,  $l > r$ ,  $p < 0.05$ ) even though absolute values are amplified.

From these patterns, it might be inferred that the flow of the left and right neural signals is interchanged three times in the binaural pathway: at the levels of the cochlear nucleus, superior olivary complex, and the lateral lemniscus, as shown in the auditory pathway schematic of Fig. 4.3. The interaction at the inferior colliculus in particular may be operative for binaural signal processing, as discussed below. In wave V as shown in Fig. 4.2e, such a reversal cannot be seen, and the relative behavior of amplitudes of the left and the right are parallel and similar. Thus, these two amplitudes were averaged and plotted in Fig. 4.6 (V symbols). For comparison, the amplitudes of wave IV (left – l and right – r) were normalized to their respective ABR amplitudes at the frontal sound incidence. These may correspond to the normalized sound pressures at the right and left ear entrances, respectively, which are also plotted.



**Fig. 4.3** High-level schematic illustration of the flow of neural signals in the ascending auditory pathway. In this schematic, the ascending pathway proceeds from left to right through successive stations. EC: external canal; ED and BC: eardrum and middle ear bone chain; BM and HC: basilar membrane, inner hair cell, and auditory nerve (not shown); CN: cochlear nucleus; SOC: superior olivary complex; LLN: lateral lemniscus; IC: inferior colliculus; MGB: medial geniculate body; AC: auditory cortex of the right and left cerebral hemispheres

**Fig. 4.4** Averaged response latencies for ABR waves I–VI as a function of horizontal angle of incidence. Data from four subjects. Sizes of circles indicate number of data points represented. The sound transmission delay between the loudspeaker and the center of the listener's head was 2 ms, indicated by the  $\alpha$  on the axis. *Filled circles*: left ABR; *empty circles*: right ABR

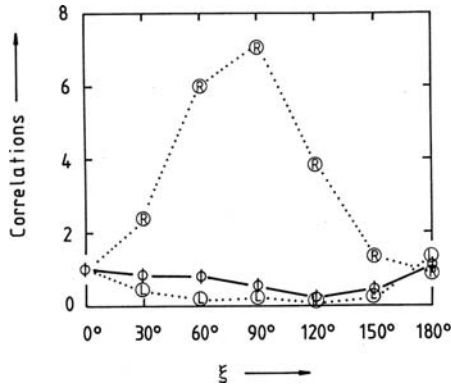


The relative latencies of ABR peaks also change with horizontal angle. Latencies of peaks of waves I through VI are computed relative to the time when the short click pulses were supplied to the loudspeaker. The shortest latencies were seen for angles around  $\xi = 90^\circ$  (Fig. 4.4). Significant differences exist ( $p < 0.01$ ) between averaged latencies at lateral locations and those in the median plane (in front  $\xi = 0^\circ$  or in back, at  $\xi = 180^\circ$ ). The differences are approximately 640  $\mu\text{s}$  on average, which corresponds to the interaural time difference created in a typical human head by a sound incident at  $\xi = 90^\circ$ . It is likely that the relative latency at waves I–III may be reflected by the interaural time difference, particularly at wave III. No significant differences could be seen between the latencies of the left and right of waves I–IV responding to such a relative factor.

#### 4.1.2 Brainstem Response Correlates of Listening Level (LL) and Interaural Crosscorrelation Magnitude (IACC)

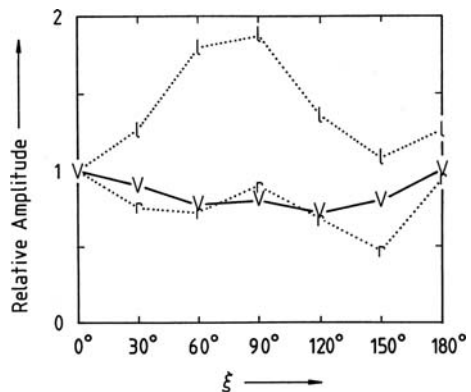
Neural auditory brainstem responses (ABRs) can be compared with interaural cross-correlation functions (IACFs) derived from acoustic measurements made at the two ears of a dummy head. We found a number of parallels between acoustic cross-correlation functions computed from sound pressure levels at the left and right ears and neural cross-correlation functions computed from ABRs that reflect the successive activation of populations of neurons in the left and right auditory pathways.

A-weighted signals were presented and free-field sound pressure measurements were taken at the two ear entrances of a dummy head as a function of the horizontal angle of the sound source. Figure 4.5 depicts the signal power at the two



**Fig. 4.5** Acoustic correlation magnitudes of sound signals arriving at the left- and right-ear entrances of a dummy head as a function of horizontal angle of incidence. Magnitudes are normalized by their respective values at  $\xi = 0^\circ$  (directly in front).  $\Phi$ : maximum interaural crosscorrelation,  $|\Phi_{lr}(\tau)|_{\max}$ ,  $|\tau| < 1$  ms. (L):  $\Phi_{ll}(0)$  measured at the left ear. (R):  $\Phi_{rr}(0)$  measured at the right ear

ears (R, L) for different angles (power is the zero-lag term of a signal’s autocorrelation function) and the maximum value of the interaural crosscorrelation function ( $\phi$ ), which are normalized only by the respective values at  $\xi = 0^\circ$ . Received signal power is greatest for the ear ipsilateral to the speaker R when it is situated at  $90^\circ$  azimuth, and least for the contralateral ear. These acoustic measures can be compared with the corresponding, neurally generated ABR potentials (Fig. 4.6). Here the neural correlate of the relative power of the received signals at the left and right ears is the average of the peak amplitudes of waves IV and V (left and right), normalized to those at the frontal incidence. Although differences in units and scaling confound direct comparison between the results in Figs. 4.5 and 4.6, there are nevertheless qualitative similarities between these acoustic and physiological responses.



**Fig. 4.6** Averaged amplitudes of ABR waves IV and V as a function of horizontal angle of incidence. Data from four subjects. Symbol “I” is wave IV<sub>l</sub>; symbol “r” is wave IV<sub>r</sub>, and the symbol “V” indicates the averaged amplitudes of waves V<sub>l</sub> and V<sub>r</sub>. Amplitudes are normalized to their values at the frontal incidence

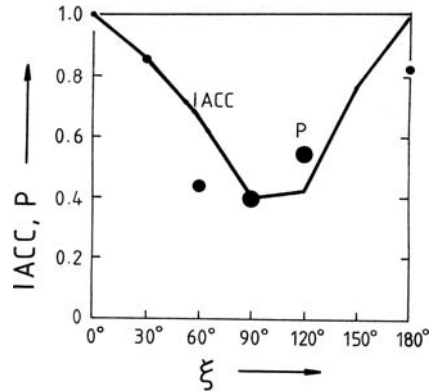


The relative behavior of wave IV (l) in Fig. 4.6 is similar to  $\Phi_{rr}(0)$  in Fig. 4.5, which was measured at the right-ear entrance r. Also, the relative behavior of wave IV<sub>r</sub> is similar to  $\Phi_{ll}(0)$  at the left-ear entrance l. In fact, amplitudes of wave IV (left and right) are proportional to  $\Phi_{rr}(0)$  and  $\Phi_{ll}(0)$ , respectively, due to the interchange of signal flow. The behavior of wave V is similar to that of the maximum value,  $|\Phi_{lr}(\tau)|_{\max}$ ,  $|\tau| < 1$  ms. Because correlations have the dimensions of the power of the sound signals, i.e., the square of ABR amplitude, the acoustic interaural correlation magnitude (IACC) defined by Equation (2.23) has a corresponding neural measure of binaural correlation  $P$ :

$$P = \frac{A_V^2}{[A_{IV,r}A_{IV,l}]} \quad (4.1)$$

where  $A_V$  is the amplitude of the wave V, which is a reflection of the “maximum” synchronized neural activity ( $\approx |\Phi_{lr}(\tau)|_{\max}$ ) in the inputs to the inferior colliculus (see Fig. 4.3). And,  $A_{IV,r}$  and  $A_{IV,l}$  are amplitudes of wave IV from the right and left, respectively. The results obtained by Equation (4.1) are plotted in Fig. 4.7. It is clear that the behavior of the IACC and  $P$  are in good agreement ( $r = 0.92$ ,  $p < 0.01$ ).

**Fig. 4.7** Values of the IACC measured due to Equation (2.23) together with (2.26) and values of  $P$  obtained by Equation (4.1). Data at  $\xi = 150^\circ$  were not available, except for one indicating a  $P$  value far over unity (unusual)



### 4.1.3 Remarks

Here we summarize the neural correlates of binaural hearing that can be observed in the relative amplitudes and latencies of waves in the left and right auditory brainstem responses (ABRs).

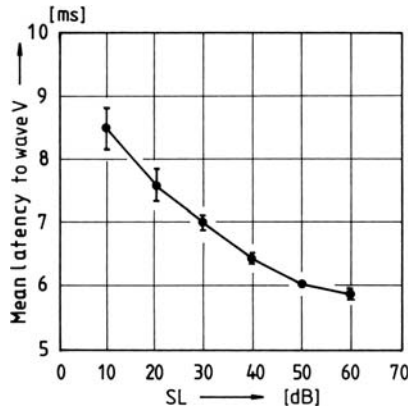
The amplitudes of the ABR clearly differ according to the horizontal angle of the incidence of sound relative to the listener (Fig. 4.2). In particular, the amplitudes of waves IV<sub>l</sub> and IV<sub>r</sub> appear to be nearly proportional to the sound pressures at the right and left ear entrances, respectively, when the amplitude is normalized to that in either the front ( $\xi = 0^\circ$ ) or back ( $\xi = 180^\circ$ ).

Due to the successive reversals of left and right amplitudes of corresponding ABR peaks discussed above in Section 4.1.1, we have inferred that a significant

part of the neural signal must cross the midline in at least four places. The first interchange of the neural signal appears to occur between right and left cochlear nuclei; the second at the level of the superior olivary complex and the third at the level of nuclei of the lateral lemniscus as shown in Fig. 4.3. Thompson and Thompson (1988), using neuroanatomical tract-tracing methods in guinea pigs, found four separate pathways connecting the two cochleae with themselves and each other via afferent and efferent pathways that pass through brainstem nuclei. This relates to the first interchange at the entrance of the cochlear nucleus, in the three interchanges in the auditory pathway.

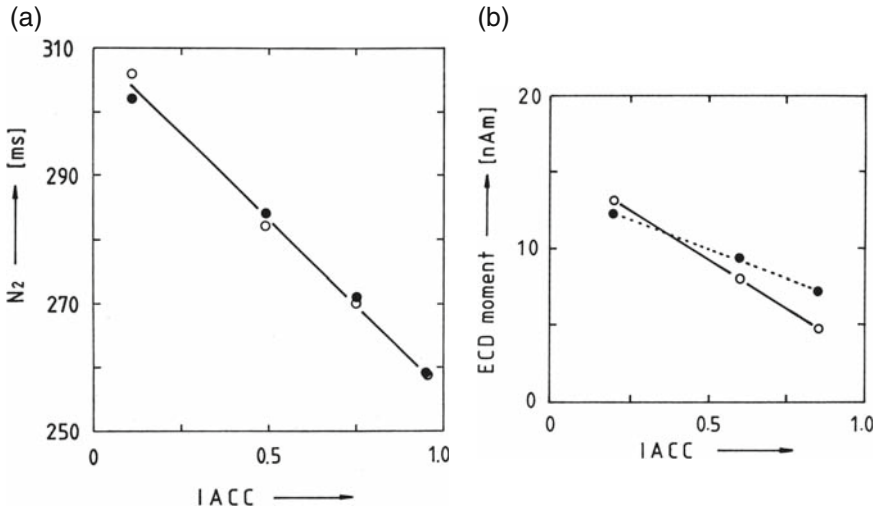
We found that the relative latencies of wave III in left and right ABRs correspond clearly to the interaural time difference. Moving up the pathway to stations above the neural binaural processors in the brainstem, the behavior of the magnitude of acoustic interaural crosscorrelation (IACC) is mirrored by “maximum” of neural activity for ABR wave V (associated with inputs to the inferior colliculus). The latency of wave V is around 8.5 ms latency relative to the time the sound signal is delivered to the loudspeakers, of which approximately 2 ms is the transit time of the sound from the loudspeakers ( $68 \pm 1$  cm away) to the ear (Fig. 4.4).

At the level of the inferior colliculus, on the other hand, Hecox and Galambos (1974) found that the latency of wave V decreases with an increasing sensation level, as shown in Fig. 4.8. This implies binaural summation for the sound energy or the sound pressure level (SPL), whose acoustic correlate can be seen in both  $\Phi_{11}(0)$  and  $\Phi_{11}(0)$  corresponding to Equation (2.24).



**Fig. 4.8** Latency of ABR wave V as a function of sensation level SL, expressed in dB re: sound detection threshold (Hecox and Galambos, 1974). This ABR latency may correspond to a binaural summation of sound energies from the left and right ears. (See also the longer and “stretched out”  $P_1$  and  $N_1$  single vertex response (SVR) latencies plotted as a function of SL in Fig. 4.14)

To summarize, the activity of the short-latency auditory brainstem responses (ABRs) that occur 0–10 ms after the sound signal has arrived at the eardrums likely reflects mechanisms in auditory binaural pathways that analyze interaural correlation patterns (e.g., the IACC).



**Fig. 4.9** (a) Relationship between the IACC and  $N_2$ -latency,  $p < 0.025$  (Ando et al., 1987a). ●,  $N_2$ -latency of SVR over the left hemisphere; ○,  $N_2$ -latency of SVR over the right hemisphere; (—), regression of these. (b) Relationship between the IACC and ECD movement,  $p < 0.01$  (Soeta, et al., 2004). ●, ECD movement over the left hemisphere; (---), regression; ○, ECD movement over the right hemisphere; (—), regression

## 4.2 Slow Vertex Responses (SVRs)

The SVR is a longer-latency (10–500 ms) averaged evoked auditory potential that reflects patterns of synchronized activity in populations of cortical neurons in the temporal lobe. We present the results of our measurements of SVRs of human listeners in the following sections. In these experiments we sought to observe the cortical neural response correlates both of acoustic parameters associated with different temporal and spatial percepts and with their preferred values. Correlates of temporal factors of the sound field reflect that reverberation patterns, such as the initial time delay gap between the direct sound and the first reflection ( $\Delta t_1$ ) and subsequent reverberation time ( $T_{\text{sub}}$ ), were found in SVRs associated with the left hemisphere. In contrast, typical spatial factors of the sound field, the IACC, and the binaural listening level (LL) were found in SVRs associated with the right hemisphere.

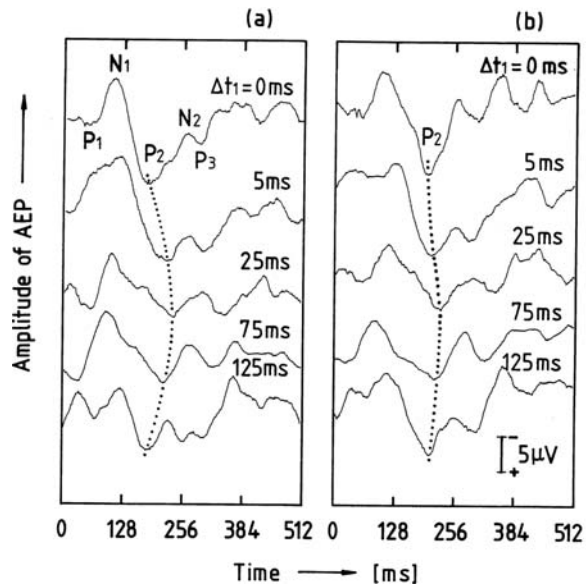
### 4.2.1 SVR Correlates of First Reflection Time $\Delta t_1$ Contrast

In experiments involving subjective preferences, we integrated the SVR for paired stimuli in a similar manner obtaining the scale value of subjective preference based on the paired-comparison method. First, subjective preferences were obtained by paired-comparison tests between different acoustic stimuli with different delay times of reflection. The preferred delay of single reflections observed in these tests

is typically in the 0–125 ms range. In order to compare slow vertex responses (SVR) with these subjective preferences, a reference stimulus was first presented, and then an adjustable test stimulus was presented. Many cortical neurons are more sensitive to changes in their inputs than those that stay the same. Typically, SVR magnitudes are greatest when stimulus patterns change abruptly, i.e., there is a pronounced contrast between paired stimuli in which spatial and/or temporal factors change substantially. The method of paired stimuli is therefore the most effective procedure because of this relativity of brain response.

Electrical responses were obtained from the left and right temporal regions (scalp locations T3 and T4, Fig. 4.24) according to the standard international 10–20 system for electrode placement (Jasper, 1958). Reference electrodes were located on the right and left earlobes and were connected together. Each subject had been asked to abstain from both smoking and drinking alcohol for 12 hours prior to the experiment. Pairs of stimuli were presented alternately 50 times through two loudspeakers which were located directly in front of the subject ( $68 \pm 1$  cm away). This arrangement ensured that the magnitude of interaural crosscorrelation (IACC) could be kept at a constant value near unity for all stimuli. The stimulus-triggered SVR waveforms for each trial were averaged together, as is done with other auditory-evoked potentials (AEPs).

Examples of such averaged SVRs are shown in Fig. 4.10. The stacked plots show averaged SVR waveforms as a function of the delay time of a single sound reflection, i.e., a sound delayed and added to itself. The source signal was a 0.9 s fragment of a continuous speech, the Japanese word, “ZOKI-BAYASHI,” which means grove or thicket. The reference “direct” stimulus sound field was the source signal without



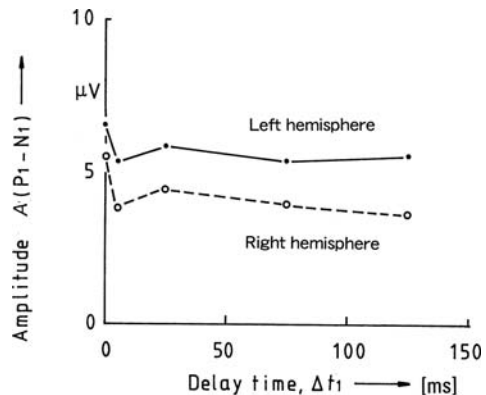
**Fig. 4.10** Example of the averaged SVR recorded for a single subject. *Dotted lines* are the loci of P<sub>2</sub>-latency for the delay time of the reflection (Ando et al., 1987b). The upward direction indicates negativity. (a) Left hemisphere. (b) Right hemisphere

any added reflections, while the second “reflection” stimulus was the source signal delayed and added to itself. The amplitude of the reflection was the same as that of the direct sound  $A_0=A_1=1$ , with delays ranging from 0 to 125 ms. Total sound pressure levels were kept constant, precisely between the direct and reflected sound stimuli by measuring the ACF,  $\Phi_p(0)$  (Section 2.1.2). From this figure, we find the longest neuronal response latencies (for the  $P_2$  wave this is the rightmost extent of the dotted line) for the stimulus that has the most preferred reflection time delay ( $\Delta t_1 = 25$  ms). This indicates the most comfortable condition for listening to continuous speech and for allowing the mind to relax. The delay time of 25 ms corresponds to the effective duration  $\tau_e$  of the continuous speech signal, as was defined in Section 2.1.2 and illustrated in Fig. 2.2 (Ando et al., 1987).

#### 4.2.2 Hemispheric Lateralization Related to Spatial Aspects of Sound

Because our sound environment always displays both temporal and spatial factors, and hemispheric dominance has been hypothesized for many different aspects of auditory function, we conducted a series of experiments to investigate possible hemispheric differences in neural response.

We observed hemispheric differences in both amplitudes and latencies of evoked electrical (SVR, EEG) and magnetic (MEG) neural responses. The amplitude of the first wave in the SVR as a function of the delay time is plotted in Fig. 4.11. This value of  $A(P_1-N_1)$  is the difference in amplitude between the first positive-going peak ( $P_1$ ) and the first negative-going trough ( $N_1$ ). Values for relative height  $A(P_1-N_1)$  from eight normal subjects who were presented the 0.9 s “ZOKI-BAYASHI” speech fragment were averaged together. The solid line indicates the initial SVR amplitude obtained from electrodes situated over the left hemisphere and the dashed line from those over the right hemisphere. The amplitude from the left is clearly



**Fig. 4.11** Averaged amplitudes  $A(P_1 - N_1)$  of the test sound field over the left and right hemispheres in change of the delay time of the reflection,  $\Delta t_1$  (eight subjects). (—): Left hemisphere. (---): Right hemisphere

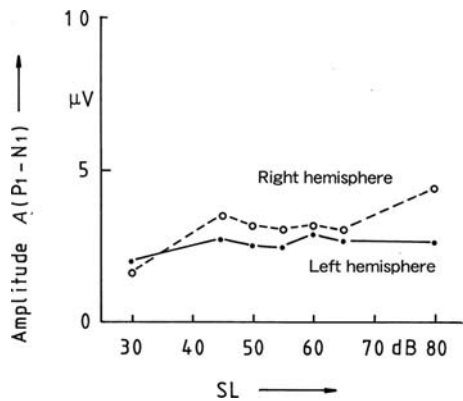
greater than that from the right ( $p < 0.01$ ), possibly implying a left hemisphere dominance or specialization in the processing of changes in the first reflection time for speech (Table 4.2) (Ando, 1992).

**Table 4.2** Amplitude differences of SVR,  $A(P_1 - N_1)$  over the left and right cerebral hemispheres. See also Tables 4.5 and 5.1

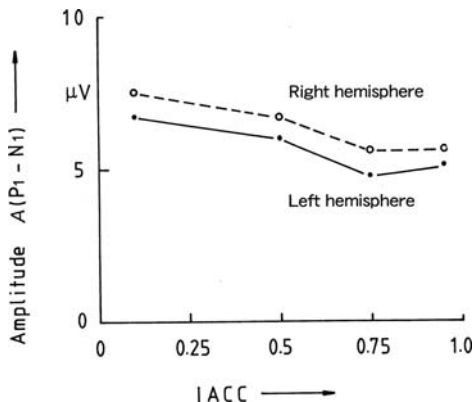
Source signal	Parameter varied	$A(P_1 - N_1)$	Significance level ( $p$ )
Speech (0.9 s)	SL	R > L	<0.01
Speech (0.9 s)	$\Delta t_1$	L > R	<0.01
Speech vowel /a/	IACC	R > L	<0.05
1/3-octave band noise	IACC	R > L	<0.05

Analogous hemispheric differences in the initial SVR amplitude were observed as a function of sensation level (SL) and the magnitude of interaural crosscorrelation (IACC). For all levels above 30 dB SL, the initial SVR amplitude from the right hemisphere was greater than that of the left,  $p < 0.01$  (Fig. 4.12) (Nagamatsu et al., 1989). We used 1/3-octave band noise with a center frequency of 500 Hz to examine the response correlates of sound direction. For all IACC values from 0.1 to 1.0 in the paired stimuli, the amplitude from the right hemisphere was always greater than that of the left,  $p < 0.01$  (Fig. 4.13) (Ando et al., 1987). To recapitulate, the sensation level response exhibits a right hemisphere dominance (Fig. 4.12). Also, there is a linear relationship between the IACC and the  $N_2$  latency observed in the slow vertex response (SVR) over both cerebral hemispheres, as shown in Fig. 4.9a (see also Fig. 4.14, right).

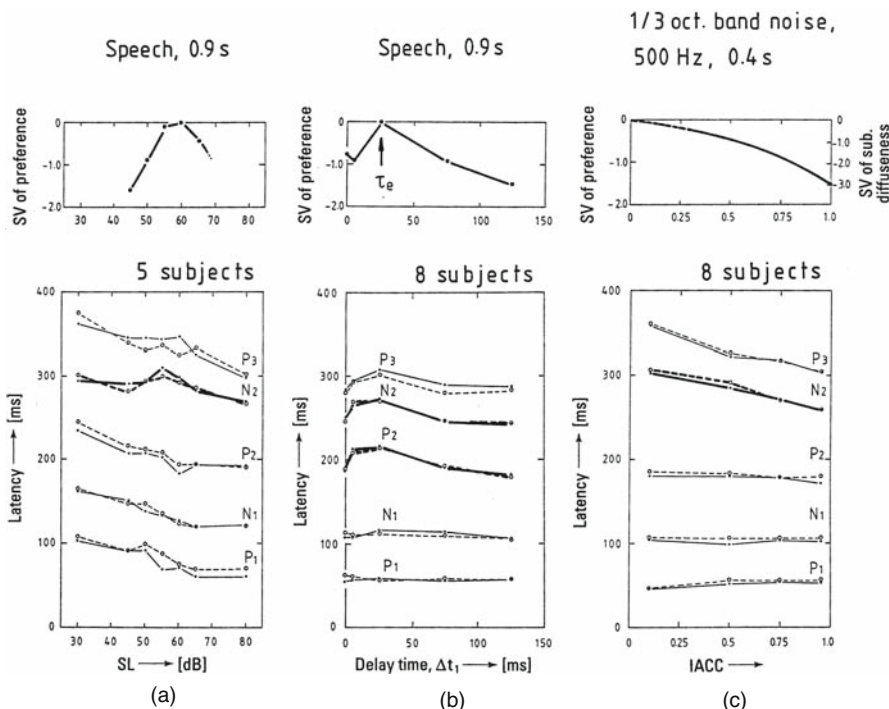
**Fig. 4.12** Averaged amplitudes  $A(P_1 - N_1)$  of the test sound field over the left and right hemispheres, in change of the sensation level (SL) (five subjects). (—): Left hemisphere. (---): Right hemisphere



By assembling the data (Table 4.2), one can see that hemispheric dominance of the SVR responses can change as a function of perceptually relevant acoustic parameters. The right hemisphere was more highly activated when spatial parameters of sounds, such as interaural magnitude (IACC), were varied in paired comparisons. The right hemisphere was also dominant when sensation level (SL) was varied, or even when a continuous speech signal /a/ was presented. On the other hand, the left



**Fig. 4.13** Averaged amplitudes  $A(P_1 - N_1)$  of the test sound field over the left and right hemispheres in change of the IACC (eight subjects). (—): Left hemisphere. (---): Right hemisphere



**Fig. 4.14** Relationships between averaged latencies of SVR and subjective preference for three factors of the sound field. (—): Left hemisphere. (---): Right hemisphere. (a) As a function of the SL. (b) As a function of the delay time of reflection,  $\Delta t_1$ . (c) As a function of the IACC

hemisphere was dominant when a temporal factor related to the sound field, the first reflection time  $\Delta t_1$ , was varied.

From classic studies, the left hemisphere appears to be more highly involved with processing speech and temporal sequences, while the right is more concerned with nonverbal identification and spatial processing functions (Kimura, 1973; Sperry, 1974). In light of our results, some aspects of hemispheric dominance may depend on context and contrast – what is changed in a pair of stimuli. The first sound stimulus sets up a neural processing context, such that the neuronal response to the second sound changes in a manner that can depend more on their contrast vis-à-vis temporal or spatial factors than on the absolute values of those factors.

### 4.2.3 Response Latency Correlates of Subjective Preference

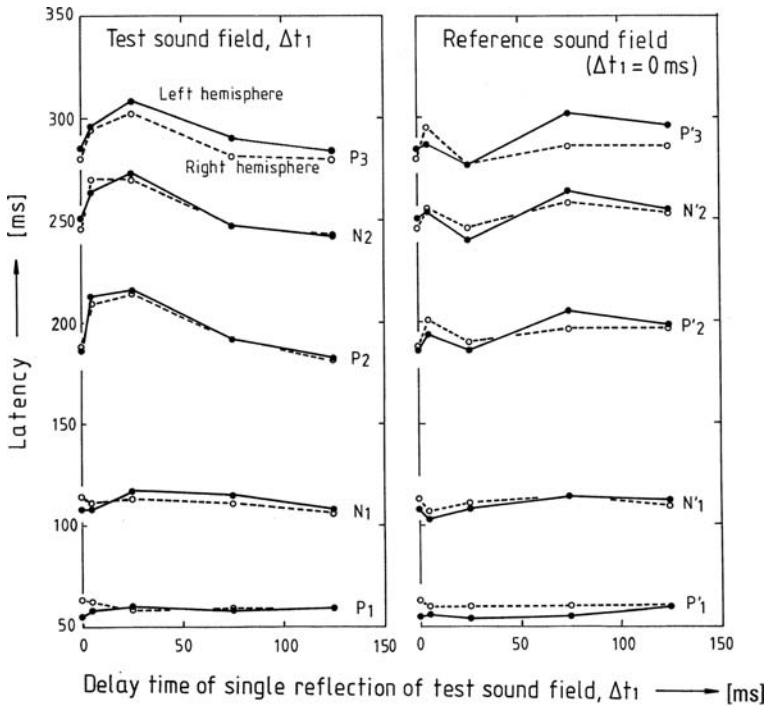
We found neural response correlates of subjective preference in the latencies of SVR peaks. The top plots of Fig. 4.14 summarize the relationship between subjective preference scale values and three acoustic parameters (sensation level SL, first reflection time  $\Delta t_1$ , and interaural correlation magnitude IACC). Applying the paired method of stimuli, both SVRs and subjective preferences for sounds fields were investigated as functions of these parameters. The upper panels show the scale preference values for the three parameters (the maxima are optimal, preferred values). The source signal was either the 0.9 s Japanese speech segment or 1/3-octave band noise. The lower panels plot the latencies of successive response components ( $P_1, N_1, P_2, N_2, P_3$ ). As shown in the left and center columns in the figure, the neural response observables related to subjective preferences for change in SL and  $\Delta t_1$  appeared typically in an  $N_2$  latency of 250–300 ms (note the corresponding peaks in the upper and lower plots).

Further details of the latencies for both the test sound field and the reference sound field, when  $\Delta t_1$  was changed, are shown in Fig. 4.15. The parallel latencies at  $P_2, N_2$ , and  $P_3$  were clearly observed as functions of the delay time  $\Delta t_1$ . However, latencies for the reference sound field ( $\Delta t_1 = 0$ ) in the paired stimuli were found to be relatively shorter, while the latencies for the test sound field with  $\Delta t_1 = 25$  ms, the most preferred delay, became longest. This pattern may indicate a kind of relative response of the brain, underestimating the reference sound field when the test sound field in the pair is the most preferred condition.

These relative long-latency responses are always observed in the subjectively preferred range of each factor. Thus, the difference of  $N_2$  latencies over both hemispheres in response to a pair of sound fields contains almost the same information obtained from paired-comparison tests for the subjective preference. In general, subjective preferences may reflect adaptive predispositions that aid in survival (“the direction of maintaining life”), such that it is perhaps not surprising that these primitive organismic responses should be present in observable neuronal responses.

The right column of Fig. 4.14 shows the effects of varying the IACC using 1/3-octave band noise (500 Hz) (Ando et al., 1987). In the upper plot, the scale preference value of subjective diffuseness is indicated as a function of IACC. Subjective





**Fig. 4.15** Averaged latencies for both the test sound field and the reference sound field for paired stimuli as a function of the delay time of the reflection,  $\Delta t_1$ . (—): Left hemisphere. (---): Right hemisphere. Maximum latencies of P<sub>2</sub>, N<sub>2</sub> and P<sub>3</sub> are found at  $\Delta t_1 = 25$  ms for the test sound field, whereas relatively short latencies of P'<sub>2</sub>, N'<sub>2</sub> and P'<sub>3</sub> are observed for the reference sound field. This is a typical brain activity showing “relativity”

preferences for subjective diffuseness behave in a similar manner for speech and music signals (Section 3.2). As with the other two factors, the neural correlate related to preference for subjective diffuseness first appears in the N<sub>2</sub> latency at 260–310 ms (the first IACC–latency curve that is correlated with the scale value for IACC and subjective diffuseness in the upper plot).

We observed a tendency for response latencies to increase when IACC decreases in both evoked electrical (SVRs, Figs. 4.9 and 4.14c) and magnetic (MEG) responses. This response pattern in Fig. 14c was observed in eight subjects (except for the left hemisphere of one subject). As already indicated in Fig. 4.9a, the relationship between the IACC and the N<sub>2</sub> latency was found to be linear and the correlation coefficient between them was 0.99 ( $p < 0.01$ ). In our MEG experiments, the activity of N1m represented by the equivalent current dipole (ECD) significantly increased with decreasing IACC, as shown in Fig. 4.9b (Soeta et al., 2004). Additionally, we found magnetic evoked correlates of another factor related to the IACC, the interaural delay  $\tau_{IACC}$ . This factor reflects the laterality (direction) of the sound image in the horizontal plane and a significant ECD correlate of this factor

was observed in the right hemisphere (Soeta and Nakagawa, 2006). Analogous to our results for evoked electrical potentials, for spatial sensations the magnetic ECD dynamic range in the right hemisphere was much larger than in the left.

Going back to Fig. 4.14, let us look at the behavior of early latencies of  $P_1$  and  $N_1$ . These were almost constant when the first reflection time and the IACC were changed. However, information related to SL or loudness is typically seen the  $N_1$  latency. This tendency agrees well with the results of Botte et al. (1975).

To summarize the neural response latency data from 40 to 170 ms latency, hemispheric dominance was found for SVR amplitude components, a difference that may reflect different functional specializations. Early latency differences corresponding to sensation level SL were found in the range of 120–170 ms. Finally, we found that the  $N_2$ -latency components in the delay range between 200 and 310 ms correspond well with subjective preferences for sensation level (SL), time delay of first reflection ( $\Delta t_1$ ), and indirectly, interaural correlation magnitude (IACC). Because the longest latency was always observed for the most preferred condition, one might speculate that the brain is most relaxed and less aroused at the preferred condition and that the slightly lower degree of neuronal excitability causes the longer observed response latencies to occur. For awake subjects, alpha rhythms observed in electroencephalography (EEG) and magnetoencephalography (MEG) usually have the longest periods of any waves that are present.

### 4.3 Electroencephalographic (EEG) Correlates of Subjective Preference

Thus far, we have discussed the neural correlates of changes in sensation level SL, first reflection time  $\Delta t_1$  and interaural correlation magnitude IACC that can be seen in averaged long-latency (0–500 ms) auditory-evoked potentials (SVR) in response to short signals less than a second long. However, for a wide range of reverberation times ( $T_{\text{sub}}$ ), no useful neural correlates of either their associated perceptual qualities or preferences could be obtained from examining slow vertex responses (SVRs). For these longer time windows one must look at ongoing neural responses to continuous sounds rather than brief, evoked auditory potentials. We therefore embarked on a series of experiments to find some distinctive feature in EEG signals that follows changes in the  $T_{\text{sub}}$  over long signal durations. In analyzing the data from these experiments we sought to find EEG correlates of subjective preferences in the duration that alpha rhythms persist, i.e., in the effective duration of alpha rhythms in the EEG signal.

#### 4.3.1 EEG Correlates of First Reflection Time $\Delta t_1$ Changes

In order to introduce the methods that we developed for analyzing EEG responses to sounds of extended duration, we shall first consider as preliminary study of the delay time of a single reflection to reconfirm the SVR results that were discussed in Section 4.2.3.

In this experiment, music motif B (Arnold's Sinfonietta, Opus 48, a 5 s segment of the third movement) was selected as the sound source (Table 2.1, Burd, 1969). The delay time of the first, single reflection  $\Delta t_1$  was alternatively adjusted to 35 ms (a preferred condition) and 245 ms (a condition of echo disturbance). The EEG from electrode positions T3 and T4 (Fig. 4.24) was recorded for about 140 s in response to ten delay-change pairs, and experiments were repeated over a total of 3 days. Eleven 22- to 26-year-old subjects participated in the experiment. Each subject was asked to close his eyes when listening to the music during the recording of the EEG. Two loudspeakers were arranged in front of the subject. Thus, the IACC was kept at a constant value near unity. The sound pressure level was fixed at 70 dBA peak, in which the amplitude of the single reflection was the same as that of the direct sound,  $A_0=A_1=1$ . The leading edge of each sound signal was recorded as a time stamp so that the recorded EEG responses could be precisely synchronized to and correlated with the presented sound. EEG signals amplified, passed through a filter with a 5–40 Hz bandwidth and slope of 140 dB/octave, and then digitally sampled at 100 Hz or more.

In order to find brain activity patterns corresponding to subjective preference, we analyzed the effective durations  $\tau_e$  of the autocorrelation functions (ACF) of EEG and MEG signals in the  $\alpha$ -wave range (8–13 Hz). The recorded signals were first passed through a 8–13 Hz digital band-pass filter, and effective durations (Fig. 2.2) were then computed from the ACFs of the filtered,  $\alpha$ -band signals. This gives a measure of how long, on the average, a temporally coherent alpha rhythm persists in a given part of the brain.

What is the length of the autocorrelation analysis window that is needed to reliably detect perceptually relevant changes in EEG signals? First, on the consideration that the subjective preference judgment needs at least 2 s to develop a psychological present, the running integration interval ( $2T$ ) was examined for periods between 1.0 and 4.0 s. A satisfactory duration  $2T$  in the ACF analysis was found only from the left hemisphere for 2–3 s, but not from the right (Ando and Chen, 1996).

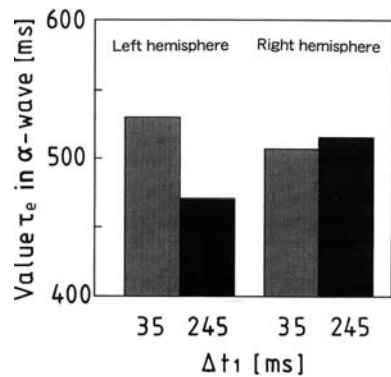
Table 4.3 indicates the results of an analysis of variance (ANOVA) for values of the effective duration  $\tau_e$  of EEG signals in the  $\alpha$ -band obtained using analysis windows of  $2T = 2.5$  s. LR in the table is analysis of differences in hemispheric response (electrode location T3 vs. T4). Though individual subject differences were

**Table 4.3** Results of the analysis of variance for the value  $\tau_e$  of EEG alpha wave with changes in  $\Delta t_1$

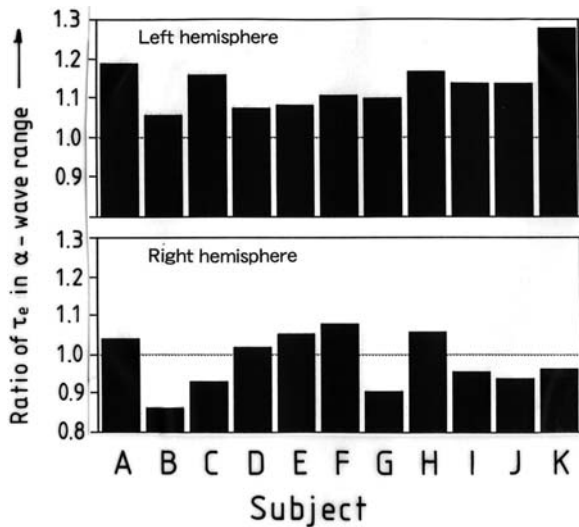
Factor	F	Significance level (p)
Subject	93.1	<0.01
Hemisphere, LR	1.0	
Delay time, $\Delta t_1$	5.8	<0.05
Subject and LR	8.9	<0.01
Subject and $\Delta t_1$	0.4	
LR and $\Delta t_1$	9.6	<0.01
Subject, LR, and $\Delta t_1$	0.4	

significant ( $p < 0.01$ ), the factor of first reflection time  $\Delta t_1$  (LR:  $p < 0.025$ ) by itself is also significant. However, it is significant for an interference effect between factors  $\Delta t_1$  and LR ( $p < 0.01$ ). Therefore, in order to analyze the data in more detail for each category,  $\Delta t_1$  and LR, we show the averaged value of  $\tau_e$  for the  $\alpha$ -band with 11 subjects in Fig. 4.16. A clear tendency is apparent. Effective durations  $\tau_e$  at  $\Delta t_1 = 35$  ms are significantly longer than those at  $\Delta t_1 = 245$  ms ( $p < 0.01$ ) only in the left hemisphere, not in the right. Ratios of  $\tau_e$  values in the  $\alpha$ -range at  $\Delta t_1 = 35$  and 245 ms, for each subject, are shown in Fig. 4.17. Remarkably, all individual data indicate that the ratios in the left hemisphere at the preferred condition of 35 ms are much longer than those in the right hemisphere. A long effective duration  $\tau_e$  of  $\alpha$ -rhythms may relate to the long  $N_2$  latency of SVR in the preferred condition that was discussed in Section 4.2.3.

**Fig. 4.16** Averaged effective durations  $\tau_e$  of EEG alpha rhythms in response to a change of first reflection time  $\Delta t_1$  to either 35 ms (the preferred time) or 245 ms (a condition of echo disturbance). Data from 11 subjects recorded from scalp electrodes on either left (“left hemisphere”) or right side (“right hemisphere”)



**Fig. 4.17** Ratios of effective durations  $\tau_e$  of EEG alpha rhythms in response to a change of first reflection time  $\Delta t_1$  for individual subjects A–K. Change in  $\Delta t_1$  was made either to the preferred value of 35 ms or to the unpreferred, echo disturbance value of 245 ms. Ratios are between the effective durations of the preferred condition divided by the durations of the unpreferred condition, i.e.,  $(\tau_e \text{ value at } 35 \text{ ms})/(\tau_e \text{ value at } 245 \text{ ms})$



Thus, the results reconfirm that, when the  $\Delta t_1$  is changed, the left hemisphere is highly activated (Fig. 4.11), and the effective durations  $\tau_e$  of  $\alpha$ -rhythms in this hemisphere correspond well with subjective preference. The  $\alpha$ -rhythm has the longest period in the EEG in the awake state and may be related to and associated with feelings of “pleasantness” and “comfort,” a condition that is universally preferred.

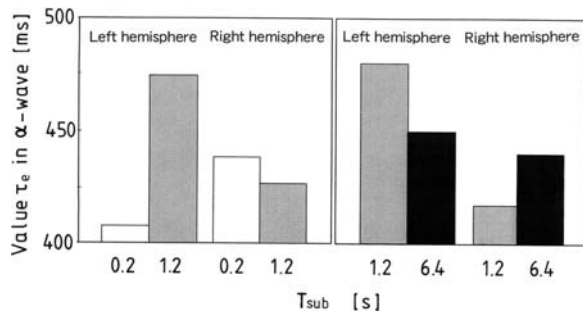
### 4.3.2 EEG Correlates of Reverberation Time $T_{sub}$ Changes

Now, let us examine how effective durations  $\tau_e$  of  $\alpha$ -rhythms change in response to reverberation times ( $T_{sub}$ ) and their subjective preference values.

Ten student subjects participated in the experiment (Chen and Ando, 1996). The sound source used was music motif B, Arnold’s lively, rollicking *allegro con brio* movement from his Sinfonetta No. 1 (Table 2.1). Ten 25- to 33-year-old subjects participated in the experiment. EEG signals from left and right hemispheres were recorded. Values of  $\tau_e$  of EEG signals in the  $\alpha$ -band were also analyzed using  $2T = 2.5$  s moving windows.

First consider the averaged values of the effective duration  $\tau_e$  of  $\alpha$ -activity, shown in Fig. 4.18. For this music, the preferred reverberation time  $T_{sub}$  is 1.2 s. Clearly, for the left hemisphere the values of  $\tau_e$  are much longer at the preferred condition  $T_{sub} = 1.2$  s than those at either of the non-preferred conditions  $T_{sub} = 0.2$  and 6.4 s. Thus, alpha rhythms persist longer in the left hemisphere for preferred reverberation time values. However, the contrary is true for the right hemisphere, where the effective durations of the non-preferred conditions  $T_{sub} = 0.2$  and 6.4 s are longer than those for the preferred conditions.

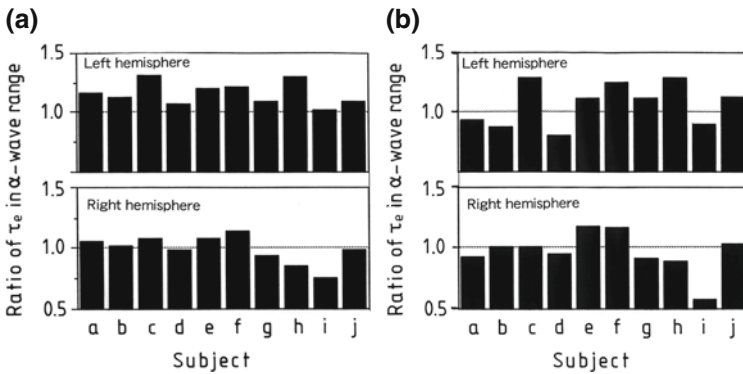
**Fig. 4.18** Averaged effective durations  $\tau_e$  of EEG alpha rhythms in response to a change of later reverberation time  $T_{sub}$  for two pairs of reverberation conditions (0.2 and 1.2 s) and (1.2 and 6.4 s). The preferred condition for late reverberation time is 1.2 s. Data from 10 subjects



The results of analysis of the variance are indicated in Table 4.4. Although there are large differences between individual subjects, a significant difference is achieved for  $T_{sub}$  in the pair of 0.2 and 1.2 s ( $p < 0.05$ ), and interference effects are observed for factors subject and LR ( $p < 0.01$ ), and LR and  $T_{sub}$  ( $p < 0.01$ ). No such significant differences are achieved for the pair at 1.2 and 6.4 s, but there are interference effects between subject and LR and subject and  $T_{sub}$ . Thus, in order to discuss the matter in more detail, the ratio of values of  $\tau_e$  for the  $\alpha$ -band is shown in Fig. 4.19 for each

**Table 4.4** Results of the ANOVA for the value  $\tau_e$  of EEG alpha wave with changes in  $T_{sub}$

Factor	Significance level (p)		Significance level (p)	
	F		F	
	(Pair of 0.2 s and 1.2 s)		(Pair of 1.2 s and 6.4 s)	
Subject	40.9	<0.01	40.2	<0.01
LR	2.1		2.0	
$T_{sub}$	6.2	<0.025	0.02	
Subject and LR	2.8	<0.01	2.0	<0.05
Subject and $T_{sub}$	1.2		2.7	<0.01
LR and $T_{sub}$	14.0	<0.01	0.2	
Subject, LR, and $T_{sub}$	1.3		0.7	

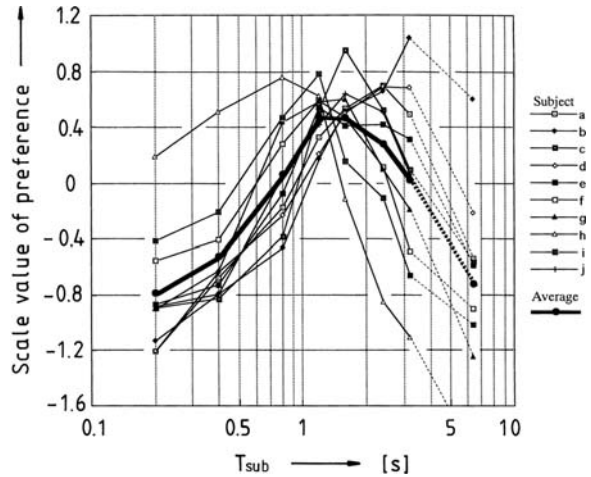


**Fig. 4.19** Ratios of effective durations  $\tau_e$  of EEG alpha rhythms in response to changes of later reverberation time  $T_{sub}$  for individual subjects a–j. The preferred value for reverberation time  $T_{sub}$  is 1.2 s. (a)  $(\tau_e \text{ value at } 1.2 \text{ s})/(\tau_e \text{ value at } 0.2 \text{ s})$ . (b)  $(\tau_e \text{ value at } 1.2 \text{ s})/(\tau_e \text{ value at } 6.4 \text{ s})$

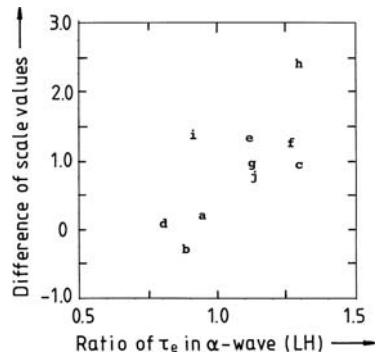
subject. All of the individual data indicate the ratios in the left hemisphere are much longer than those in the right hemisphere at  $T_{sub} = 1.2 \text{ s}$  in reference to  $T_{sub} = 0.2 \text{ s}$  (Fig. 4.19a). However, this is not the case for  $T_{sub} = 1.2 \text{ s}$  relative to  $T_{sub} = 6.4 \text{ s}$ , indicating large individual differences (Fig. 4.19b).

In fact, these individual results correspond well to the scale values of individual subjective preference. Figure 4.20 shows the scale values of preference as a function of  $T_{sub}$  for each subject. The most preferred values for  $T_{sub}$ , which were different for each subject, average at about 1.2 s. The ratio of values for  $\tau_e$  of the  $\alpha$ -wave at 1.2 and 6.4s is well correlated to the difference of the scale values of subjective preference for each individual, also reflecting large individual differences, as shown in Fig. 4.21 ( $r = 0.70, p < 0.01$ ).

**Fig. 4.20** Scale values of subjective preferences for later reverberation time  $T_{sub}$  obtained by paired comparisons for subjects a–j. Scale values at 6.4 are extrapolations using the formula of 3/2 power of  $\log_{10}(T_{sub})$ , according to Equations (3.7) and (3.10)



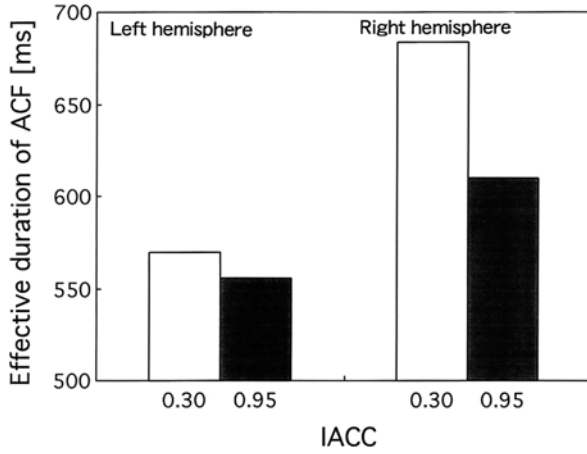
**Fig. 4.21** Relationship between difference of scale values for reverberation time preference [SV (1.2 s)–SV (6.4 s)] and the corresponding ratios of the effective durations of EEG alpha rhythms associated with the left hemisphere in response to a change in reverberation time  $T_{sub}$  for each of 10 subjects, a–j ( $r = 0.70, p < 0.01$ )



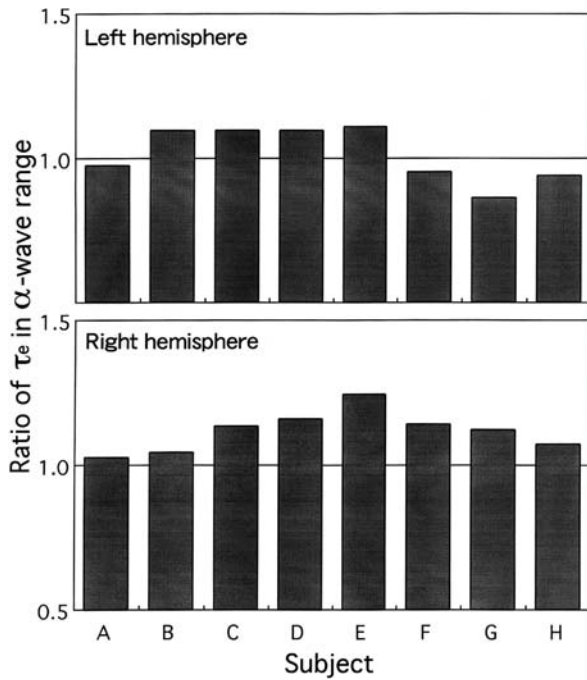
### 4.3.3 EEG Correlates of Interaural Correlation Magnitude (IACC) Changes

We also investigated EEG responses to changes in interaural correlation magnitude IACC. This acoustic parameter of the interaural correlation function (IACF) is related to the perception of the subjective diffuseness and apparent source width (ASW) of sound sources, which in turn are related to the degree to which one feels enveloped by a sound. The preferred condition is a minimal value of IACC, as close to zero as possible, that corresponds to maximal diffuseness and envelopment.

Eight student subjects participated in this paired-comparison experiment (Sato et al., 2003) and music motif B was again used as the stimulus. For interaural correlation magnitudes (IACCs), the effective duration  $\tau_e$  of  $\alpha$ -band activity found to be substantially longer in the preferred condition (IACC = 0.30) in the right hemisphere (Fig. 4.22). A significant difference is achieved in the right hemisphere for the pair of sound fields IACC = 0.95 and 0.30 ( $p < 0.01$ ). In seven of eight subjects, the ratios of effective durations  $\tau_e$  for  $\alpha$ -band responses to IACC change,



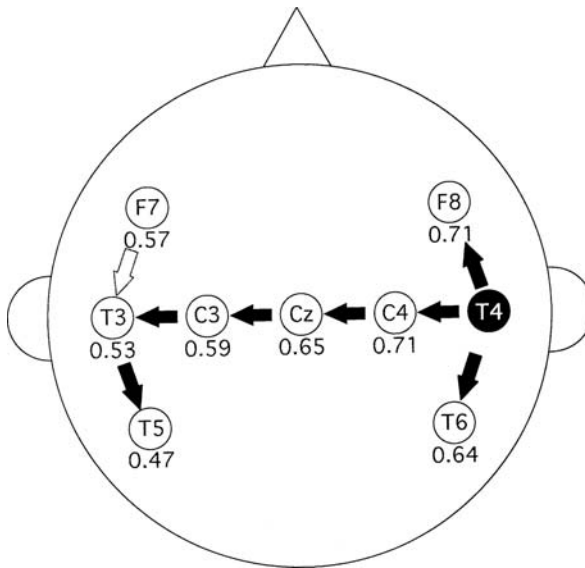
**Fig. 4.22** Averaged effective durations  $\tau_e$  of EEG alpha rhythms in response to a change of interaural correlation magnitude IACC for the pair of IACC = 0.30 and 0.95. Here the smaller value of IACC = 0.30 is preferred



**Fig. 4.23** Ratios of effective durations  $\tau_e$  of EEG alpha rhythms in response to changes of IACC in signals from the left hemisphere (T3) and the right hemisphere (T4) for eight subjects A–H. The ratio displayed is  $(\tau_e \text{ value at IACC} = 0.30)/(\tau_e \text{ value at IACC} = 0.95)$



$[\tau_e(\text{IACC} = 0.3)/\tau_e(\text{IACC} = 0.95)]$ , in the right hemisphere were greater than those in the left hemisphere (Fig. 4.23). Thus, as far as the IACC is concerned, the more preferred condition with a smaller IACC is related to longer  $\alpha$ -rhythm effective durations in the right hemisphere. A wave of  $\alpha$ -rhythm activity in the right hemisphere ( $T_4$ ) at  $\text{IACC} = 0.3$  later propagates toward the left hemisphere ( $T_3$ ) as shown in Fig. 4.24. One can examine its generation and movement by carrying out a cross-correlation analysis of the EEG signals at different electrodes. Using appropriate filtering one isolates the  $\alpha$ -band components of the signals at the different electrodes, computes crosscorrelation functions (CCFs) between signals from pairs of electrodes, and finds the time lag associated with maximum CCF values,  $|\phi(t)|_{\max}$ . The maximum amplitudes of the CCFs provide indices of the extent of coherent  $\alpha$ -rhythm activity in a given region. For the IACC, these maximum values are higher in right hemisphere than in the left hemisphere.



**Fig. 4.24** Propagation of the alpha wave flow from the right hemisphere to the left in response to a change of interaural correlation magnitude IACC. Real numbers reflect the median values of alpha rhythm correlation magnitudes between alpha band EEG signals from the indicated electrodes and electrode T4. The correlation magnitudes here are the maximum absolute values observed in the crosscorrelation functions computed from the two EEG signals

Table 4.5 summarizes hemispheric dominance results obtained by analysis of the effective durations  $\tau_e$  of  $\alpha$ -rhythms, with respect to changes in first reflection time  $\Delta t_1$ , longer reverberations  $T_{\text{sub}}$ , and the interaural correlation IACC. This finding suggests that the effective duration  $\tau_e$  in the  $\alpha$ -band is an objective index for designing excellent human acoustic environments (see also Section 4.4.1).

**Table 4.5** Hemisphere dominance determined by analyses for the value  $\tau_e$  of EEG alpha wave with changes in  $\Delta t_1$ ,  $T_{\text{sub}}$ , and the IACC. See also Tables 4.2 and 5.1

Source signal	Parameter varied	Ratio of values of $\tau_e$ for $\alpha$ waves	Significance level (p)
Music motif B	$\Delta t_1$	L > R	<0.01
Music motif B	$T_{\text{sub}}$	L > R	<0.01
Music motif B	IACC	R > L	<0.01

## 4.4 Magnetoencephalographic (MEG) Correlates of Preference and Annoyance

In magnetoencephalographic (MEG) studies, the weak magnetic fields produced by electric currents flowing in neurons are measured with multichannel SQUID (superconducting quantum interference device) gradiometers. These devices enable the safe and non-invasive study of many interesting properties of the working human brain.

Although both MEG and EEG have excellent temporal resolution, the two techniques measure somewhat different aspects of neuronal activity. MEG accurately detects superficial tangential currents, whereas EEG is sensitive to both radial and tangential current sources and reflects activity in the deepest parts of the brain. Only currents that have a component tangential to the surface of a spherically symmetric conductor produce a sufficiently strong magnetic field outside of the brain; radial sources are thus externally silent. Therefore, MEG mainly measures neuronal activity from the fissures of the cortex, which often simplifies interpretation of the data. Fortunately, all the primary sensory areas of the brain (auditory, somatosensory, and visual) are located within fissures. The advantages of MEG over EEG result mainly from the fact that the skull and other extracerebral tissues are practically transparent to magnetic fields, but do substantially alter electrical current flows. Thus, magnetic patterns outside the head are less distorted than electrical potentials on the scalp. Further, magnetic recording is reference free, whereas electric brain maps depend on the location of the reference electrode.

### 4.4.1 Preferences and the Persistence of Alpha Rhythms

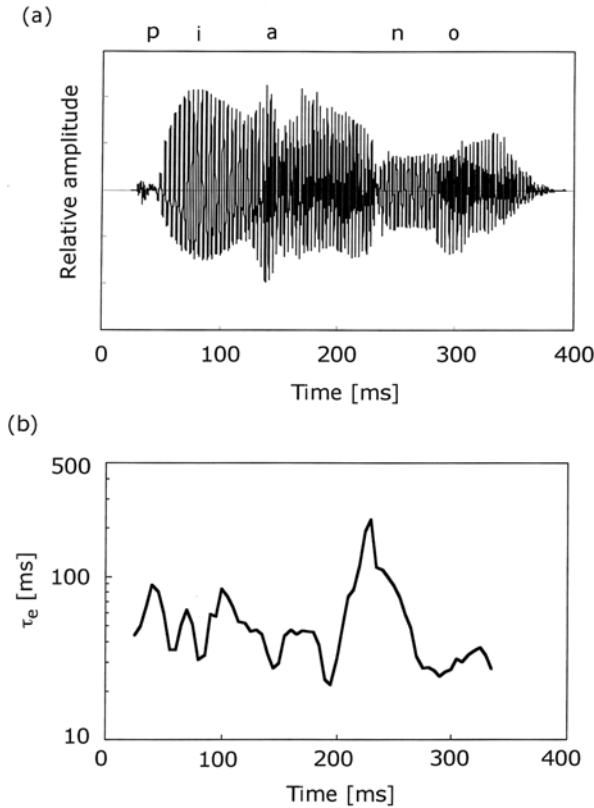
The general design of MEG experiments for finding the neural correlates of subjective preferences was similar to that used in EEG experiments. Measurements of MEG responses were performed in a magnetically shielded room using a 122-channel whole-head neuromagnetometer (Neuromag-122<sup>TM</sup>, Neuromag Ltd., Finland, see Photo 4.1) (Soeta et al., 2002).

The source signal was the word “piano,” which had a 0.35 s duration [Fig. 4.25(a)]. The minimum value of the moving effective duration  $\tau_e$ , i.e.,  $(\tau_e)_{\text{min}}$ , was about 20 ms (Fig. 4.25b). It is worth noting that this value is close to the most



**Photo 4.1** Magnetometer used in recording the MEG

preferred delay time of the first reflection of sound fields with continuous speech [Ando and Kageyama, 1977; Fig. 3.1(b)]. In the present experiment, the delay time of the single reflection ( $\Delta t_1$ ) 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.

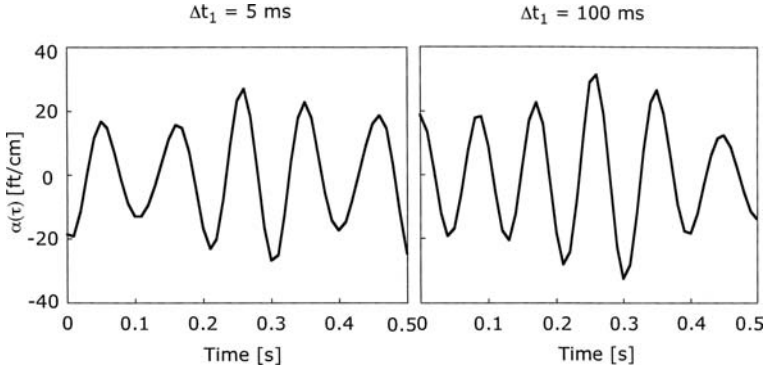


**Fig. 4.25** (a) Waveform of sound signal “piano.” (b) Effective duration of the running ACF,  $\tau_e$  ( $2T = 30$  ms at 5-ms intervals) of the signal,  $(\tau_e)_{\min} \approx 20$  ms

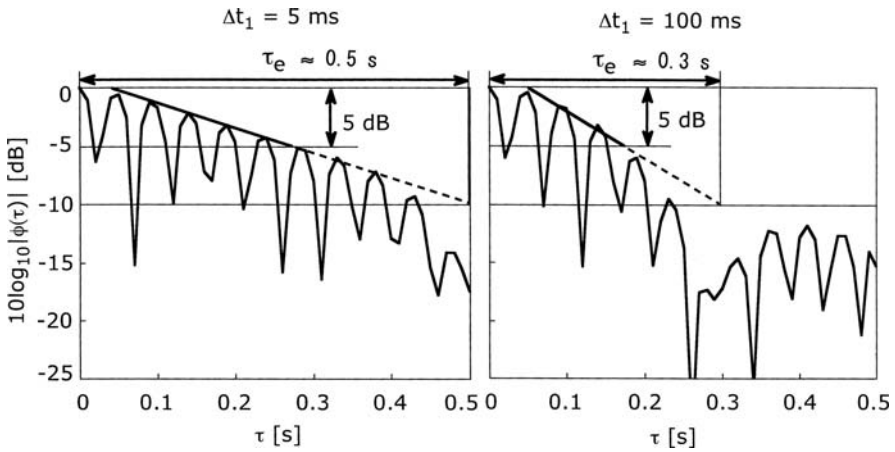
The sound–pressure level, which was measured at the end of the tubes, was fixed at 70 dBA.

Seven 23- to 25-year-old subjects, all with normal hearing, participated in the experiment. Using a paired-comparison test methodology, each subject compared ten possible stimulus pairs per session, and a total of ten sessions were conducted for each subject. Similar to the EEG measurements obtained in Section 4.3; 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 signals were analyzed. The magnetic data were recorded continuously with a filter of 0.1–30.0 Hz and digitized with a sampling rate of 100 Hz. Figure 4.26 shows an example of recorded MEG alpha waves.

The computation of the effective duration of MEG  $\alpha$ -band signals is shown in Fig. 4.27. Eight channels that had larger amplitudes of N1m response in each hemi-



**Fig. 4.26** Examples of recorded alpha band MEG signals (Soeta et al., 2002)

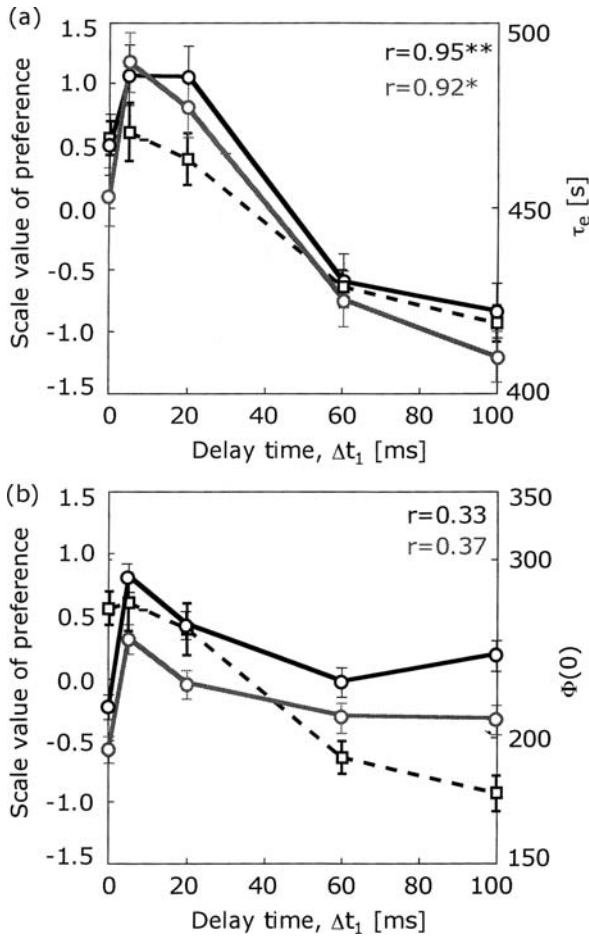


**Fig. 4.27** Examples of the normalized autocorrelation function (NACF) of alpha band (8–13 Hz) filtered MEG signals. The linear fit to the ACF envelope in the logarithmic scale and its extrapolation to 10% of its maximal value (–10 dB) illustrates the definition of effective duration  $\tau_e$ . At the preferred first reflection time ( $\Delta t_1 = 5$  ms), the effective duration  $\tau_e$  of the MEG alpha rhythm (about 0.5 s) is much longer than that observed (0.3 s) for the less preferred first reflection time ( $\Delta t_1 = 100$  ms) in which there is echo disturbance. All other things being equal, as a rule, alpha rhythms in both EEG and MEG signals persist over longer durations under preferred listening conditions

sphere were selected for an autocorrelation (ACF) analysis. We analyzed the MEG  $\alpha$ -band signal for each of the paired stimuli for each subject, computed their normalized autocorrelation functions (ACFs), and plotted them using logarithmic units of magnitude. The envelope of the ACF is fitted with a straight line and its intercepts with the –5 and –10 dB magnitude values are determined. The –10 dB (10% of maximum value) intercept is defined as the effective duration. From the figure, for the preferred condition where the first reflection time is short ( $\Delta t_1 = 5$  ms, left plot) the computed effective duration value of the MEG alpha band response is long

( $\tau_e \approx 0.5$  s), compared to the effective duration of  $\tau_e \approx 0.3$  s that is seen for the unfavorable condition of echo disturbance ( $\Delta t_1 = 100$  ms, right plot).

Results from the eight subjects confirmed a linear relationship between the averaged effective duration  $\tau_e$  values in the MEG alpha band 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 (Fig. 4.28a). As in



**Fig. 4.28** Alpha rhythm persistence vs. alpha band power as a predictor of preference. Relationships between scale preference values for first reflection time  $\Delta t_1$ , effective durations of alpha rhythms  $\tau_e$ , and amplitude of MEG alpha band signals. Averaged values of (a) alpha rhythm effective duration  $\tau_e$  and (b)  $\Phi(0)$  of the MEG alpha-band signals from left and right hemispheres as a function of  $\Delta t_1$ . Scale values of first reflection time preferences (dashed lines), left hemisphere MEG response (dark solid lines), right hemisphere MEG response (gray solid lines). Error bars are the 95% confidence interval. Correlations between left and right hemisphere responses and scale preference values are shown in the upper right corners of the plots, with significant correlations asterisked

the SVR and EEG studies of first reflection time  $\Delta t_1$ , the left-hemisphere response is dominant. An almost direct relationship between individual scale values of subjective preference and the  $\tau_e$  values over the left hemisphere was found in each of the eight subjects. Results for each of the subjects are shown in Fig. 4.29. Remarkably, the correlation coefficient,  $r$ , was 0.94 over all subjects. However, as shown in Fig. 4.28b, there was only a weak correlation between the scale values of subjective preference and power in the  $\alpha$ -band,  $\Phi(0)$ , for either hemisphere ( $r < 0.37$ ).

The effective duration  $\tau_e$  in the alpha wave band reflects the persistence of alpha rhythms in time, so that under the preferred listening conditions, the brain repeats a similar rhythm of alpha activity for a longer period of time. This tendency for a longer effective duration  $\tau_e$  of the alpha rhythm in the preferred condition is much more significant than the aforementioned results in Section 4.3 that were obtained through similar analyses of EEG signals.

#### ***4.4.2 Preferences and the Spatial Extent of Alpha Rhythms***

Magnetic responses were also analyzed using crosscorrelation functions (CCF) between 36 reference channels and 35 test channels. In MEG measurements using the word, “piano” as the source signal, 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 50 times alternately at a constant 1-s interstimulus interval (Soeta et al., 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 paired comparison (PCT) also.

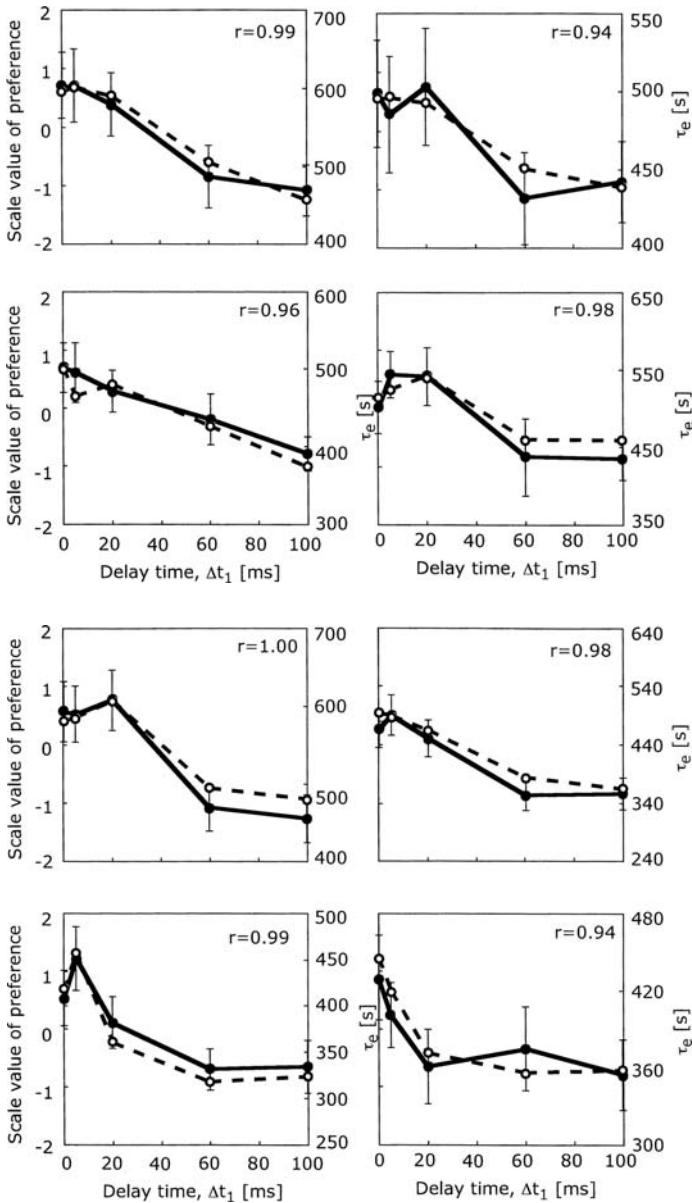
Results from this experiment showed that (1) the maximum amplitude of the crosscorrelation function CCF,  $|\phi(t)|_{\max}$ , between alpha band signals (8–13 Hz) recorded at two different channels increases with increasing subjective preference, and (2) the maximum amplitudes of channel crosscorrelations decrease with increasing channel distance. These imply that when listeners are stimulated using preferred sound fields, alpha rhythms persist over wider cortical territories, and that there is a higher degree of alpha rhythm coherence over these larger areas.

MEG experiments also reconfirmed, using the same speech signal with changing interaural correlation magnitudes (IACC = 0.27, 0.61, and 0.90), that effective durations  $\tau_e$  and maximum MEG channel crosscorrelation amplitudes increased when the IACC decreased (Soeta et al., 2005).

#### ***4.4.3 Alpha Rhythm Correlates of Annoyance***

In addition to the neural correlates of preferred listening conditions, one can also study distinctly non-preferred, annoying sounds and listening conditions. We undertook a series of MEG experiments similar to those described in the last section to measure neural responses to annoying stimuli: pure tones and band-pass noises with center frequencies of 1,000 Hz (Soeta et al., 2004).

In order to control the ACF of the source signal, the bandwidth of the noise, centered on 1,000 Hz, was varied at five levels (0, 40, 80, 160, and 320 Hz) using

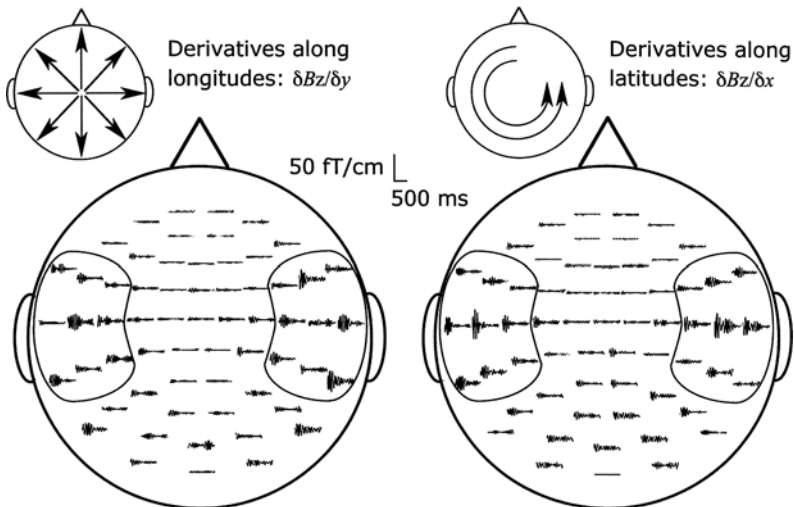


**Fig. 4.29** Correlations between scale values for first reflection time  $\Delta t_1$  preference and effective durations  $\tau_e$  of MEG alpha rhythms values for eight individual subjects. The individual preference curves (*dashed lines and open circles*) and the effective durations  $\tau_e$  of MEG alpha rhythms from the left hemisphere of the same individual subject (*solid lines, filled circles*) are superimposed. The averaged effective duration  $\tau_e$  value and the scale value was the highest correlation over the eight channels. Error bars show standard errors. Correlations between preferences and alpha rhythm effective durations are shown for each individual



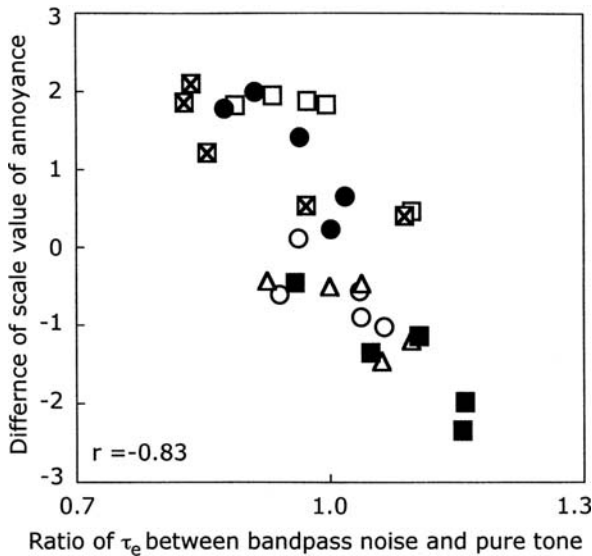
an extremely sharp filters with slopes of 2,000 dB/octave (Sato et al., 2002). We employed such filters in consideration of the sharpening effect previously observed at the level of the inferior colliculus for higher frequencies (Katsuki et al., 1958). It is worth noting that a filter with a slope of 60 dB/octave is too small to apply to any acoustic measurement. The “0 Hz” bandwidth signal was produced with an actual filter, for which the cutoff frequencies of a high-pass filter and a low-pass filter were both set at 1,000 Hz, so that only the slope component remains. The sound pressure level for all source signals was fixed at 74 dBA by measurement of the ACF,  $\Phi(0)$ . Source signals were characterized by the ACF temporal factors:  $\tau_1$ ,  $\phi_1$ , and  $\tau_e$ . The dependent factors,  $\phi_1$  and  $\tau_e$ , can be controlled by the bandwidth of the source signal.

Seven 22- to 28-year-old subjects participated in this experiment. Paired comparisons were performed for all combinations of 15 pairs in a single session, and subjects were run in a total of ten sessions. The signal duration was 2 s with rise/fall times of 10 ms. Subjects were asked to judge which of the two sound signals was more annoying, and thus the scale value of annoyance for each individual subject was obtained (Ando, 1998). The same subjects who participated in annoyance tests also participated in parallel MEG recording experiments in which the paired stimuli which were presented in exactly the same way. Combinations of the reference pure-tone stimulus and a test noise stimulus were presented alternatively 30 times and their MEG responses were recorded. Eighteen channels located around the temporal area in each hemisphere were selected for the auto- and crosscorrelation analysis of alpha wave activity in the 8–13 Hz range. Examples of recorded MEG signals are shown in Fig. 4.30.

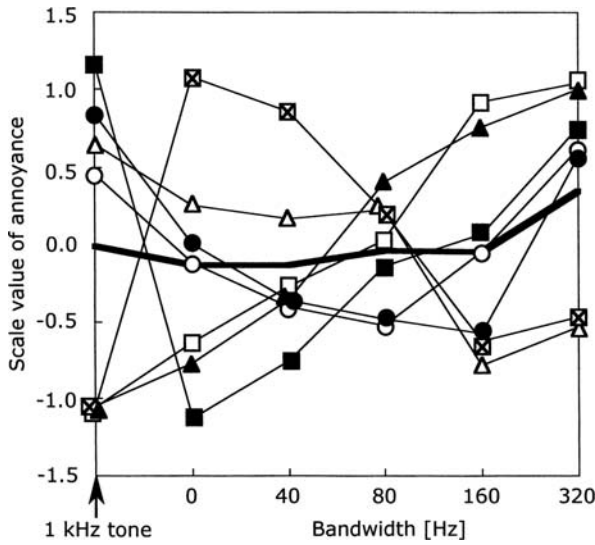


**Fig. 4.30** Examples of recorded MEG signals in response band-pass noise with center frequency of 1,000 Hz and an extremely narrow bandwidth (Soeta et al., 2004). Eighteen channels located around the left and right temporal regions (enclosed areas) were selected for alpha-band autocorrelation and crosscorrelation analysis

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, shown in Fig. 4.31, is that relative degree of annoyance (difference in scale value) was inversely correlated with the effective duration  $\tau_e$  of the band-pass noise relative to that of the tone (the ratio of these effective durations). The correlation coefficient of this relation was  $r = -0.83$ . Not unexpectedly, annoyance behaves in a manner opposite that of subjective preference, because effective duration  $\tau_e$  increases with increasing subjective preference. Thus, effective durations  $\tau_e$  were shorter for annoying stimuli. Also, it was found that the MEG channel crosscorrelation magnitudes  $|\phi(\tau)|_{\max}$  in the alpha band decreased with increasing annoyance. Thus, when an annoying stimulus is presented, the brain is not relaxed either temporally or spatially. Previous studies of EEG and MEG responses showed that effective duration  $\tau_e$  increased significantly with increasing preference. This signifies that the brain is repeating a similar (alpha) 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  $\tau_e$  were observed in the right hemisphere for all of subjects who participated. This implies a right hemispheric dominance for responses to noise and other non-verbal stimuli (Chon, 1970).



**Fig. 4.31** Relation of alpha rhythm persistence and annoyance. Differences in annoyance ratings of band-pass noise and pure tones are plotted as a function of the ratios of the effective durations of evoked MEG alpha rhythms for the two stimuli. Taking the difference of scale values of annoyance [SV(band-pass noise)–SV(pure tone)] and ratios of effective duration alpha rhythm durations normalizes these responses so that their relationship can be compared across subjects. Different symbols signify results of different subjects. A strong negative correlation ( $r = -0.83$ ) between the scale value of annoyance and the effective durations of MEG alpha rhythms was found. The shorter period of time the alpha rhythm persists, the greater the annoyance rating



**Fig. 4.32** The scale value of annoyance as a function of the bandwidth of band-pass noise of individual subjects. Different symbols signify results of different subjects

Although there is a deep relationship between annoyance and effective durations  $\tau_e$  of the MEG alpha band responses in each subject, annoyance itself differs between individuals. As shown in Fig. 4.32, large individual differences in the scale values of annoyance resulted as a function of bandwidth within the critical band. In this context of critical bands and loudness summation, it is also worth noting that evoked magnetic responses show N1m amplitudes that correspond well to subjective loudness judgments in the frequency range between 250 and 2,000 Hz (Soeta et al., 2006).

## Chapter 5

# Model of Temporal and Spatial Factors in the Central Auditory System

In this chapter, a workable neuropsychological signal processing model is proposed that links temporal and spatial acoustic factors with their corresponding perceptual attributes via observable response properties of the central auditory system. In the model, temporal factors that are observed to be associated with left cerebral hemisphere responses may be extracted from central autocorrelation processors. Similarly, spatial factors observed to be associated with the right hemisphere may arise from the action of central binaural crosscorrelators in the auditory pathway. Thus, the subjective attributes of sound fields can be described in terms of these temporal and the spatial factors and their corresponding specializations in the two cerebral hemispheres.

## 5.1 Signal Processing Model of the Human Auditory System

### 5.1.1 Summary of Neural Evidence

The central auditory signal processing model is based on several related sets of acoustical, mechanical, and neural evidence: the physical characteristics of the ear, auditory brainstem responses, slow vertex responses, EEG recordings, and MEG recordings.

#### 5.1.1.1 Physical Characteristics of the Ear

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 sound transmission path includes its propagation through external space as well as human head and pinna, the external canal, and the eardrum and the ossicular bone chain. The transfer function of this cascade system largely determines the sensitivity of the human ear. Because the A-weighting network is modeled to represent this sensitivity function, for the sake of practical convenience, it can be utilized in place of the transfer function of the physical system.

### 5.1.1.2 Left and Right Auditory Brainstem Responses (ABRs)

Characteristics of typical human auditory brainstem responses (ABR) imply:

1. Amplitudes of ABR waves  $I_{l,r}$  and  $III_{l,r}$  correspond roughly with the respective sound pressure levels at the two ears as a function of the horizontal angle of incidence to the listener,  $\xi$  as shown in Fig. 4.2a, c.
2. Amplitudes of waves  $II_{l,r}$  and  $IV_{l,r}$  correspond roughly to sound pressure levels as a function of the contralateral horizontal angle,  $\xi$  as shown in Fig. 4.2b and d. The left–right reversal of ABR wave amplitudes implies three major interchanges of neural signal flow between the left and right auditory pathways (Fig. 4.3).
3. Analysis of these ABR peaks suggest that they reflect neuronal responses at the level of the inferior colliculus that correspond well with interaural correlation magnitude values (IACCs, Figs. 4.5–4.7).

### 5.1.1.3 Left and Right Hemisphere Slow Vertex Responses (SVRs)

Recordings of left and right slow vertex responses (SVRs) have revealed the following:

4. The left and right peak-trough amplitudes of the early SVR,  $A(P_1-N_1)$  reflect left and right hemispheric dominance with respect to temporal and spatial factors, respectively. The temporal factor is the first reflection time-lag  $\Delta t_1$  (Fig. 4.11) and the spatial factors here are the sensation level SL (Fig. 4.12) and spatial compactness IACC (Fig. 4.13). At first, from a physical viewpoint, we considered classifying sensation level SL (or listening level LL) as a temporal-monaural factor. However, slow vertex responses indicated that SL is right hemisphere dominant. Classification of the LL as a spatial factor is natural because it is measured by the geometric average of sound energies arriving at the two ears (Equation 2.24).
5. Both left and right latencies of the  $N_2$  wave covary with interaural correlation magnitudes IACC (Fig. 4.9), and thus these are related to the listener preferences regarding sound diffuseness.

### 5.1.1.4 Left and Right Hemisphere EEG Responses

Analysis of EEG signals recorded from the left and right cerebral hemispheres reconfirmed that

6. Neuronal responses related to first reflection time  $\Delta t_1$  and later reverberations  $T_{\text{sub}}$  are relatively dominant in the left hemisphere (Figs. 4.16–4.19), while those related to the IACC are relatively dominant in the right hemisphere

(Figs. 4.22–4.24). This suggests to us that there is a high degree of independence between the left and right hemispheric factors.

7. When temporal and spatial factors of the sound field are varied, scale values of subjective preferences correspond well with the durations of alpha-coherence, i.e., the effective duration  $\tau_e$  extracted from the ACF of the  $\alpha$ -band EEG.

#### **5.1.1.5 Left and Right Hemisphere MEG Responses**

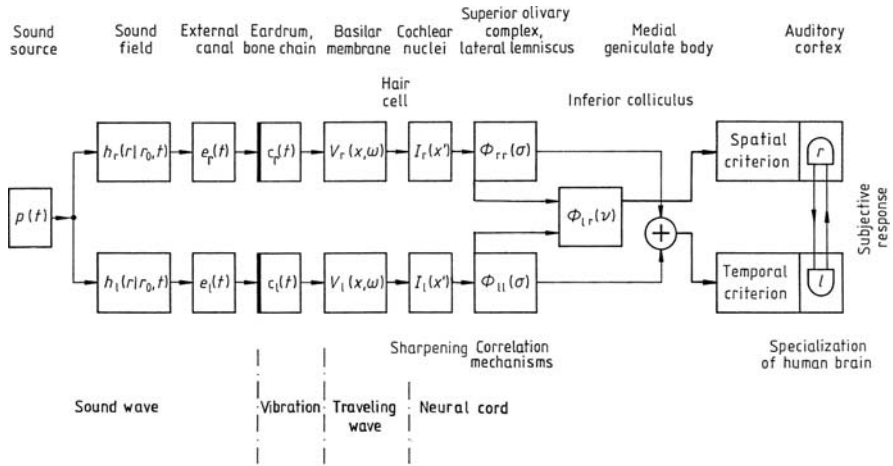
Analysis of MEG signals recorded from the left and right cerebral hemispheres reconfirmed that

8. Amplitudes of MEG signals recorded when first reflection time  $\Delta t_1$  is changed show left hemisphere specialization (Figs. 4.28 and 4.29).
9. As with EEG signals, the scale values of individual subjective preferences correspond well with durations of alpha-coherence in MEG signals from the left hemisphere (Fig. 4.29).

### ***5.1.2 Auditory Signal Processing Model***

Based on the above-mentioned physical system and physiological responses, a high level signal processing model of the central auditory system can be constructed to relate acoustics, observable aspects of neuronal response, temporal and spatial perceptual attributes, and subjective preferences (Ando, 1985). The model consists of monaural autocorrelation mechanisms, the interaural crosscorrelation mechanism between the two auditory pathways, and the specialization of human cerebral hemispheres for temporal and spatial factors of the sound field (Fig. 5.1).

The signal processing pathway in the figure roughly corresponds to the successive auditory stations and major levels of processing in the auditory pathway. Obviously the neuroanatomical, neurophysiological, and neurocomputational descriptions of each of these auditory stages are much more detailed than those found in this high level signal processing model. At every station there are many different cell types, neurotransmitters, projection patterns, neurophysiological response types, and information processing roles. For example, the “cochlear nucleus” in our high-level auditory signal processing model is, on closer investigation, a collection of three major nuclei, each of which has subdivisions, complex internal structure, distinctive response types, and separate functional roles. One can find more detailed, comprehensive accounts of the structure and function of the auditory system in a number of excellent monographs and edited books (Aitkin et al., 1984; Irvine, 1986; Yost and Gourevitch, 1987; Aitkin, 1990; Popper and Fay, 1992; Webster et al., 1992; Dallos et al., 1996; Hawkins, 1996; Ehret and Romand, 1997; Oertel et al., 2002). Many of these are volumes of the Springer Handbooks of Auditory Research series.



**Fig. 5.1** Central auditory signal processing model for auditory percepts and subjective preferences. The signal processing chain begins with a source sound signal in the time domain  $p(t)$  which is mechanically filtered successively by the outer ears, external canals, eardrums, middle ear bone chains, and basilar membranes, half-wave rectified and transduced into ionic currents by cochlear hair cells that are converted into trains of action potentials (“spikes”) in auditory nerve fibers that are time-locked to the filtered stimulus waveforms. Effectively, the auditory nerve (not shown in this schematic) is an array of roughly 3,000 frequency channels each of which conveys the fine temporal structure of the cochlea-filtered signals up to  $\sim 5$  kHz. The stimulus-driven temporal patterning of spikes supports autocorrelation-like representations based on population-wide interspike interval distributions and crosscorrelation-like representations based on the interactions of time-structured spike trains in brainstem binaural crosscorrelators. The signal processing model uses two central representations: a monaural autocorrelation function (ACF) and a binaural, interaural crosscorrelation function (IACF) that are very tentatively located in the upper brainstem and midbrain (nuclei of the lateral lemniscus, inferior colliculus). The two monaural autocorrelation representations may be combined  $\oplus$  in the inferior colliculus. From evidence for cortical lateralizations of responses related to temporal and spatial factors and therefore to these two representations, it is surmised that there are preferential signal paths that connect the monaural autocorrelation (ACF) representation with the left auditory cortex and the binaural interaural correlation (IACF) representation with the right auditory cortex (Ando, 1985). This signal processing flow leads to representations of ACF-related temporal factors in the left auditory cortex (Chapter 6) and representations of IACF-related spatial factors in the right auditory cortex (Chapter 7). The overall subjective responses, for example, subjective preference and annoyance, may be processed in both hemispheres in relation to their respective temporal and spatial factors (Ando, 2002)

The early and middle parts of the signal processing chain in Fig. 5.1 model the transformation from sound source to neural representations for monaural and binaural hearing at the level of the inferior colliculus. A sound source  $p(t)$  is located at position  $r_0$  in a three-dimensional space and a listener is sitting at position  $r$ , which is defined by the location of the center of the his or her head.  $h_{l,r}(r|r_0, t)$  are the impulse responses of the two sound paths through the room between the sound’s origin  $r_0$  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. Mechanical–electrical transduction of the traveling wave is carried out in the inner hair cells of the cochlea, which half-wave rectify the signals in the process. The axons of primary sensory neurons in the auditory system, called “auditory nerve fibers,” collectively make up the auditory nerve (also called the cochlear or acoustic nerve). Hair cell synapses on auditory nerve fibers produce postsynaptic currents that trigger action potentials (“spikes”) which are then conducted to secondary neurons in the cochlear nuclei. The subsequent ascending auditory pathway continues via several diverging and converging paths through the superior olivary complex (including the nuclei of the medial superior olive MSO, the lateral superior olive LSO, and the trapezoid body), the auditory midbrain (inferior colliculus), auditory thalamus (medial geniculate body), and auditory cortex. The main monaural ascending auditory neural signal pathway crosses the midline in the brainstem, between the cochlear nucleus and the inferior colliculus. The set of auditory pathways is bilateral at all levels, culminating in the auditory cortical regions of the two cerebral hemispheres.

The nature of the signals in the auditory system is a critical consideration. Neural representations of both the stimulus power spectrum and its autocorrelation function (ACF) exist in early auditory stations. The input power density spectrum of the cochlea  $I(x')$  can be roughly mapped at a certain cochlear place  $x'$  (Katsuki et al., 1958; Kiang et al., 1965). Neural frequency maps (“tonotopy” or “cochleotopy”) that can support coarse frequency distinctions (between octaves and large fractions of octaves) can be found at all levels of auditory pathway from brainstem to cortex. The neural representation of the stimulus autocorrelation function is less widely appreciated largely because it exists in the temporal firing patterns of auditory neurons (i.e., a temporal code) rather than in which neurons are most strongly excited (a rate-place code). The spike timing information that supports the temporal correlation representations and operations exists to its fullest frequency extent in the auditory nerve, brainstem, and the inputs to midbrain, but drastically declines in quality and frequency extent as one ascends the pathway to the auditory thalamus and cortex.

Temporal codes in the auditory nerve, brainstem, and midbrain support both temporal autocorrelation representations of monaural percepts (temporal sensations) and temporal crosscorrelation representations of spatial auditory percepts (spatial sensations). Patterns of vibrations mechanically filtered through the cochlea and transduced into electrical currents by cochlear hair cells impress their temporal fine structure on the spike trains of auditory nerve fibers. At moderate to high sound levels, the signal processing resembles coarse band-pass filtering followed by half-wave rectification. Jitter introduced by chemical synapses between inner hair cells and auditory nerve fibers degrades the fidelity of the stimulus-locked spike timing such that an effective direct temporal representation of stimulus periodicities appears to exist only up to roughly 4–5 kHz. Such temporal discharge patterns create spike trains that are correlated with stimulus waveforms. A striking visual example of these stimulus-driven temporal firing patterns can be found in the time domain analysis of firing rates of cat auditory nerve fibers in response to a multi-formant



synthetic vowel (Secker-Walker and Searle, 1990, see also Cariani, 1999). This correlation of neural temporal discharge patterns with incoming acoustic signals, albeit taking into account cochlear filtering and rectification, is the deep reason that the neuronal correlation-based representations and operations resemble their acoustic counterparts (ACF, IACF). The systematic similarity allows one to discuss the signal processing properties that these representations have in common.

Support for temporal autocorrelation representations can be found in interspike interval distributions in early auditory stations. Interspike intervals are the time intervals between spikes. First-order interspike intervals are time intervals between consecutive spikes, whereas all-order intervals are intervals between both consecutive and nonconsecutive spikes. As a consequence of phase-locking, the interspike intervals generated in the auditory nerve precisely and robustly reflect stimulus periodicities. Histograms of these “all-order interspike intervals” produced by a given auditory nerve fiber resemble the autocorrelation functions of acoustic signals in the form they are presented to that fiber, i.e., after the acoustic signals have been subjected to cochlear filtering, half-wave rectification, and the effects of synaptic jitter.

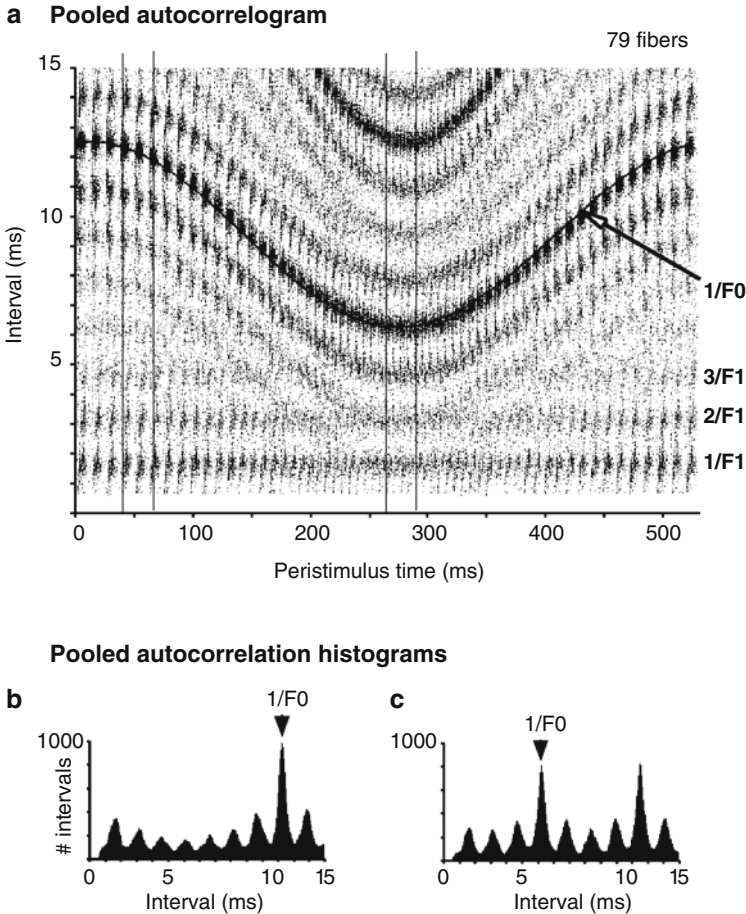
The interspike interval statistics for the entire auditory nerve constitute a neural population code for the stimulus autocorrelation function (ACF). When the interspike intervals of auditory nerve fibers associated with all cochlear places are combined, the result is called variously a “summary autocorrelation function” (Meddis and Hewitt, 1991, 1998), a “pooled all-order interspike interval distribution” (Cariani and Delgutte, 1996a, b), or simply a “population–interval distribution” (Cariani, 1999, 2001). For the same reasons that the power spectrum of a signal can be decomposed into the sum of the spectra of a set of overlapping band-pass filters, the autocorrelation of a signal can be decomposed into the sum of the autocorrelations of a similar set of filters (Cariani and Delgutte, 1996a; Cariani, 1999). Just as the neural spatial excitation pattern across the cochlea can serve as a representation of the power spectrum of the acoustic stimulus, albeit with distortions, population–interval distributions can serve as neural “autocorrelation-like” representations of the stimulus, i.e., a neural ACF. The main differences between these neural representations and their signal processing counterparts are due to floor and ceiling (saturation) nonlinearities of auditory nerve firing rates, effects of cochlear filtering on interactions between partials, weakening of phase-locking at higher frequencies and periodicities, and cochlear distortion products.

Thus an “autocorrelation-like” neural representation exists in the time domain, in the temporal discharge patterns in the auditory nerve. This temporal representation complements traditional frequency-domain “rate-place” representations for periodicities up to about 4–5 kHz. Because the temporal representation cannot account for the perception of pure tones above this limit, i.e.,  $f > 5$  kHz, several auditory theorists have postulated a dual or “duplex” auditory representation that includes both neural ACF and power spectrum representations. The original duplex model of J.C.R. Licklider (1951, 1959) was implemented several decades later in the running “correlogram” computer simulations of Slaney and Lyon (1993).

Some differences exist between the temporal and the spectral representations. Unlike the rate-place representation, the neural ACF is robust over the entire dynamic range of hearing. This 4–5 kHz limit of availability of temporal information corresponds well to the existence region of musical tonality, as evidenced by the frequency limits of pitches of complex tones, octave matching, and the recognition of musical intervals and melodies (Inoue et al., 2001; Cariani, 2001, 2002). In contrast with neural representations that are based on neural firing rates, which tend to change their form and degrade at higher sound levels, representations based on temporal spike information remain largely invariant in form over the whole dynamic range of hearing and thus mirror the behavior of auditory percepts. For low-frequency components that are resolved by cochlear filters, population–interval distributions resemble the autocorrelation of the stimulus itself. For higher-frequency components that are not resolved by cochlear filters, population–interval distributions resemble the autocorrelation of their waveform envelopes (Cariani and Delgutte, 1996a; Cariani, 1999). The behavior of population–interval distributions parallels perceptual differences that are observed between sets of low, resolved harmonics and high, unresolved ones.

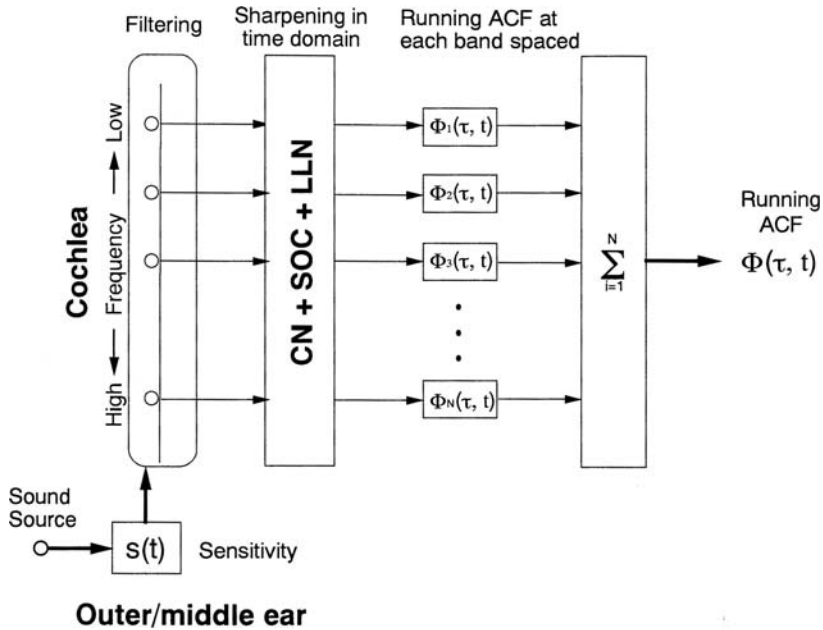
A case can be made that this neural ACF representation conveys most of the fine periodicity and spectral information that is used for music and speech perception. This includes the pitches of pure and complex tones (Cariani and Delgutte, 1996a, b), timbral qualities of sustained musical notes (Cariani, 2001) and voiced speech (Palmer, 1992; Hirahara et al., 1996), and tonal consonance (Tramo et al., 2001). If the neural ACF is extended in time to encompass longer time delays that are present in neural responses, then it can also handle rhythm, prosody, and reverberations. A running neural ACF representation could potentially also handle transient patterns of amplitude, frequency, and phase that distinguish the timbres of nonstationary sounds, such as characteristic onset dynamics of musical instruments and consonantal distinctions in speech. An example of using features of the running ACF to account for the perception of the timbres of distorted electric guitars is presented in Section 6.6.

The predominant interspike interval present in the population–interval representation reliably predicts both the pitch of low-frequency pure tones (< 4–5 kHz) and the pitch that is heard at the fundamental of harmonic, complex tones. The upper plot of Fig. 5.2 shows a running population–interval distribution of 79 cat auditory nerve fibers in response to single-formant vowel whose fundamental frequency oscillates between 80 and 160 Hz. The stimulus was presented 100 times at 60 dB SPL. Each dot in the plot represents 10 or more interspike intervals of a given duration (ordinate) whose last spike ended at a given time relative to the stimulus onset (abscissa). The dark interval bands track the fundamental period (12.5–6.25–12.5 ms) and its double-period (in the middle of the plot) as it oscillates over time. The lower plots are cross-sections of the upper plot that show the population–interval distribution of intervals ending in a short (20 ms) time slice. Here the major peaks are associated with the fundamental period, which is the period of the pitch that is heard. The minor peaks are associated with the dominant harmonics of the single vowel (e.g., those in vowel formant regions), which determine the vowel’s tonal quality/timbre/phonetic



**Fig. 5.2** Population-wide all-order interspike interval distributions in the auditory nerve in response to a single-formant vowel with an oscillating fundamental frequency ( $F_0=80\text{--}160\text{ Hz}$ ). **(a)** Autocorrelogram plot showing the global interspike interval distribution of 79 auditory nerve fibers as a function of time after the stimulus onset (peristimulus time). The stimulus was presented 100 times to each fiber at 60 dB SPL. The plot is based on data from roughly 100,000 spikes. Each dot represents 10 or more intervals of a given duration ( $y$ ) that ended at the given peristimulus time ( $x$ ). **(b)** and **(c)** Population-wide interspike interval distributions for the two 20 ms stimulus segments demarcated in the autocorrelogram when  $F_0=83$  and 160 Hz. The peaks closely follow the fundamental frequency  $F_0$  of the stimulus and effectively predict the low pitch that is heard (Cariani and Delgutte, 1996a). The form of the population-interval distributions resembles that of the stimulus autocorrelation function (ACF) and so can serve as a general purpose autocorrelation-like representation of the stimulus (Cariani, 1999, 2002)

identity. Thus, the autocorrelation-like population-interval distribution can effectively represent pitch and those aspects of timbre that are related to low-frequency spectra.



**Fig. 5.3** A tentative two-dimensional model of the ACF processor in the auditory pathways. Each single neuron can act as a filter with its characteristic frequency. It is worth noting that there are no scientific evidences for the fixed range of frequency band like the 1/3octave or 1/10octave filters, because cut-off frequencies of filters might be arbitrary. CN, Cochlear nuclei; SOC, Superior olivary complex; LLN, Lateral lemniscus

The signal processing model of the running autocorrelation processor is depicted in Fig. 5.3. The nature of the representation and processing of monaural temporal information in the auditory brainstem and above is still poorly understood (Cariani, 1999). Stimulus-related spike timing information becomes successively sparser and less in evidence as one ascends the auditory pathway above the midbrain level. Current evidence suggests that the most plausible location for a neural autocorrelation analysis in the left and right monaural pathways ( $\Phi_{ll}(\sigma)$  and  $\Phi_{rr}(\sigma)$  in Fig. 5.1) is at the level of the nuclei of the lateral lemniscus or the inferior colliculus.

Moving up the signal processing chain, subsequent neural populations at the cortical level produce responses that covary in various ways with perceptual preferences. In comparison to the right, the temporal regions of the left cerebral hemisphere generate relatively larger neuronal responses for stimulus parameters that are associated with temporal sensory attributes. They are therefore likely to be driven by the outputs of internal autocorrelation-like representations and processing mechanisms. This left hemisphere response dominance could be caused by preferential ascending connections from monaural autocorrelators lower in the pathway or by differential amplification of their outputs via recurrent cortico-thalamic or cortico-collicular circuits.

The early temporal code in the auditory system also supports binaural crosscorrelation mechanisms at the level of the superior olive and the inferior colliculus (Fig. 5.1). As was discussed in the preceding chapter, there exist correlates of the interaural correlation magnitude IACC in neural activity (Fig. 4.7) at the level of the brainstem and midbrain (wave V together with waves  $IV_L$  and  $IV_R$ ). A great deal of anatomical and neurophysiological evidence points to interaural crosscorrelation representations at the levels of the superior olive and the inferior colliculus. We surmise that the output signal of the interaural crosscorrelation mechanism that computes the IACC may be dominantly connected to the right hemisphere. Representation of sound pressure level may also be preferentially processed in the right hemisphere. Sound pressure level can be described in terms of a geometrical average of the ACFs for the two ears at the origin of time ( $\sigma = 0$ ), but like other aspects of spatial hearing, changes in sound pressure level produce changes in neuronal response latencies at the level of the inferior colliculus. Differences in neuronal responses across the two hemispheres could be due to specializations for fine analyses of interspike interval patterns vs. neuronal response latencies that minimize cross-talk between the two types of representation. The autocorrelation and crosscorrelation mechanisms may also preferentially access the two hemispheres for reasons of economy, in order to minimize neural processing effort and maximize efficiency.

**Table 5.1** Hemispheric specializations of temporal and spatial factors determined by analyses of the AEPs, EEG, and MEG

Factors changed	AEP (SVR) $A(P_1 - N_1)$	EEG, ratio of ACF $\tau_e$ values of alpha wave	AEP (MEG) N1m	MEG, ACF $\tau_e$ value of $\alpha$ -wave
<b>Temporal</b>				
$\Delta t_1$	L > R (speech) <sup>1</sup>	L > R (music)		L > R (speech)
$T_{sub}$	–	L > R (music)	–	
<b>Spatial</b>				
LL	R > L (speech)	–	–	
IACC	R > L (vowel /a/)	R > L (music) <sup>2</sup>	R > L (band noise) <sup>3</sup>	
	R > L (band noise)			
$\tau_{IACC}$			R > L (band noise) <sup>3</sup>	
Head-related transfer functions			R > L (vowels) <sup>4</sup>	

<sup>1</sup> Sound source used in experiment is indicated in parentheses.

<sup>2</sup> Flow of EEG alpha 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 alpha waves recorded at different electrodes.

<sup>3</sup> Soeta and Nakagawa (2006).

<sup>4</sup> Palomaki et al. (2002).

We conclude that the listening level (LL) and the IACC are associated with the right cerebral hemisphere, and the temporal factors,  $\Delta t_1$  and  $T_{\text{sub}}$ , and the sound field in a room are associated with the left (Table 5.1). Hemispheric specializations for temporal and spatial sensations may relate to the highly independent contribution their factors play in the perception of subjective attributes. For example, the “cocktail party effect,” the ability to “hear out” individual speakers amidst many background voices, may be explained at least partially on terms of such specializations, because speech is mainly processed in the left hemisphere, in a manner somewhat independent of the spatial information that is represented in the right hemisphere (Section 9.2).

Based on the model, we will describe the perceptual, subjective attributes of sounds and sound fields in terms of signal representations and processing operations, relating these where possible to neuronal signal processing in the ascending auditory pathways and auditory regions of the two cerebral hemispheres.

## 5.2 Temporal Factors Extracted from Autocorrelations of Sound Signals

The autocorrelation function ACF of the sound signals at the left and the right ear entrances, respectively, is defined by,

$$\begin{aligned}\Phi_{ll}(\tau) &= \frac{1}{2T} \int_{-T}^{+T} f_l'(t)f_l'(t + \tau)dt \\ \Phi_{rr}(\tau) &= \frac{1}{2T} \int_{-T}^{+T} f_r'(t)f_r'(t + \tau)dt\end{aligned}\tag{5.1}$$

where  $p'(t) = p(t)*s(t)$ ,  $s(t)$  is the ear sensitivity, which is essentially the transfer function of the physical system from the free field to the oval window of the cochlea. The auditory temporal integration window  $2T$  is described in the following section. Rigorously, although the sensitivities of individual listeners vary, for practical purposes, the impulse response of an A-weighted network is commonly used as a generic substitute.

The ACF and the power density spectrum mathematically contain the same information, and both kinds of representations are present at the level of the auditory nerve. For reasons discussed in the preceding sections, we believe that the fine spectral and temporal information needed for the analysis of music and speech is mainly conveyed through distributions of interspike intervals that constitute autocorrelation-like representations. This means that features in the neural ACF have extensive correspondences with those of the ACFs of acoustic sound sources, both in isolation and in filtered through sound fields. We believe that the correlation mechanisms in the neural system may work in a manner analogous to the features and operations on the ACF in the model.

As is mentioned in Section 2.1, five temporal factors are extracted from the ACF analysis, many of which are associated with corresponding auditory percepts:

1. ACF amplitude at zero lag (loudness). Energies at two ear entrances represented are given by  $\Phi_{ll}(0)$  and  $\Phi_{rr}(0)$ . The geometrical mean of the sound energies arriving at the two ears yields the binaural listening level (LL) that is given by

$$LL = 10\log[\Phi_{ll}(0)\Phi_{rr}(0)]^{1/2}/\Phi_{\text{Ref}}(0)$$

where  $\Phi_{\text{Ref}}(0)$  is the reference level corresponding to the pressure 20  $\mu\text{Pa}$  that is the reference value of the sound pressure level (SPL).

2. Width of the ACF peak about zero lag. 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  $W_{\phi(0)}$ , given that  $\phi(\tau)$  is an even function. In the fine structure of the autocorrelation function, which consists of peaks, troughs, and delays, for instance,  $\tau_1$  and  $\phi_1$ , respectively, are the delay time and the amplitude of the first ACF peak. After the first peak there are subsequent peaks and delays,  $\tau_n$  and  $\phi_n$  being the delay time and the amplitude of the  $n$ th local peak ( $n > 1$ ). Usually there are certain correlations between  $\tau_n$  and  $\tau_{n+1}$  and between  $\phi_n$  and  $\phi_{n+1}$ .

Two of the most significant factors that are related to the perceived timbre of signals lie at the first ACF maximum (aside from the central peak at zero-lag origin). These are:

3. Factor  $\tau_1$ , the delay of the first ACF peak, which reflects the frequency of the highest partial present in the signal, i.e., signal bandwidth.
4. Factor  $\phi_1$ , amplitude of the first ACF peak, which reflects the magnitude of this partial.
5. Effective duration. The effective duration of the envelope of the normalized ACF,  $\tau_e$ , is defined by the tenth-percentile delay and represents a repetitive feature containing the sound source itself (Ando, 1985, 1998). Intuitively, the effective duration is a measure of the duration of periodic structure in the signal.

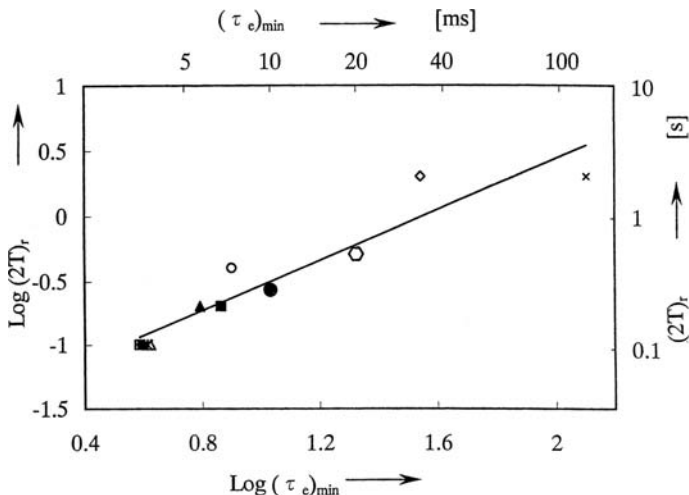
The normalized ACF is defined by

$$\phi_p(\tau) = \Phi_p(\tau)/\Phi_p(0) \quad (5.2)$$

As in Fig. 2.2, effective duration  $\tau_e$  is obtained by fitting a straight line to extrapolate the delay time when the ACF amplitude would be 10% of its zero-lag value (i.e., at -10 dB). The linear extrapolation in effect assumes that the initial envelope of the ACF decays exponentially. It is remarkable that these temporal factors or features of the ACF can effectively describe four temporal sensations, i.e., loudness, pitch, timbre, and duration. These are discussed further in Chapter 6 and in Ando (2002).

### 5.3 Auditory Temporal Window for Autocorrelation Processing

In analyzing the running autocorrelation function ACF, we need to know the temporal integration time of auditory signals, the so-called auditory-temporal window  $2T$  in Equation (5.1). The initial part of the ACF within the effective duration  $\tau_e$  of the



**Fig. 5.4** Relationship between the temporal integration window  $(2T)_r$  and the minimum value of the effective duration  $(\tau_e)_{\min}$  of the running ACF. Different symbols signify experimental results using different sound sources

ACF contains the information about the signal that is most relevant to its perception. In order to determine the auditory-temporal window, we can examine the time duration over which the perceptual effects of sounds are integrated. Successive loudness judgments in pursuit of the running listening level LL have been conducted (Mouri, Akiyama and Ando, 2001). Results are shown in Fig. 5.4, so that the recommended signal duration  $(2T)_r$  to be analyzed is approximately given by

$$(2T)_r \approx 30(\tau_e)_{\min} \tag{5.3}$$

where  $(\tau_e)_{\min}$  is the minimum value of  $\tau_e$  obtained by analyzing the running ACF. The auditory-temporal integration window should thus be about 30 times the minimum effective duration of the signal. This signifies an adaptive temporal window that depends on the temporal characteristics of the sound signal presented. Therefore, the range of the optimal temporal window duration differs according to whether the source material consists of musical pieces, with long effective durations, or continuous speech, with short ones. For music, the optimal auditory-temporal window range  $(2T)_r$  is 0.5–5 s, whereas for continuous speech, it is 50–100 ms for vowels and 5–10 ms for consonants. The auditory-temporal window roughly corresponds to “fast” or “slow” time constant settings in contemporary sound meters. Ideally, for a given type of sound source, such parameters would be replaced by an optimal temporal window computed from the effective durations of the signals themselves. The running time stepsize  $(R_s)$ , which signifies the degree of overlap of successive segments of the signal that are analyzed, is not so critical. The analysis stepsize  $K_2(2T)_r$  is typically a fraction  $K_2$  of the auditory-temporal window, with  $K_2$  being in the range of 1/4–1/2.



## 5.4 Spatial Factors and Interaural Crosscorrelation

In the auditory signal processing model, features in the interaural crosscorrelation function IACF underlie most perceptual attributes associated with spatial hearing, i.e., the spatial sensations. The interaural crosscorrelation function IACF for the sound pressures,  $p(t)_{l,r}$ , arriving at the two ear entrances is given by

$$\Phi_{lr}(\tau) = \frac{1}{2T} \int_{-T}^{+T} f'_l(t) f'_r(t + \tau) dt \quad (5.4)$$

where  $f'_{l,r}(t) = p(t)_{l,r} * s(t)$ ,  $s(t)$  being the ear sensitivity, which is essentially the transfer function of the outer and middle ears. The normalized IACF in the possible range of maximum interaural delay times is given by,

$$\Phi_{lr}(\tau) = \Phi_{lr}(\tau) / [\Phi_{ll}(0) \Phi_{rr}(0)]^{1/2}, \quad -1 \text{ ms} < \tau < +1 \text{ ms} \quad (5.5)$$

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

From the IACF analysis, four spatial factors are computed (Fig. 2.8):

1. The magnitude IACC of the interaural crosscorrelation function (IACF, denoted  $\phi_{lr}$ ) is defined by the maximum value of the crosscorrelation function

$$\text{IACC} = |\phi_{lr}(\tau)|_{\max} \quad (5.6)$$

for possible interaural time delays in the physiological range:  $-1 \text{ ms} < \tau < +1 \text{ ms}$ .

2. The interaural delay time at which the IACF has its maximal value, i.e., the delay at which IACC is defined is  $\tau_{\text{IACC}}$ .
3. The effective bandwidth of the IACC,  $W_{\text{IACC}}$  is the width of the IACF peak whose maximum defines the IACC. The bandwidth is measured by finding the interaural delay interval around the IACF peak for which the value is at least  $\delta$  below its maximal value, the IACC. This interaural correlation bandwidth  $W_{\text{IACC}}$  may correspond to the just-noticeable-differences (JND) of the interaural correlation magnitude IACC and its associated percepts: apparent source width, subjective diffuseness, and envelopment. This value mainly depends on the frequency composition of the source signal. One perceives sounds that are spatially well defined and localized in the horizontal plane at a lateral position associated with the interaural time delay  $\tau_{\text{IACC}}$  when there is a sharp peak in the IACF and a correspondingly small value of  $W_{\text{IACC}}$ . On the other hand, when listening to a sound field with a low value of IACC that is less than 0.15, then a subjectively diffuse sound is perceived (Damaske and Ando, 1972).
4. The denominator of Equation (5.5), defined by the binaural listening level LL, is the geometrical mean of the sound energies arriving at the two ears.

In Chapter 7, the spatial sensations of the sound field are described in terms of these factors extracted from the IACF. These spatial sensations become immediately apparent when we enter a sound field, because our binaural system evaluates the IACF within a relatively short temporal window, as discussed in the next section. This situation is quite different from the adaptive temporal window for the sound signals themselves, which varies according to the effective duration of the ACF of the sound-source signal.

## 5.5 Auditory Temporal Window for Binaural Processing

The interaural correlation function IACF contains the information needed for localizing a sound in the horizontal plane. When a sound signal is moving such that it changes direction in the horizontal plane, we must identify a suitable “temporal window”  $2T$  for analyzing the running IACF. This binaural temporal integration window corresponds to our ability to track the moving image of the localized sound. The range of  $\tau_{IACF}$  values extracted from the IACF determines the range of conditions under which such a moving image can be followed. The binaural temporal integration window ( $2T$ ) must be shorter than the period of movement, or the image will be smeared and no movement will be detected. It should also be longer than the interaural delays used for binaural localization because its value fluctuates greatly when  $2T$  is shorter than 1 ms, which is the largest value of  $\tau_{IACF}$  that is possible within the natural physiological range. For a sound source moving sinusoidally in the horizontal plane with a cycling rate of less than 0.2 Hz,  $2T$  may be selected in a range from 30 to 1,000 ms. When a sound source is moving with a cycling rate of 4.0 Hz or less, a binaural integration window of 30–100 ms is acceptable (Mouri, Fujii, Shimokura and Ando, unpublished). In order to obtain reliable results, the recommended temporal window for the IACF that covers a wide range of movement velocities in the horizontal localization is around 30 ms. For measurement of spatial factors at each audience seat location in a concert hall using a sound source that is fixed on the stage, the value of the binaural integration window may be selected to be longer than 1.0 s.

## 5.6 Hemispheric Specialization for Spatial Attributes of Sound Fields

Characterization of the independent influences of the aforementioned temporal and spatial factors on subjective preference judgments has been achieved (Ando, 1983, 1985, Chapter 4).

Recordings over the left and right hemispheres of SVR, EEG, and MEG have revealed the following evidence for functional specialization with respect to these perceptual attributes (Table 5.1).

1. The left- and right-relative peak amplitudes of the first major SVR waves,  $A(P_1 - N_1)$ , indicate lateralization of responses associated with the temporal factor ( $\Delta t_1$  in the left hemisphere) and spatial factors (LL and IACC in the right hemisphere). Correlates of loudness, the sensation level SL and the binaural listening level LL, were classified as temporal-monaural factors from a physical viewpoint. However, the neural correlates of these parameters were observed in slow vertex responses (SVRs) predominantly at electrode locations over the right hemisphere. Thus, in terms of central processing mechanisms, SL and LL should be reclassified as spatial factors.
2. Both the left and right latencies of  $N_2$  covary with the IACC.
3. Results of EEG measurements that showed hemispheric lateralization of the temporal factors, i.e.,  $\Delta t_1$  and  $T_{\text{sub}}$  reconfirmed left hemisphere dominance for these factors. Similarly, spatial factors associated with interaural magnitude IACC showed right hemispheric dominance. Thus, there appears to be a high degree of independence between the factors that predominant in the neural responses in the left and right hemispheres.
4. Scale values of subjective preferences are well predicted from the values of the effective durations  $\tau_e$  of EEG  $\alpha$ -band activity recorded over the left and right hemispheres. Neural correlates of preferences related to changing temporal factors dominate in left hemisphere EEG signals, while those associated with changing spatial factors dominate in the right hemisphere responses.
5. Recorded MEG amplitudes reconfirmed the left hemisphere specialization for first reflection time  $\Delta t_1$ .
6. Scale values of individual subjective preferences relate directly to the effective duration  $\tau_e$  of MEG  $\alpha$ -band activity. Note that it is the effective durations of the  $\alpha$ -rhythms in EEG and MEG recordings, and not the absolute *amplitudes* of these waves, that correspond to subjective preferences.
7. A right hemisphere specialization was reconfirmed using MEG recordings for the IACC and  $t_{\text{IACC}}$  (Soeta and Nakagawa, 2006).

In addition to the above-mentioned temporal response patterns, spatial patterns of cortical neural response were analyzed by examining crosscorrelations between  $\alpha$ -band activity at different scalp locations in the two hemispheres using EEG (see Fig. 4.24 and Section 15.1.2 in this volume; Sato et al., 2003; Okamoto et al., 2003) and MEG signals (Soeta et al., 2003). The results show that larger areas of the cerebral cortex show  $\alpha$ -rhythm activity when preferred sound fields are presented. These observations imply that the brain repeats a similar temporal rhythm in the  $\alpha$ -frequency band over greater numbers of neurons when preferences are better satisfied.

It has also been reported that the left hemisphere is mainly associated with identification of temporal sequences (Zatorre and Belin, 2001, Wong, 2002), while the right hemisphere is fundamentally concerned with spatial identifications. It is worth noting that the spatial factor  $W_{\text{IACC}}$  extracted from the IACF is closely related to spectral features, and thus might be expected to show right hemisphere dominance. Left-hemispheric specialization for speech signals has been reported by a number

of authors using EEG and MEG recordings (e.g., Eulitz et al., 1995, Näätänen et al., 1997, Alho et al., 1998). Because speech signals can be characterized by temporal features of autocorrelation functions (such as  $\tau_1$ ,  $\phi_1$ ,  $\tau_c$ , and patterns of local peaks  $\tau_2$ ,  $\phi_2$ ,  $\tau_3$ ,  $\phi_3$ ) neural responses to speech sounds might be expected to be associated with the left hemisphere under monaural conditions such as telephone listening. A left hemisphere specialization for speech-coded information and a right hemisphere specialization for nonverbal information (Opitz et al., 2000) have been shown using functional magnetic resonance imaging (fMRI). This reinforces the hypothesis that the left hemisphere is mainly associated with speech and identification of temporal sequences, while the right is more concerned with nonverbal and spatial identification (Kimura, 1973; Sperry, 1974).

These observations notwithstanding, when the IACC was changed for speech and music signals that would be expected to produce left hemisphere-dominant responses, the response to this change in a spatial factor was predominantly from the right hemisphere (Table 5.1). Therefore, rather than having an absolute response bias for particular kinds of signals, hemispheric response dominance is relative and depends on the nature of the factor being changed in the comparison pair.

## Chapter 6

# Temporal Sensations of the Sound Signal

The basic perceived attributes of sound can be divided into those related to a sound's perceived location in space (spatial sensations) and those qualities that distinguish different sounds independent of location (temporal sensations). Spatial sensations include sound location, size (apparent source width, ASW), and diffuseness. These subjective attributes are mainly subserved by the binaural system and can be described in terms of the interaural correlation function IACF (Chapter 7). Nonspatial and subjective attributes of sound include pitch, loudness, timbre, and duration. These perceptual qualities are grouped under the rubric of “temporal sensations” because they can be described in terms of temporal factors extracted from the monaural autocorrelation function ACF. Factors associated with temporal sensations typically predominate in neuronal responses from the left cerebral hemisphere (Fig. 5.1), while those associated with spatial sensations predominate in responses from the right hemisphere (Ando, 2006).

When a source signal is produced in a sound field, the properties of the sound field can influence the perception of nonspatial attributes. For example, reverberation time can affect loudness (Ando, 1998) and reverberations can degrade the pitches of unresolved harmonics (Sayles and Winter, 2008). Here we will discuss only the perceptual qualities of sounds in the absence of reverberation.

### 6.1 Combinations of Temporal and Spatial Sensations

The model outlined in Chapter 5 quite naturally leads to a division of primary sensations into two main categories: temporal sensations and spatial sensations. To begin with, we will discuss the relationship of subjective sensations to physical factors.

Neuropsychological models of perception attempt to describe relations between the physical attributes of external stimuli and the internal sensations they evoke. Whereas physical attributes can be publicly measured, sensations must be revealed to us either by direct experience or the overt perceptual judgments of others. Models of perception therefore involve finding mappings between physical attributes and mental, perceptual variables. Mental variables in turn are reflections of underlying

functional brain states. Psychophysical models of perception therefore include both phenomenological models that map physical stimulus variables directly onto patterns of perception and causal, mechanistic psychoneural models that first map stimulus attributes to brain states, and then brain states to mental states (experiences).

In psychophysical models, perceptual attributes associated with a given sensation can depend upon multiple physical factors, such that a single sensation  $j$  may not be well described only in terms of a single factor. Let  $X_i$  ( $i = 1, 2, \dots, I$ ) be physical factors representing cues influencing any primary sensation (temporal and spatial sensations as mentioned above), in where  $J$  is a number of significant physical factors and  $I$  is the total number of physical factors, then similar to Equation (3.5), a sensation  $S_j$  may be expressed by

$$S_j = f(x_1, x_2, \dots, x_I), j = 1, 2, \dots, J \quad (6.1)$$

If physical factors are orthogonal to each other and contribute independently to a given sensation, then  $S_j$  may be expressed by a linear combination, such that,

$$S_j = f(x_1) + f(x_2) + \dots + f(x_I), j = 1, 2, \dots, J \quad (6.2)$$

For example, let us consider the scale value of loudness, which might be described by not only the sound energy  $\Phi(0)$  and the pitch  $\tau_1$ , but also by repetitive feature of the signal ( $\phi_1$  and/or  $\tau_e$ ) and the duration of the signal  $D$ , as expressed by Equation (6.9).

Here, the question arises as to whether or not a single sensation is independent of other sensations. The simplest case with two physical factors  $x_1$  and  $x_2$  is described in Equation (6.2) so that

$$\begin{aligned} S_1 &= f_1(x_1) + f_1(x_2), j = 1 \\ S_2 &= f_2(x_1) + f_2(x_2), j = 2 \end{aligned} \quad (6.3)$$

The correlation coefficient between  $S_1$  and  $S_2$  is given by

$$r_{12} = \overline{S_1 S_2} = f_1 \overline{(x_1) f_2(x_1)} + f_1 \overline{(x_2) f_2(x_2)} + f_1 \overline{(x_1) f_2(x_2)} + f_1 \overline{(x_2) f_2(x_1)} \quad (6.4)$$

and is not zero, because in general, the first and second terms of the right-hand side are not always zero. Previously, it was believed that these perceptual attributes are largely independent of each other. For example, timbre is independent and usually not much affected by loudness (50–90 dB SPL), pitch ( $F_0$ : 200–400 Hz), or duration (100–600 ms), within limits. And we can readily identify musical instruments by their timbres irrespective of how loudly and how long any arbitrary notes are being played. However, more rigorously, we shall discuss sensations in relation to the possible dimension of physical factors.

In our auditory model, we consider differences in response patterns between the cerebral hemispheres of human listeners. Here, temporal factors are more prominent in the left hemisphere, whereas spatial factors are more prominent in the right hemisphere. Models in which an internal variable associated with each hemisphere

is represented as the linear combination of hemisphere-specific factors can explain these differences. Temporal sensations  $S_L$  and spatial sensations  $S_R$  can thus be modeled in terms of the contributions of different factors that dominate in the neural response:

$$\begin{aligned} S_L &= f_L(x_{11}) + f_L(x_{21}) + \dots + f_L(x_{11}), 1 = 1, 2, \dots L \\ S_R &= f_R(x_{1r}) + f_R(x_{2r}) + \dots + f_R(x_{1r}), 1 = 1, 2, \dots R \end{aligned} \quad (6.5)$$

where  $L + R = J$ .

Individual differences in the weighting of factors can also produce differences in sensation and preference. Even for such temporal and spatial sensations, there are substantial individual differences due to multiple physical factors, as expressed by Equations (6.1) and (6.2). Individual differences can be caused both by differing individual sensitivities and/or unique responses to the various factors. These differences of sensation and preference can be seen as characteristics of different individual listeners who have distinct auditory and visual “personalities.”

Subjective responses that are related to the overall intensity of the evoked perceptual experience (e.g., preference or annoyance) can be expressed by both temporal and spatial factors,  $S_L$  and  $S_R$ , so that

$$S = S_L + S_R \quad (6.6)$$

Going back to the theory of subjective preference of the sound field described in Section 3.3, each of the scale values in Equation (6.6) may be given by  $S_L = S_2 + S_3$  and  $S_R = S_1 + S_4$ . It is worth noting that on the subjective preference judgments for the sound field, factors such as  $\tau_1$ ,  $\phi_1$  extracted from the ACF and  $W_{IACC}$  extracted from the IACF play a minor influence. In the following sections, we shall discuss temporal sensations according to the guideline given by Equation (6.6) with limited significant temporal factors.

## 6.2 Pitch of Complex Tones and Multiband Noise

Pitches that are heard at the fundamental frequencies of harmonic complex tones have relatively straightforward correlates in patterns of major peaks in their autocorrelation functions (ACFs). The pitch period corresponds to the time delay ( $\tau_1$ ) of the first major peak. The pitch of multiband “complex noise” is also described by the value of  $\tau_1$ , and its strength is related to the value  $\phi_1$ . The autocorrelation model for pitch sensation holds for fundamental frequencies below  $\sim 4$ -5 kHz and for missing fundamentals below  $\sim 1200$  Hz.

### 6.2.1 Perception of the Low Pitch of Complex Tones

Most of the sounds in tonal music that constitute the notes of melodies and harmonies are harmonic complex tones rather than pure tones. A harmonic complex

tone consists of a series of partials whose frequencies ( $f_1, f_2, f_3, \dots, f_m$ ) are integer multiples ( $n = 1, 2, 3, \dots, m$ ) 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) can still 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  $\sim 5000$  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, whereas 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 crosscorrelation that had been proposed 3 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. Licklider’s later “triplex” model (1959) added a binaural crosscorrelation stage to the duplex model. In a similar vein, Cherry and Sayers (1956) combined autocorrelation and crosscorrelation operations to deal with issues related to aural fusion, sound separation, and directional hearing.

After a series of turns in the evolution of pitch theory (for a historical review, see de Boer, 1976), temporal models were neglected in favor of spectral pat-



tern 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 (Goldstein, 1973; Wightman, 1973a,b; 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. Because 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 1980s, 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 Noorden, 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,b). 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, neurophysiological 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 neurophysiological 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 nonspectral pitch of amplitude-modulated noise.

Analogous phenomena have also been observed for nonperiodic, inharmonic complex tones as well as nonstationary 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 et al., 1978; Yost, 1996a,b) used rippled noise stimuli to probe pitch strength, peripheral weighting, and the effects of the dominance region for pitch (Ritsma, 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, 2004).

Here we propose a model for pitch that is based on a central autocorrelation representation (ACF). The ACF model predicts the pitches not only of complex tones and ripple noise, but also of multiband complex noise with missing fundamentals. Pitch can be calculated by the delay  $\tau_1$  associated with the first major ACF peak, where pitch strength corresponds to the amplitude  $\phi_1$  of this peak. The main purpose

of the next set of experiments described below is to apply the ACF model to predict the pitch of a harmonic complex with a missing fundamental.

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 Fig. 6.1. As test signals, the two waveforms of complex tones, (a) in-phases and (b) random-phases, were applied as shown in Fig. 6.2. Starting phases of all components of the in-phase stimuli were set at zero. The phases of the components of random-phase stimuli were randomly set to avoid any periodic peaks in the real waveforms. As shown in Fig. 6.3, 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 Fig. 6.3, their NACFs are identical. The time delay at the first maximum peak of the NACF,  $\tau_1$ , equals 5 ms (200 Hz), corresponding to the fundamental frequency. Five 20- to 26-year-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 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  $0.8 \text{ m} \pm 1 \text{ cm}$ .

Pitch matching results for the five subjects are shown in Fig. 6.4. The histograms show matching frequencies within each semitone (1/12 octave) band for in-phase and random-phase stimuli. The dominant pitch match of 200 Hz is absent from the spectrum of both stimuli, and this periodicity is not at all apparent in the waveform of the random-phase signal. However, it is readily apparent in the autocorrelation functions of the two stimuli, which are identical to each other (Fig. 6.3). For both in-phase and random-phase conditions, about 60% of the responses clustered within a semitone of the fundamental. There are no major differences in the distributions of pitch-matching data between the two conditions.

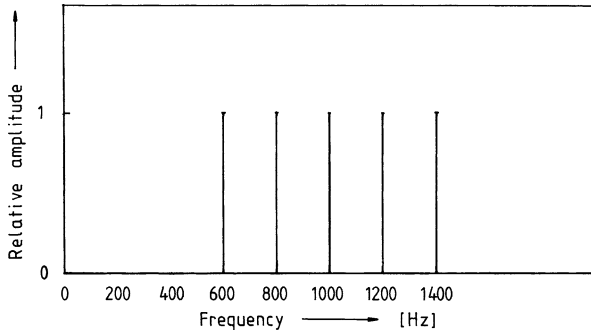
For more detail, the averaged values and standard deviations (SD) of the data obtained from each subject at frequencies near 200 Hz are listed in Table 6.1. Results obtained for pitch under the two conditions are definitely similar. In fact, the pitch strength remains invariant under both conditions. Thus, pitch of complex tones can be predicted from the time delay at the first major peak  $\tau_1$  of the NACF. This conclusion is in agreement with the findings of Yost (1996a), who demonstrated that pitch perception of iterated rippled noise determined by the first major ACF peak of the stimulus signal.

From Equation (6.6), pitch as one of temporal sensations may be expressed by

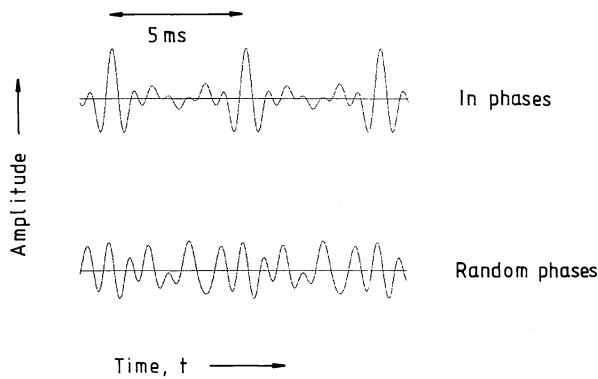
$$S = S_L = f_L(\tau_1) \approx 1/\tau_1(\text{Hz}), \quad (6.7)$$

when  $\phi_1 = 1$ .

Individual differences in pitch perception were also found. The results for each subject are indicated in Fig. 6.5. Subjects B and D matched only around the fundamental frequency (200 Hz). About 20% of the responses were clustered around 400 Hz, and the NACF has a distinct dip at  $\tau = 2.5$  ms (Fig. 6.3). However, an octave

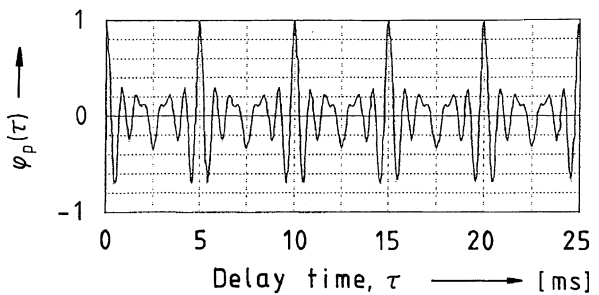


**Fig. 6.1** Complex tone presented with pure-tone components of 600, 800, 1000, 1200, and 1400 Hz without the fundamental frequency of 200 Hz



**Fig. 6.2** Waveforms of 200 Hz missing-fundamental complex tones consisting of in-phase components (*top*) and random-phase components (*bottom*)

**Fig. 6.3** Normalized autocorrelation function (NACF) of the two complex tones with different phase components,  $\tau_1 = 5$  ms (200 Hz)



shift for a phase change (Lundeen and Small, 1984) was not observed in the results obtained from these subjects. Subjects A and E matched at the fundamental frequency and at the frequency an octave higher. This octave change might be caused by a similarity for the octave relation. The time delay of the ACF for this pitch is

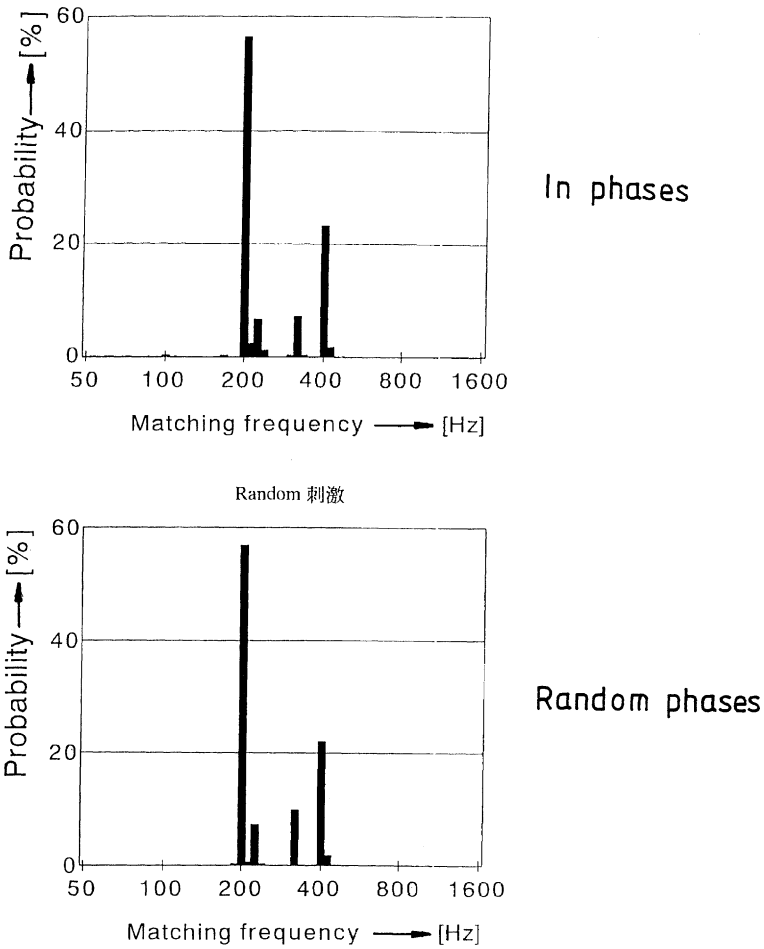


Fig. 6.4 Results of pitch-matching tests for the two complex tones,  $\tau_1 = 5$  ms (five subjects)

Table 6.1 Mean and standard deviation (SD) of the pitch-matching test for each subject

Subject	Matched frequency, mean value (Hz)		SD (Hz)	
	In-phase	Random phase	In-phase	Random phase
A	202.6	201.0	1.89	2.44
B	199.1	198.3	1.70	1.42
C	202.5	202.1	1.18	1.76
D	203.7	201.7	2.29	1.65
E	202.2	202.2	1.87	2.07
Total	201.9	201.0	2.43	2.38

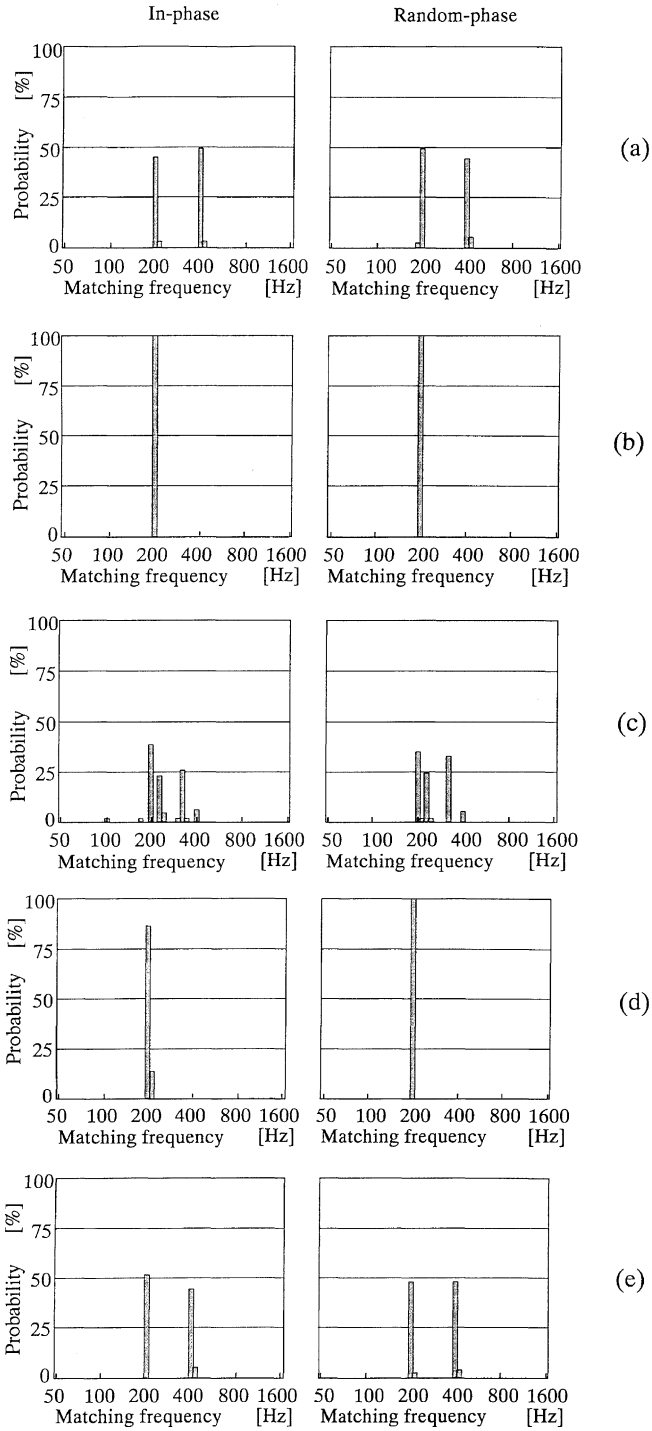


Fig. 6.5 Results of the pitch-matching tests for each of five subjects. (a-e)

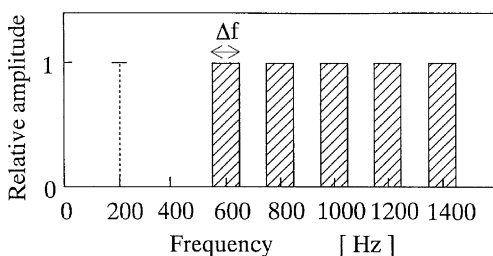
2.5 ms, so this pitch cannot be predicted because of a dip in the ACF structure. None of the subjects matched at  $\tau_1 = 10$  ms (100 Hz), which is an octave lower than the fundamental frequency, though there is a peak at  $\tau_1 = 10$  ms (Fig. 6.3). Subject C matched in three categories of center frequencies (200.0, 224.5, and 317.5 Hz). Subject C may have sought a harmonic relation because he is a musician who uses the key of E-flat. Two notes in the E-flat major triad, E-flat and G, correspond to the semitone bins that had center frequencies of 317.5 and 200 Hz respectively. Despite these categorical errors, subject C's pitch-matches in the vicinity of 200 Hz (Table 6.1) were comparable in accuracy to those of the other subjects.

## 6.2.2 Pitch of Multiband “Complex Noise”

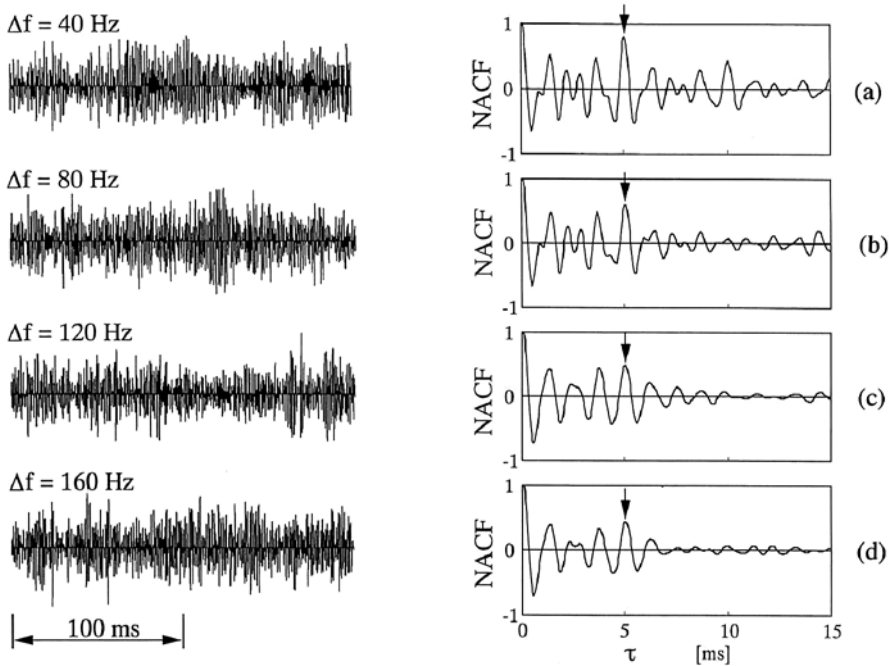
The purpose of this experiment using complex noise was to determine if the amplitude  $\phi_1$  of the first major autocorrelation peak determines the perceived strength of the pitch.

The experimental method was the same as that of the experiment described in the previous section. The bandwidths of each partial noise, which consist of the band-pass white noise with a cutoff slope of 1080 dB/octave, were changed. The center frequencies of the band noise components were 600, 800, 1000, 1200, and 1400 Hz. The complex signal consisting of band-pass noises with different center frequencies is called here “complex noise.” The bandwidths ( $\Delta f$ ) of the four components were 40, 80, 120, and 160 Hz (Fig. 6.6). Their waveforms (Fig. 6.7, left plots) have no obvious envelope periodicities. Measured results of the NACF for four conditions are shown on the right of Fig. 6.7. The amplitude of the maximum peak (indicated by arrows in the figures) in the NACF is increased with decreasing  $\Delta f$ . Four musicians from the first test and a new musician, 20 to 25 years old, participated as subjects in this experiment.

**Fig. 6.6** Multiband complex noise containing five passbands with center frequencies: 600, 800, 1,000, 1,200, and 1,400 Hz. The fundamental frequency is centered on 200 Hz



The probabilities of the matching data counted for each 1/12-octave band are shown in Fig. 6.8. All histograms show that there is a strong tendency to perceive a pitch of 200 Hz for each stimulus. This agrees with the prediction based on the value of  $\tau_1$ . These results indicate that a stimulus with a narrow bandwidth gives a stronger pitch corresponding to 200 Hz than does a stimulus with a wide bandwidth. The standard deviation (SD) for the perceived pitches increased because the value



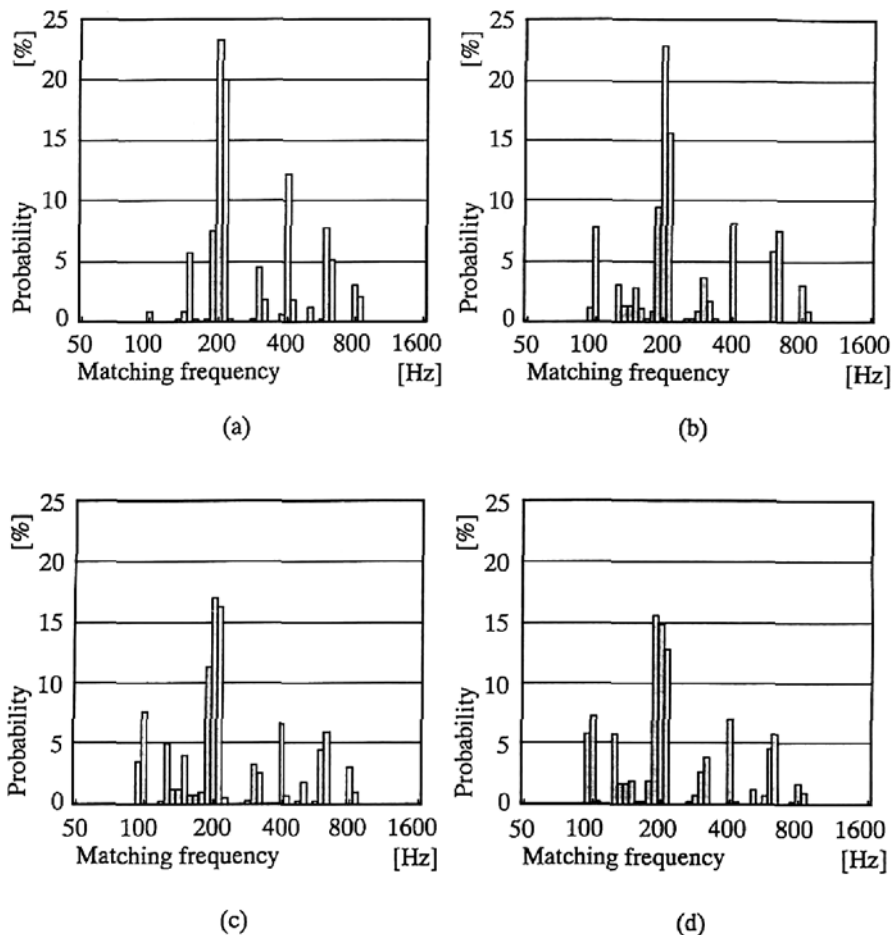
**Fig. 6.7** Waveforms and the NACF of the four complex noises applied with  $\Delta f =$  (a) 40 Hz, (b) 80 Hz, (c) 120 Hz, and (d) 160 Hz

of  $\phi_1$  decreased as  $\Delta f$  increased. The probabilities matched to 400 Hz (one octave higher than 200 Hz) keep increasing as the bandwidth becomes narrower. This is caused by the similarity of the octave relation under the pitch perception, which also appears in the first experiment. The probability of pitch around 200 Hz being identified is plotted in Fig. 6.9, as the function of the  $\phi_1$ . For narrower-band noise, the probability of a pitch of the fundamental frequency increases as the magnitude of the 5-ms peaks in the NACF increases. Thus, as  $\phi_1$  increases, pitch strength also increases ( $r = 0.98$ ). In this figure, the pitch-matching result from the previous section using the complex tones is also plotted at  $\phi_1 = 1.0$ .

Individual differences were also observed in the results obtained in tests with complex noises (Sakai et al., unpublished data). To summarize the results of this experiment, we found that the ACF model also successfully predicts the pitch of multiband complex noise stimuli with missing fundamentals.

### 6.2.3 Frequency Limits of Missing Fundamentals

We conducted a pitch-matching experiment to determine the upper frequency limit of pitches evoked by harmonic complex tones with missing fundamentals. Pitch-matching tests were conducted for two conditions: (1) for complex tones consisting



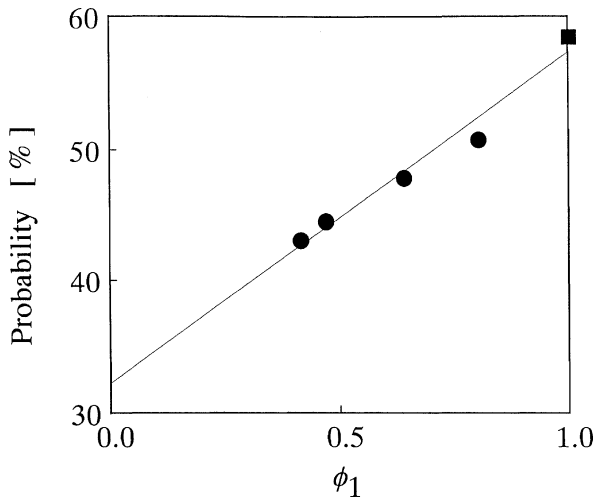
**Fig. 6.8** Results of pitch-matching tests, with all five subjects.  $\Delta f$ : (a) 40 Hz, (b) 80 Hz, (c) 120 Hz, and (d) 160 Hz

of harmonics 2-4 (i.e.  $2F_0$ ,  $3F_0$ ,  $4F_0$ ) of fundamental frequencies  $F_0$ s of 500, 1000, 1200, 1600, 2000, or 3000, Hz and (2) for complex tones consisting only of harmonics 2 and 3 (i.e.  $2F_0$ ,  $3F_0$ ). It was found that (1) the ACF model holds for missing fundamental frequencies up to roughly 1200 Hz; (2) within this frequency range, the pitch can be reliably matched to the missing fundamental frequency even if the harmonic complex consists of only two tones.

For fundamental frequencies of 500, 1000, 1200, 1600, 2000, and 3000 Hz, stimuli consisting of two or three pure-tone components were produced in a computer (Inoue et al., 2001). The two-component stimuli consisted of the second and third harmonics of the fundamental frequency, and the three-component stimuli consisted of the second, third, and fourth harmonics. The starting phase of all components was

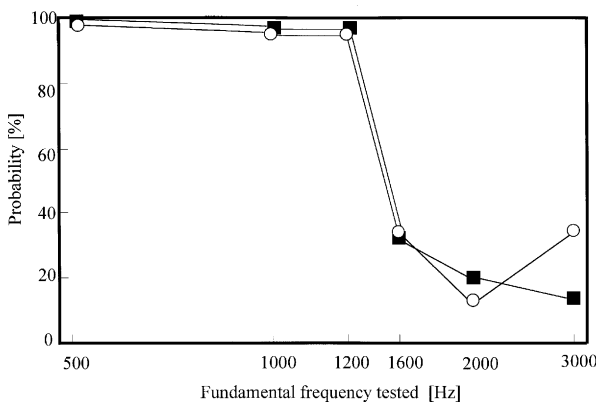


**Fig. 6.9** Relationship between  $\phi_1$  and probability of the pitch being within  $200 \pm 16$  Hz ( $r = 0.98$ ,  $p < 0.01$ ). For reference, the plot (■) at  $\phi_1 = 1$  is the result with the pure tone



adjusted to zero (in phase). The total SPL at the center of the listener’s head was fixed at 74 dB. The NACF of all stimuli was calculated obtaining the peak  $\tau_1$  related to the fundamental frequency. The loudspeaker was placed in front of a subject in an anechoic chamber. The distance between the center of the subject’s head and the loudspeaker was 0.8 m. Three 21- to 27-year-old musicians participated as subjects in the experiment. Pitch-matching tests were conducted using complex tones as test stimuli and a pure tone generated by a sinusoidal generator as a reference.

Pitch matches for all subjects are shown in Fig. 6.10. Whenever the missing fundamental frequency of the stimulus was 500, 1000, or 1200 Hz, more than 90% of



**Fig. 6.10** Probability that three subjects adjusted a pure tone near the fundamental frequency of complex tones. *Empty circles* are results for two harmonics, and *full squares* are those for three harmonics

the responses obtained from all subjects under both conditions clustered around the fundamental frequency. When missing fundamentals were 1600, 2000, or 3000 Hz, however, the probability that the subjects matched the frequency of the pure tone to the fundamental frequency was much lower. These results imply that the ACF model is applicable when stimuli have missing fundamentals of 1200 Hz or less.

The reasons for this upper limit are fairly straightforward. According to neuronal autocorrelation models, in order to evoke a “missing fundamental,” one needs to satisfy at least one of two conditions that involve respectively either individual, cochlear-resolved harmonics or envelopes of unresolved, interacting harmonics (Cariani and Delgutte, 1996a,b). In the first mode, interspike intervals associated with individual harmonics are produced and summed together across the auditory nerve. Here one needs at least two resolved harmonics that are below the limit of significant phase-locked temporal information ( $\sim 4000$  Hz), such that interspike intervals associated with their common subharmonic, the fundamental, will predominate in the pooled ACF representation. In the second mode, pairs of unresolved adjacent harmonics beat together to produce interspike intervals associated with their beat period, which is the fundamental period. For several reasons, this mechanism that is based on interval representation of the stimulus envelope is less effective at producing intervals close to the fundamental period, and as a consequence, the pitches evoked are weaker than those associated with the first mechanism. In the current context, in order to represent a 1500-Hz missing fundamental using the envelope-based mechanism, one would need several pairs of unresolved harmonics, all at 9000 Hz or above ( $n > 5$ , for  $F_0 = 1500$  Hz,  $f_n > 9000$  Hz). Because there are relatively few auditory nerve fibers in humans that are responsive to such high frequencies, and intervals from all regions are pooled together, intervals associated with envelopes in these frequency regions are dwarfed by the spontaneous activity in the rest of the auditory nerve. The result is that the interval peaks associated with the  $F_0$  envelope period are very shallow and do not rise above the signal/background threshold required for an audible low pitch [in Cariani and Delgutte (1996a), AM tones with 6400-Hz carriers and  $F_0$ s from 80 to 320 Hz presented at 60 dB SPL did not generate enough intervals to exceed this signal/background threshold].

On the low-frequency side of fundamental pitch perception, in psychophysical experiments, the lowest periodicities that produce clear pitches capable of supporting melodic recognition are approximately 30 Hz (Pressnitzer et al., 2001). This may be a consequence of a limitation in the longest interspike intervals that central auditory pitch processors analyze. Many current ACF models of pitch and consonance (e.g., Cariani, 2001, 2002, 2004) therefore use a tapering interval weighting system that eliminates from consideration intervals longer than  $\sim 33$  ms.

It is worth noting that results of evoked magnetic response (N1m latency) correspond to the fundamental frequency down to 19 Hz (Yrttiaho et al., 2008). Thus, Equation (6.7) could hold for the fundamental frequency,  $19 \text{ Hz} < f_L(\tau_1) \leq 1200 \text{ Hz}$ .

So far, we have come to the following two conclusions:

1. For low pitches of complex tones, the ACF model is applicable when a missing fundamental frequency is below 1200 Hz, and probably above 19 Hz.

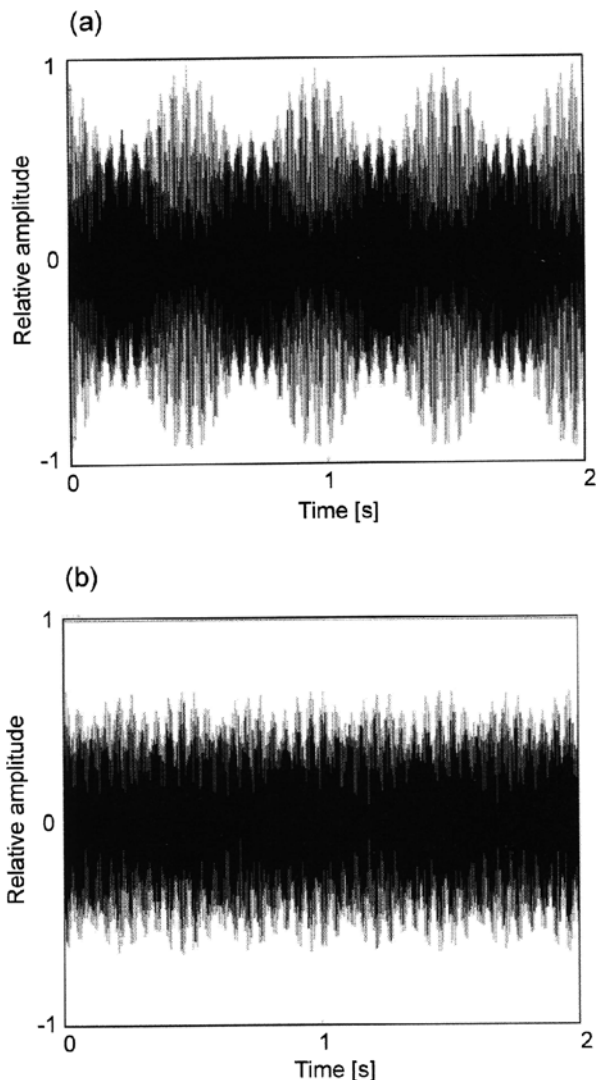
2. Within this periodicity range, even if a complex tone with a missing fundamental has only two harmonics ( $n=2,3$ ), the pitch corresponds to the delay time of the first major peak in the NACF.

### 6.3 Beats Induced by Dual Missing Fundamentals

It has been observed that a monaural beat can be induced by two complex tones that have slightly different missing fundamentals. This is true even when the envelope-beat component is realized by random-phase components. The beat stimulus is constructed by mixing two complex tones,  $A$  and  $B$ , that have missing fundamental frequencies at  $F0_a$  and  $F0_b$  respectively. When we listen to these complex tones  $A$  and  $B$  together, a beat is heard that corresponds to the frequency difference between the fundamentals ( $\Delta f = F0_b - F0_a$ ). When all components are in phase, the composite tone ( $A + B$ ) has a waveform repetition and envelope periodicity corresponding to  $\Delta f$ . When the components are in random phase, however, the envelope periodicity  $\Delta f$  disappears. Experimental results show that, in both cases, beats of  $\Delta f \leq 4$  Hz were perceived clearly for stimuli with missing fundamentals up to 256 Hz. These results show that beats that are independent of the envelope component can be detected. These phenomena can be explained in terms of the delay time of the maximum peak extracted from the ACF of the sound signal.

An experiment on monaural beats induced by two complex tones with missing fundamentals was conducted (Shimokura and Ando, 2004). Each stimulus signal consisted of two complex tones,  $A$  and  $B$ , mixed together. Let  $F0_a$  and  $F0_b$  be the fundamental frequencies of  $A$  and  $B$ , respectively, each consisting of upper harmonics ( $n \geq 8$ ).  $\phi_a$  and  $\phi_b$  are phases of complex frequency components. Amplitudes of all components were equal. Fundamental frequencies  $F0_a$  of the first tone  $A$  were either 32, 64, 128, 256, or 512 Hz, while the fundamentals  $F0_b$  of the second tone  $B$  differed from the first by 2, 4, 8, or 16 Hz, respectively, so that  $F0_b = F0_a + \Delta f$ . The lowest component of  $A$  was always fixed at 1,024 Hz. For example, when  $F0_a = 128$  Hz and  $\Delta f = 2$  Hz, then the components of  $A$  consisted of three harmonics  $n = 8-10$  of 128 Hz: 1,024, 1,152, and 1,280 Hz, while the components of  $B$  ( $F0_b = 130$  Hz) were harmonics  $n = 10-12$  of 1,300, 1,430, and 1,560 Hz. The total peak sound pressure level measured at the center position of the center of the head was fixed at 74 dB SPL.

Figure 6.11a shows the waveforms of the two complex tones with  $F0_a = 128$  Hz and  $\Delta f = 2$  Hz. When components are in-phase, the envelope has a periodicity corresponding to  $\Delta f$ . When phases are random, however, the envelope regularity disappears as shown in Fig. 6.11b. Figure 6.12 shows the results of the ACF analysis of the two complex tones. As is well known, the ACF is identical for in-phase and out-of-phase signals. The maximum peak,  $\tau_1 = 0.5$  s, corresponds to 2 Hz. In the ACF for  $\tau < 10$  ms shown in Fig. 6.12b, two initial fundamental frequencies (128 and 130 Hz) are indicated by arrows.



**Fig. 6.11** Waveforms of two complex tones. (a) In-phase condition. (b) Random condition

Three 23- to 24-year-old subjects participated in the beat matching test. Each subject was seated in the listening room, and the same sound signal was fed to the two ears via headphone (Sennheiser, HE60). First, subjects were presented with the two combined complex tones and were asked to listen for a single beat in the sound signal. Then, subjects were presented a train of pulse tones generated by an oscillator and asked to adjust the pulse rate to match the beat perceived for the combined complex tones. This process was repeated until subjects could match an identical beat.

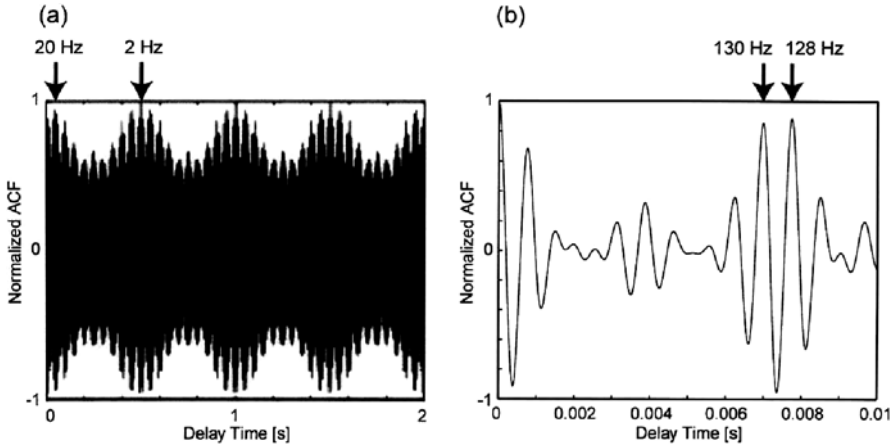


Fig. 6.12 NACF analyzed for both conditions. (a)  $\tau \geq 2.0$  s. (b)  $\tau \leq 0.01$  s

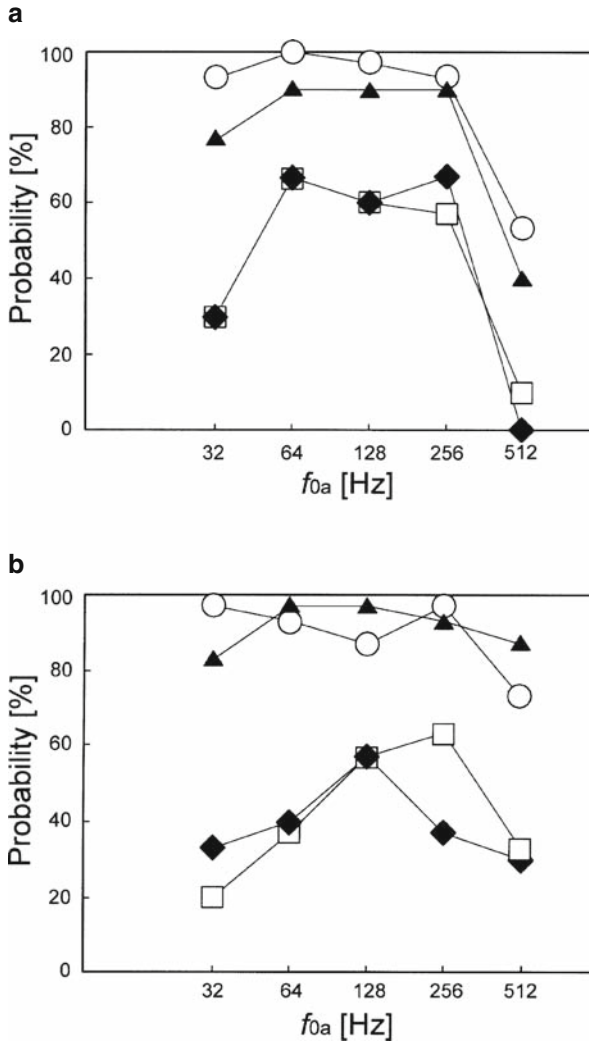
Beat-matching tests for the set of five stimuli were presented in random sequence, with each composite complex tone being presented a total of 10 times.

Figure 6.13 shows the probability of the subjects correctly matching the perceived beat frequency by adjusting the pulse rate  $\Delta f$  within the one-third octave that was generated separately. This is shown as a function of the fundamental frequency  $F0_a$  and as a parameter of the beat frequency  $\Delta f$ . What we find remarkable is that when  $\Delta f = 2$  to 4 Hz, the probabilities for the frequency range of  $F0_a = 32$  to 256 Hz almost always exceeded 80% for both in-phase and out-of-phase conditions. When  $\Delta f = 8$  to 16 Hz, the probabilities all decreased below 65%. For in-phase conditions, beat-matching probabilities were smaller than those for out-of-phase conditions, only except for one condition,  $F0_a = 512$  Hz ( $p < 0.025$ ). This beat that is perceived is independent of the envelope of the waveform; consequently it was distinguished from an envelope beat.

As discussed previously, the pitch-matching test of the single complex tone shows that the listeners hear a pitch at the fundamental frequency, which can be described by the delay time of the first peak in the ACF below 1200 Hz (Inoue et al., 2001). However, the beat phenomenon induced by the dual missing fundamentals was observed in the range of  $32 \text{ Hz} < F0_a < 256 \text{ Hz}$ . The periodicity-limiting mechanism of this fundamental frequency range for  $F0_a$  is unknown.

These experiments demonstrated that:

1. Fundamental frequencies of multiple complex tones can induce an additional secondary fundamental frequency that is perceived as a beat. The perceived beat rate corresponds to the delay time of the maximum peak of the ACF of the whole signal.
2. The perceived beat was independent of the existence of regularities and fluctuations in the waveform envelope of the two-tone stimulus. This rule holds for fundamental frequencies below 256 Hz.



**Fig. 6.13** Probability match to the beat calculated by  $\tau_1$  as a function of the fundamental frequency  $F_{0a}$ . **(a)** In-phase condition. **(b)** Random condition. The different symbols indicate the beat frequency as the secondary fundamental frequencies,  $\Delta f$ :  $\circ$ , 2 Hz;  $\blacktriangle$ , 4 Hz;  $\square$ , 8 Hz;  $\blacklozenge$ , 16 Hz

## 6.4 Loudness

### 6.4.1 Loudness of Sharply Filtered Noise

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 ( $\tau_e$ ) 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.

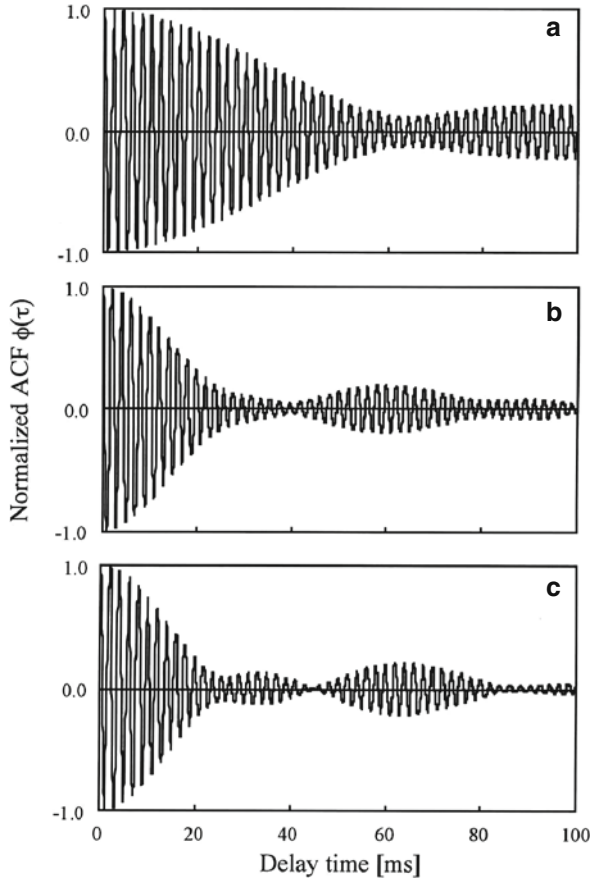
Previous studies on the relationship between loudness and the bandwidth of noise have concluded that for sounds having the same SPL, loudness remains constant as bandwidth increases, up until the bandwidth reaches the “critical band.” For bandwidths larger than the critical band, loudness increases with bandwidth (Zwicker et al., 1957). The spectral characteristics of the filters used in those studies were not specified, except by Greenwood (1961a,b). 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 response of the filter outside its  $-3$  dB bandwidth greatly affects the repetitive feature of the signal (temporal coherence), represented by the effective duration  $\tau_e$  that is extracted from ACF representations (Fig. 5.1; Ando, 1998; Ando et al., 1999). Such sharp filters may exist in the auditory system for high frequencies (Katsuki et al., 1958). In any case, a roll-off of more than 1000 dB/octave is required. The loudness of a sharply (1080 dB/octave) filtered noise increases as the effective duration of the normalized ACF ( $\tau_e$ ) increases, even if the bandwidth of the signal is within the critical band. It is worth noting that we observed that when the subsequent reverberation time ( $T_{\text{sub}}$ ) of a sound field increases, effective duration  $\tau_e$  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 SPL is fixed at a constant value, the scale value of loudness  $S$  is expressed by

$$S = S_L = f_L(\tau_1, \phi_1, \tau_e, D) \quad (6.8)$$

where the factors are defined in Section 5.2;  $W_{\phi(0)}$  is excluded in the above equation because the center frequency of the noise is fixed and is represented by  $\tau_1$ , 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 the maximum audio frequency. Thus, the value  $10\log\Phi(0)/\Phi(0)_{\text{ref}}$  is far more accurate than any factor based on the envelope of the waveform,  $\Phi(0)_{\text{ref}}$  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 SPL 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 white noise and then filtered it. The source signal of band-pass noises is characterized in terms of their ACFs as shown in Fig. 6.14. Bandwidth ( $\Delta f$ ) was changed by using a sharp filter with the cutoff

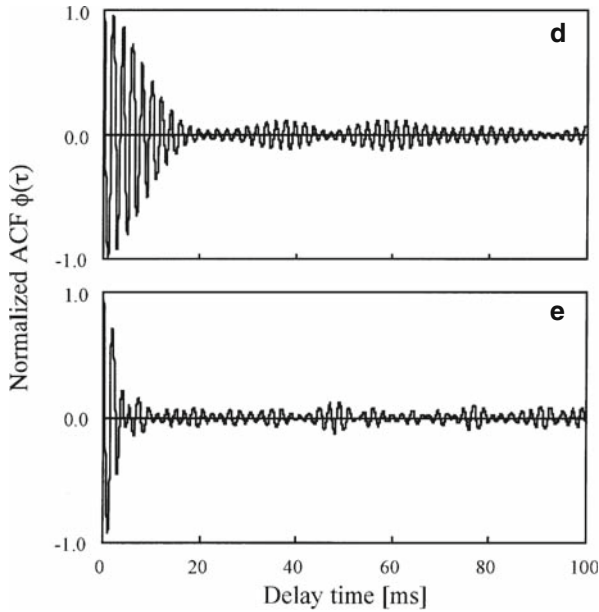


**Fig. 6.14** Examples of the NACF analyzed for the center frequency of 500 Hz. The filter bandwidth  $\Delta F$ : (a) 0 Hz, (b) 40 Hz, (c) 80 Hz, (d) 160 Hz, (e) 320 Hz

slope of 2068 dB/octave, which was realized by a combination of two filters. Factors of  $\tau_1$ ,  $\tau_e$ , and  $\phi_1$  analyzed are shown in Fig. 6.15. In fact, the filter bandwidth of 0 Hz included only its slope component. All source signals were the same SPL at 74 dBA, which was accurately adjusted by measurement of the ACF at the origin of the delay time,  $\Phi(0)$ . As one can readily see, whereas filter bandwidth has absolutely no effect on the signal's dominant periodicity (as reflected by  $\tau_1$ ), it has a profound effect on the slope of the ACF envelope (effective duration, as reflected by  $\tau_e$ ) and a lesser effect on the relative height of the peak associated with the dominant periodicity  $\phi_1$ .

Loudness judgement experiments utilized paired comparison tests in 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



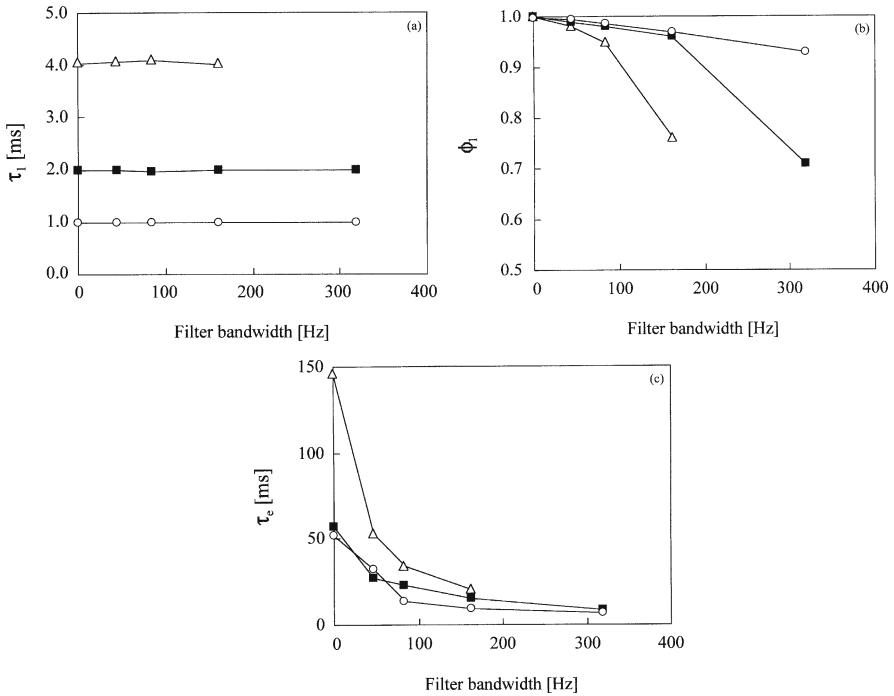


**Fig. 6.14** (continued)

hearing participated in the experiment. They were seated in an anechoic chamber and asked to judge which of two paired sound signals was 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  $\times$  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 ( $p < 0.05$ ) consensus 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 Fig. 6.16. The scale value difference of 1.0 corresponds to 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  $\tau_e$  and large bandwidths, with minima at smaller bandwidths (40, 80, 160 Hz, respectively). From the dependence of  $\tau_e$  on filter bandwidth, we found that loudness increases with increasing  $\tau_e$  almost within the "critical bandwidth." Results of analysis of variance for the scale values of loudness indicated that for all center frequencies tested, the scale values for the loudness of pure tones were significantly larger than those for other band-pass noises within the critical band



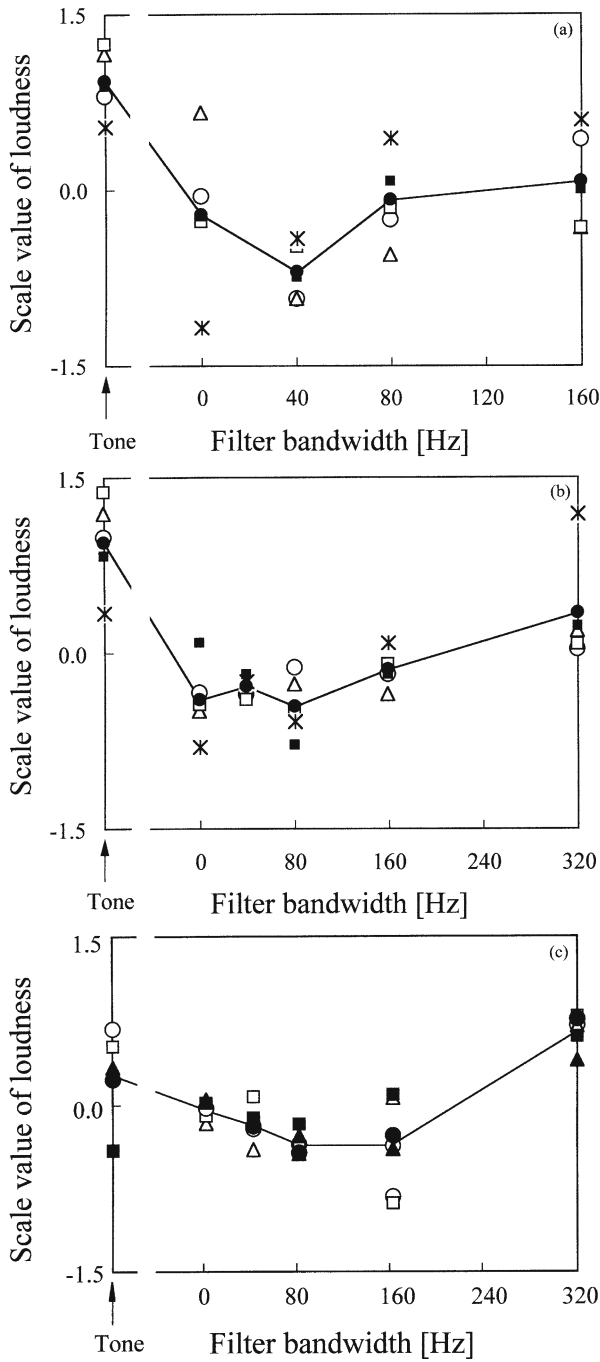
**Fig. 6.15** Measured factors extracted from the ACF of the source signal as a function of the bandwidth. Different symbols indicate different frequencies.  $\Delta$ , 250 Hz;  $\blacksquare$ , 500 Hz;  $\circ$ , 1000 Hz. (a) Delay time of the first peak of ACF ( $\tau_1$ ). (b) Amplitude of the first peak of the ACF ( $\phi_1$ ). (c) Effective duration of ACF ( $\tau_e$ )

( $p < 0.01$ ). When the effects of changes in the reverberation time  $T_{\text{sub}}$  of the sound field on loudness are taken into account, the conclusion is that the factor  $\tau_e$ , a measure of repetitive features of the sound signal, contributed to the loudness that is perceived (Merthayasa et al., 1994).

Consequently, loudness of the band-pass noise with identical SPL 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  $\tau_e$  within the critical band. Therefore, Equation (6.8) within the critical band may be reduced by

$$S = S_L = f_L(\tau_1) + f_L(\tau_e) \quad (6.9)$$

In fact, MEG records of auditory-evoked magnetic fields showed that the N1m magnitude decreases with increasing bandwidth when the bandwidth is less than the critical bandwidth. This N1m peak magnitude also increases with increasing bandwidth beyond the critical band (Soeta et al., 2005).



**Fig. 6.16** Scale values of loudness as a function of the bandwidth of noise. Different symbols indicate the scale values obtained with different subjects. (a)  $f_c = 250$  Hz. (b)  $f_c = 500$  Hz. (c)  $f_c = 1000$  Hz

### 6.4.2 Loudness of Complex Noise

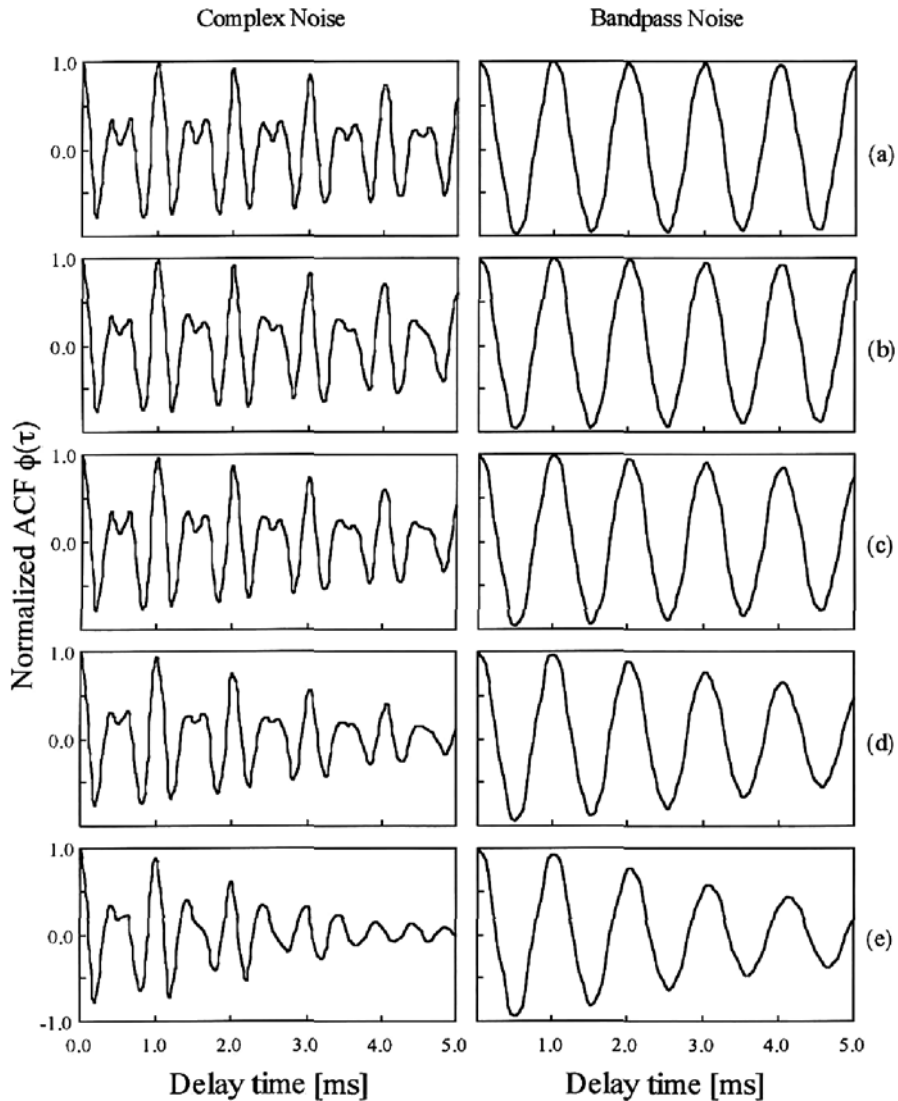
The loudness of spectrally complex noise, which includes band-pass noises, with multiple pass bands and center frequencies, is examined as a function of the effective duration of the ACF ( $\tau_e$ ). The center frequencies of two components of the complex noise were 2000 and 3000 Hz, so that the perceived pitch was centered on 1000 Hz due to the missing fundamental phenomenon as discussed in Section 6.2.2. To control the  $\tau_e$  of the source signal, the bandwidth of each component noise was modified using a 2068 dB/octave sharp filter. Scale values for loudness obtained using the PCT were similar to those for single noise components centered on 1000 Hz. As with single noise components, loudness increases with the value of  $\tau_e$  of the source signal, and the loudness of the complex noise with identical SPL is not constant within the critical band, 160 Hz.

This study examined the loudness of spectrally complex noise (Sato et al., 2001). The complex noises used consisted of multiple band-pass noises whose passband center frequencies were harmonics of a 1000-Hz fundamental. There were no correlations between the noise bands. The perceived pitch was centered on 1000 Hz, the “missing fundamental” of the noise bands. Perceptual judgments by listeners were compared with those for the single band-pass noise of 1000 Hz and 2000 Hz center frequencies in terms of the factors extracted from the ACF.

Source signals in the experiments included: (1) a complex noise stimulus with two band-pass noise components whose center frequencies were 2000 and 3000 Hz, and (2) a complex tone with pure-tone components of 2000 and 3000 Hz. All partial components had the same SPL, measured by taking the square root of the ACF of the recorded signal at the origin of the delay time,  $\Phi(0)$ . To control the  $\tau_e$  of the ACF of the complex noise, the bandwidths of each partial noise ( $\Delta f$ ) were changed to 0, 40, 80, 160, and 320 Hz with the cutoff slope of 2068 dB/octave. In the 0 Hz bandwidth condition only the slope component of the filter was used. Figure 6.17 shows the normalized ACF of complex noise with fundamental frequencies of 1000 Hz and that of a single noise component centered on 1000 Hz. As shown in Fig. 6.18a, all of the ACFs indicate the maximum peak at  $\tau_1 = 1.0$  ms. Figure 6.18b and c show the measured  $\phi_1$  and  $\tau_e$  of the source signals as a function of the bandwidth.

Loudness judgments were performed using paired comparison tests (PCT). Pair-wise comparisons were made using the complex tones and five complex noises ( $\Delta f = 0, 40, 80, 160, 320$  Hz). The same source signal was presented to both ears through headphones. The magnitude of the IACC was thus kept constant at unity. All stimuli were fixed at the same SPL at 74 dBA by measurement of  $\Phi(0)$ . SPL was calibrated by using a dummy head with 1/2-inch condenser-type microphones at both ears. Input signals were digitized at 24 kHz sampling frequency. Fluctuation of the measured  $\Phi(0)$  for all stimuli were within  $\pm 0.06$  dB when the duration of the signals was lengthened beyond 0.8 s; therefore, the stimulus duration was chosen at 1.0 s.

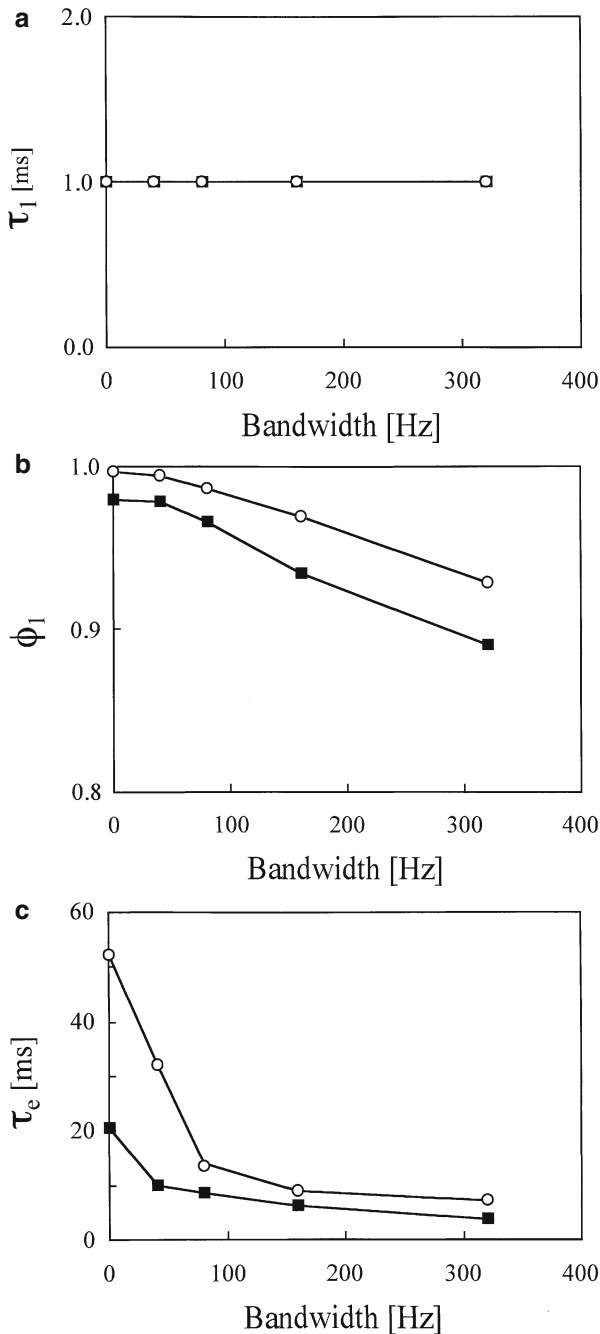
Four subjects with normal hearing ability were seated in the anechoic chamber and asked to judge which of two sound signals they perceived to be louder. The rise and fall times were 50 ms, and the silent interval between the stimuli was 0.5 s. Each



**Fig. 6.17** Normalized ACFs of complex multiband noises with fundamental frequencies of 1000 Hz (*left-hand side*) and single band-pass noises with 1000 Hz center frequency (*right-hand side*) for different passband bandwidths ( $\Delta f$ ). (a)  $\Delta f = 0$  Hz. (b)  $\Delta f = 40$  Hz. (c)  $\Delta f = 80$  Hz. (d)  $\Delta f = 160$  Hz. (e)  $\Delta f = 320$  Hz

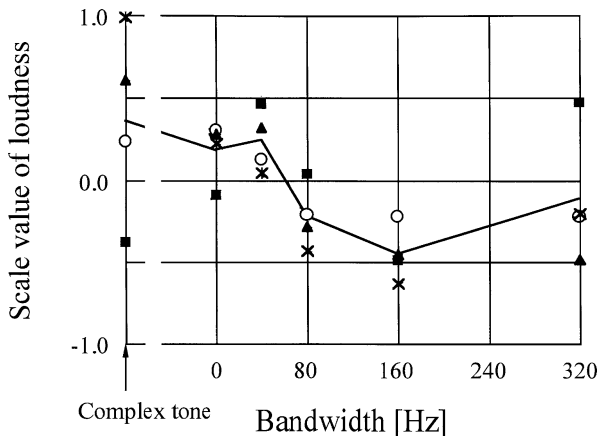
pair of stimuli was separated by an interval of 3.0 s, and the pairs were presented in random order. A single test session consisted of 15 pairs [ $N(N - 1)/2$ ;  $N = 6$ ] of stimuli and lasted about 1.5 minutes. Ten sessions were performed for each subject.

Forty responses (4 subjects  $\times$  10 sessions) to each stimulus were obtained. Consistency tests indicated that all subjects had a significant ( $p < 0.01$ ) ability to



**Fig. 6.18** Measured factors extracted from the ACF of the source signal as a function of the bandwidth: ■, complex multiband noises with fundamental frequencies of 1000 Hz; ○, single band-pass noises of 1000 Hz center frequency. (a)  $\tau_1$ . (b)  $\phi_1$ . (c)  $\tau_e$

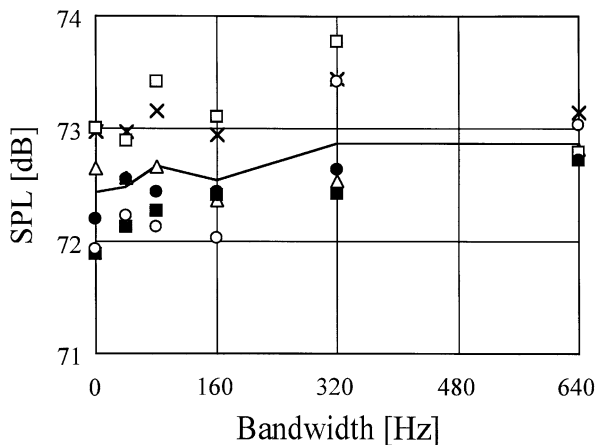
discriminate loudness. The test of agreement also indicated that there was significant ( $p < 0.05$ ) agreement among all subjects. Scale values for loudness were obtained by applying the law of comparative judgment. The relationship between the scale value of loudness and the bandwidth of each partial component of the complex noise with its fundamental frequency of 1000 Hz is shown in Fig. 6.19. The minimum loudness was observed for a bandwidth of 160 Hz. Loudness increased with increasing  $\tau_e$  of the source signal within the bandwidth of 160 Hz, as shown in Fig. 6.19 for the complex noise. Analysis of the variance for the scale values of loudness showed that there were significant differences between the pairings of a complex tone and 160 Hz; 0 and 80 Hz; 0 and 160 Hz; 40 and 80 Hz; and 40 and 160 Hz. We find it remarkable that the pattern of the perceived loudnesses of single band-pass noises with 1000 Hz center frequencies and varying bandwidths closely resembled that of multiband complex noises with pseudo-fundamentals at 1000 Hz (compare Fig. 6.19 with Fig. 6.16c).



**Fig. 6.19** Scale values of loudness as a function of the bandwidth for complex noises with fundamental frequencies of 1000 Hz. Different symbols indicate the scale values obtained with different subjects (four subjects)

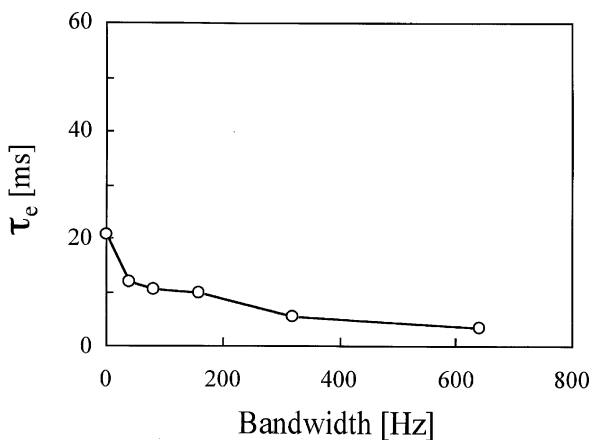
All of the source signals used in this experiment had a fundamental frequency of 1000 Hz, and in fact the measured  $\tau_1$  was 1.0 ms. In our preliminary experiment, a different set of subjects matched the pitch of the complex noise to 1000 Hz. The pitch of a complex tone consisting of the second and third harmonics corresponds to  $\tau_1$  for fundamental frequencies below 1200 Hz. For missing fundamentals above 1200 Hz, the probability of accurate pitch matching rapidly decreases (see Section 6.2.3).

As shown in Fig. 6.20, the loudness of the sharply (2068 dB/octave) filtered single band-pass noise centered on 2000 Hz that was obtained by constant method is flat up to 160 Hz, although the  $\tau_e$  increases as bandwidth decreases, as shown in Fig. 6.21. Thus, loudness may be described in relation to effective duration,  $\tau_e$ , for fundamental frequencies below 1200 Hz, which is the upper limit of missing



**Fig. 6.20** Loudness of single band-pass noises centered at 2000 Hz obtained by the constant method comparing the 2000-Hz tone as a function of the bandwidth. Different symbols indicate the loudness obtained with different subjects (six subjects)

**Fig. 6.21** Measured factor  $\tau_e$  extracted from the ACF of the band-pass noise of 2000-Hz center frequency as a function of the bandwidth



fundamental percepts. In these experiments, we found that loudness for the complex noise with fundamental frequencies of 1000 Hz is similar to that of the single noise component centered on 1000 Hz. This is because both signals have the same  $\tau_1$ . Also, loudness increases with the increasing value of  $\tau_e$  within the critical band of 1000 Hz. Loudness for the band-pass noise, centered on 2000 Hz, is not affected by the value of  $\tau_e$  because of the limitation on the ACF model. It is worth noting we should take the spatial factors extracted from the IACF into consideration when estimating the perceived loudness of a sound in the context of a sound field (e.g., Edmonds and Culling, 2009).



## 6.5 Duration Sensation

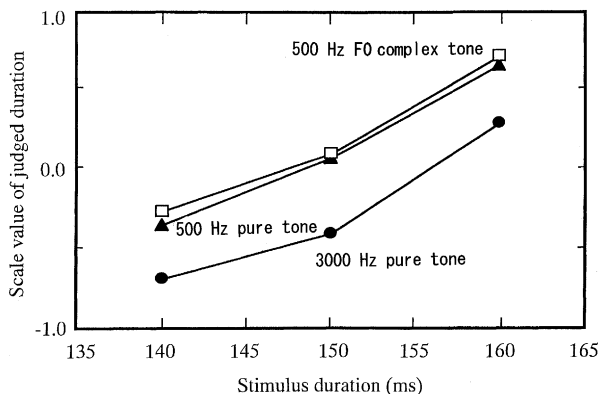
The sensation of temporal duration (DS) is introduced here. Every auditory event that is detected, including musical notes and speech sounds, has a perceived duration. Obviously, the sense of duration depends most directly on the physical signal duration,  $D$ . But apparent duration can be influenced by other acoustic factors as well. In terms of internal auditory representations, we show here that perceived duration also covaries with tone frequency and with its corresponding temporal factor in the ACF,  $\tau_1$ , and consequently, with the pitch that is perceived (Ando et al., 2002).

This experimental study probed differences in the perception of duration for pure and complex tones for two frequencies (Saifuddin et al., 2002). The experiment used paired comparisons (PCT). Sound pressure level was fixed at 80 dBA throughout and all waveform amplitudes were ramped during onsets and offsets with rise/fall times of 1 ms, i.e., the time required to reach a threshold 3 dB below the steady level.

Perceived durations of the two-component complex tones (3,000 and 3,500 Hz) having a fundamental at 500 Hz were compared with those evoked by pure tones with frequencies of 500 or 3,000 Hz. Pairs of 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 directly in front of the center of the seated listener's head at a horizontal distance of  $74 (\pm 1)$  cm. Ten 22- to 36-year-old subjects with normal hearing 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 nine stimuli are shown in Fig. 6.22. Whereas 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, whereas durations for pure tones of different frequencies (500 vs. 3000 Hz) differed significantly. This difference was approximately 10 ms (judging from equivalent scale values, the 500 Hz pure tone appeared about 10 ms longer than the 3,000 Hz tone). Thus, the perceived duration (DS) of the higher frequency pure tone (3000 Hz;  $\tau_1 = 0.33$  ms) was found to be significantly shorter ( $p < 0.01$ ) than that of either the lower frequency pure tone ( $f = 500$  Hz;  $\tau_1 = 2$  ms) or the complex tone ( $F_0 = 500$  Hz;  $\tau_1 = 2$  ms). Also, the scale values of duration sensation between the two pure tones:  $\tau_1 = 2$  ms (500 Hz) and 0.33 ms (3,000 Hz) are almost parallel, so that the effects of periodicity ( $\tau_1$ ) and signal duration ( $D$ ) on the apparent duration (DS) are independent and additive. These relations are expressed in Equation (6.10), where  $S$  is the scale value of the perceived duration.

$$S = S_L = f_L(\tau_1, D) = f_L(\tau_1) + f_L(D) \quad (6.10)$$



**Fig. 6.22** Scale values of DS obtained by the PCT:  $\square$ , complex tone ( $F_0 = 500$  Hz) with 3000-Hz and 3500-Hz pure-tone components;  $\blacktriangle$ , 500-Hz pure tone;  $\bullet$ , 3000-Hz pure tone

Here  $\tau_1$  is extracted from the stimulus ACF. Figure 6.23 shows the normalized stimulus ACF in which  $\tau_1$  corresponds to the missing fundamental, i.e., the pitch that is heard for fundamental periodicities below roughly 1,200 Hz (Section 6.2.3).

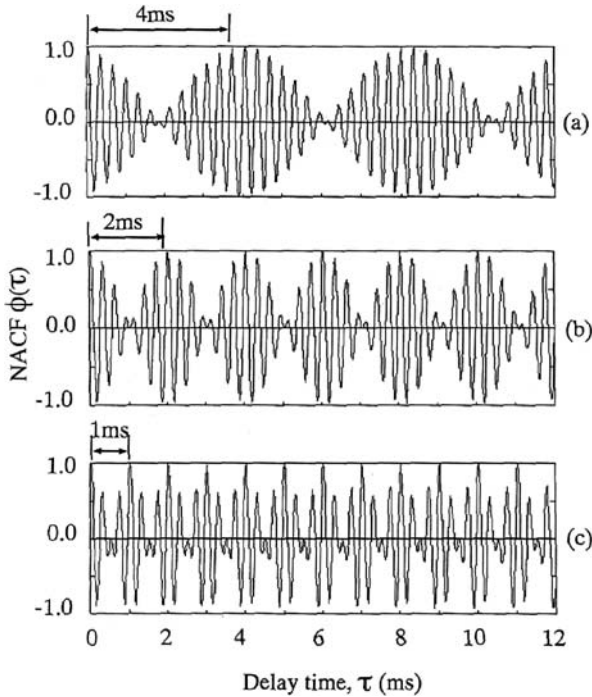
Scale values of individual listeners were also compiled (see Section 9.1 in Ando, 1998). Goodness of fit results for the two-factor model of duration perception are listed in Table 6.2 for 10 subjects. These individual data confirmed the above-mentioned results within the range of 1 standard deviation, except for subjects M.K. and K.A., whose value of  $d$  reflected poor fits, exceeding 22.2% and 19.4% ( $K > 7$ ), respectively.

The significant results of this study are summarized below.

1. Apparent stimulus duration DS depends primarily on the duration of the signal and secondarily on signal periodicity  $\tau_1$  (pure-tone frequency or complex-tone fundamental frequency).
2. Effects of the  $\tau_1$  extracted from the ACF on DS are almost the same on the scale value for the pure-tone ( $\tau_1 = 2$  ms) and complex-tone ( $\tau_1 = 2$  ms) stimuli. The apparent duration DS of the pure-tone stimulus ( $\tau_1 = 0.33$  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 ( $\tau_1 = 2$  ms =  $1/500$  Hz).
3. Apparent duration DS can be readily expressed as a function of  $D$  and  $\tau_1$  for both pure and complex tones.

## 6.6 Timbre of an Electric Guitar Sound with Distortion

Timbre is defined as an aspect of sound quality that is independent of loudness, pitch, and duration. It encompasses those perceived qualities of sound that distinguish two notes of equal pitch, loudness, and duration that are played on



**Fig. 6.23** Demonstrations of the NACF analyzed for the complex tone. (a) Complex tone with the components of 3000 Hz and 3250 Hz ( $F_0 = 250$  Hz). (b) Complex tone with the components of 3000 Hz and 3500 Hz ( $F_0 = 500$  Hz). (c) Complex tone with the components of 3000 Hz and 4000 Hz ( $F_0 = 1000$  Hz)

**Table 6.2** Results of tests of goodness of fit for 10 subjects. (For the method of goodness of fit, see Ando and Singh, 1996; Ando, 1998)

Subject	$K^1$	$d$ (%) <sup>2</sup>
M.K.	8	22.2
D.G.	6	16.7
S.K.	6	16.7
M.N.	4	11.1
K.A.	7	19.4
N.K.	6	16.7
D.B.	4	11.1
N.A.	5	13.9
M.A.	5	13.9
S.S.	6	16.7

<sup>1</sup>  $K$  is the number of poor responses.

<sup>2</sup>  $d = 2 K/F(F - 1)$ , where  $F$  is the number of stimuli used for the judgment. In this investigation,  $F = 9$ . Thus, if  $K = 8$ , then  $d = 28\%$ .

different musical instruments. Timbre is often described in terms of sound texture or coloration.

In this experiment, we investigated differences in timbre that are produced from electric guitar notes that were processed using different distortion effects. We discuss the relationship between these timbral differences and a temporal factor extracted from the ACF  $W_{\phi(0)}$ . As shown in Fig. 2.1, this factor  $W_{\phi(0)}$  reflects the relative width of the ACF peak at its zero-lag origin.  $W_{\phi(0)}$  is defined by the first delay time  $\phi(\tau)$  at which the normalized ACF declines to half its maximal value (i.e., 0.5). It is worth noting that this factor  $W_{\phi(0)}$  in the monaural autocorrelation function (ACF) is analogous to factor  $W_{IACC}$  in the interaural correlation function (IACF).

An electric guitar with “distortion” is a primary instrument of pop and rock music. Previously, Marui and Wartens (2005) investigated timbral differences using of three types of nonlinear distortion processors with differing levels of Zwicker Sharpness (Zwicker and Fastl, 1999). In this study, we examined whether timbre can be described in terms of temporal factors extracted from the running ACF of the source signal. We wanted to determine whether one can distinguish notes that are played with different degrees of distortion despite their identical pitch, loudness, and duration.

### 6.6.1 Experiment 1 – Peak Clipping

The purpose of this experiment is to find the ACF correlate of distortion. We changed the strength of distortion by the use of a computer. The distortion of music signal  $p(t)$  was processed by a computer program that peak clipped the signal to keep it within a given cutoff amplitude range ( $\pm C$ ) and below a corresponding cutoff sound pressure level (CL). The signal was hard-limited in amplitude such that for  $|p(t)| \leq C$

$$p(t) = p(t) \quad (6.13)$$

and for  $|p(t)| > C$

$$p(t) = +C, p(t) \geq C; p(t) = -C, p(t) \leq -C \quad (6.13b)$$

where  $C$  is the cutoff pressure amplitude, and its cutoff level CL is defined by

$$CL = 20 \log_{10} (C / |p(t)|_{\max}) \quad (6.14)$$

with  $|p(t)|_{\max}$  being the maximum amplitude of the signal.

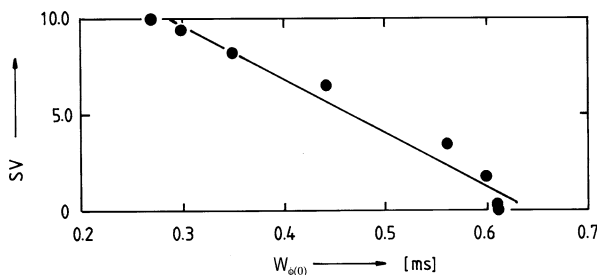
The value of cutoff level CL relative to the unclipped sound pressure level was varied from 0 to  $-49$  dB in 7 dB steps, yielding a set of eight test stimuli. As indicated in Table 6.3, pitch, signal duration, and listening level were fixed. The subjects were 19 students (male and female, all 20 years of age). Subjects listened to three stimuli and judged timbral dissimilarity. The number of stimulus combina-

**Table 6.3** Conditions of Experiments 1 and 2

Condition	Experiment 1	Experiment 2
(1) Conditions fixed		
Note (pitch)	A4 (220 Hz) By use of third string and second fret	A4 (220 Hz) By use of third string and second fret
Listening level in $L_{AE}$ (dB)	80	70
Signal duration (s)	4.0	1.5
(2) Conditions varied		
CL (dB) by Equation (6.12)	Eight signals tested changing the cutoff level for 0–49 dB (7 dB step)	
Distortion type	–	Three different types: VINT, CRUNCH, and HARD
Drive level	–	Three levels due to the strength of distortion: 50, 70, 90 by the effector Type ME-30 (Boss, Roland, Hamamatsu, Japan)

tions in this experiment was  ${}_8C_3 = 56$  triads. The dissimilarity matrix was constructed according to the dissimilarity judgments. The value 2 was assigned to the most different pair, 1 to the neutral pair, and 0 to the most similar pair. After multi-dimensional scaling analysis, we obtained the scale value (SV). This value is different from the scale value obtained by the method of comparative judgment (PCT).

We analyzed contributions to the scale value SV of other factors, for example, the mean value of  $W_{\phi(0)}$ , the decay rate of SPL (dBA/s), and the mean value of  $\phi_1$  (pitch strength). It was found that the most significant factor contributing to the SV 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 representative. The scale value of perceived timbral dissimilarity as a function of the mean value of  $W_{\phi(0)}$  is shown in Fig. 6.24. The correlation between the SV and the value of  $W_{\phi(0)}$  is 0.98 ( $p < 0.01$ ).

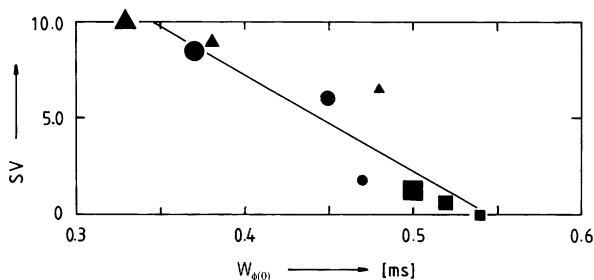


**Fig. 6.24** Results of regression analysis for SV and the mean value of  $W_{\phi(0)}$  (Experiment 1)

### 6.6.2 Experiment 2 – Commercial Effects Box

The purpose and methodology of this experiment were similar to the last, except that we used a commercial guitar effects unit rather than numerical clipping. As indicated in Table 6.3, we used a multi-effect box Type ME-30 (BOSS/Roland Corporation, Los Angeles) to produce a set of nine stimuli that had three kinds of effect types (VINT, CRUNCH and HARD) and three drive levels related to the degree of distortion: 50, 70, and 90. The subjects were 20 students who listened to three stimuli at a time to judge their dissimilarity. Thus, the number of stimulus combinations of this experiment was  ${}_9C_3 = 84$  triads. Results achieved were similar to Experiment 1, as shown in Fig. 6.25. The correlation between the scale value for timbral dissimilarity SV and the value of  $W_{\phi(0)}$  is 0.92 ( $p < 0.01$ ).

**Fig. 6.25** Results of regression analysis for SV and the mean value of  $W_{\phi(0)}$  (Experiment 2)



### 6.6.3 Concluding Remarks

In these two experiments we found that the main ACF correlate of timbral contrast due to distortion is  $W_{\phi(0)}$ . This factor is deeply related to the frequency composition of the source signal. In this context, it is interesting that Ohgushi (1980) showed that the lowest and highest frequency components present are the primary determinants of timbre.

# Chapter 7

## Spatial Sensations of Binaural Signals

Spatial sensations include the apparent location, apparent source width, and subjective diffuseness (envelopment) of sounds. These are described by the multiple spatial factors extracted from the IACF for the signal arriving at the two ear entrances. Neuronal correlates of spatial factors (LL, IACC and  $\tau_{IACC}$ ) observed in SVR, EEG and MEG signals (Section 5.6) were predominantly associated with the right hemisphere.

### 7.1 Sound Localization

#### 7.1.1 Cues of Localization in the Horizontal Plane

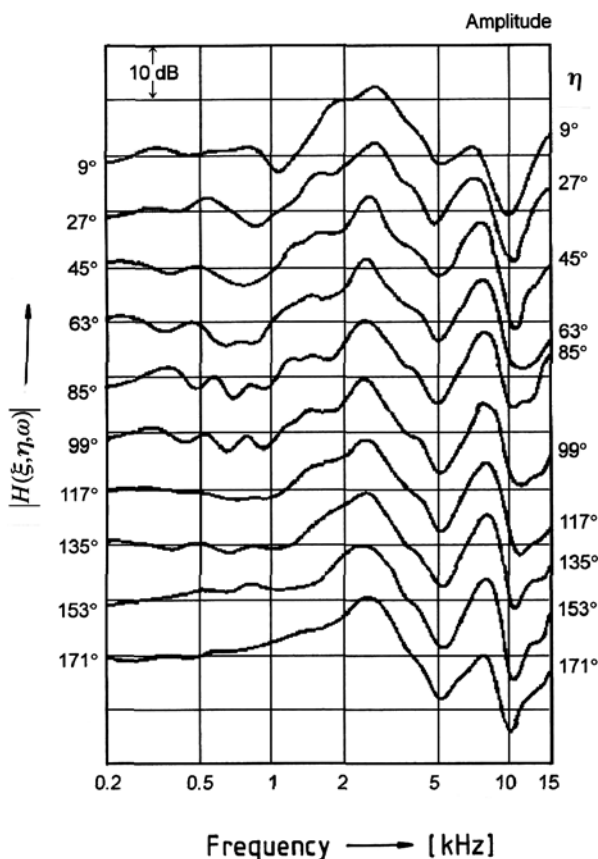
The perceived direction of a sound source in the horizontal plane can be expressed in terms of the spatial factors extracted from the IACF, such that

$$L_{\text{Horizontal}} = S_R = f_R(\Phi_{\text{ll}}(0), \Phi_{\text{rr}}(0), \text{IACC}, \tau_{\text{IACC}}, W_{\text{IACC}}) \quad (7.1)$$

where  $\Phi_{\text{ll}}(0)$  and  $\Phi_{\text{rr}}(0)$  signify sound energies of the signals arriving at the left- and right-ear entrances. Interaural delay time,  $\tau_{\text{IACC}}$ , is the most significant factor for horizontal localization among the five spatial factors in Equation (7.1). A second major factor, is the interaural level difference, which is a function of  $\Phi_{\text{ll}}(0)$  and  $\Phi_{\text{rr}}(0)$  based on the difference between the energies at the two ears. A well-defined, clear direction is perceived when the normalized IACF has one sharp maximum with a large value of IACC, and with a narrow (small) value of  $W_{\text{IACC}}$ . These conditions are typically produced by high frequency components above 2 kHz. On the other hand, subjective diffuseness ranging from the lack of a clear direction to the absence of any impression of spatial direction corresponds to a low value of IACC, e.g.  $\text{IACC} < 0.15$  (Damaske and Ando, 1972) and a wide set of associated delay times  $W_{\text{IACC}}$ . These conditions are typically produced by low frequency components.

### 7.1.2 Cues of Localization in the Median Plane

Apart from these five spatial factors of the localization in the horizontal plane given by Equation (7.1), of particular interest is the localization in the median plane. In this plane, these spatial factors from the IACF are not significantly changed as a function of elevation (i.e., due to almost symmetric shape of the head and pinnae), so that  $IACC$  is almost in unity,  $\tau_{IACC}$  is zero, and  $\Phi_{II}(0) = \Phi_{II}(0) = \text{constant}$ , and  $W_{IACC} = \text{constant}$ , which depend only on the spectrum of the source signal. It has been believed for a long time that the cues must be found in the spectrum of the sound signals. However, it is hard to find distinct cues in the head-related transfer function (HRTF) in the spectrum as shown in Fig. 7.1 (Mehrgardt and Mellert, 1977). In fact, the temporal factors extracted from the early delay range of the ACF of a sound signal arriving at the ear entrances might act as cues (Sato et al., 2001).



**Fig. 7.1** Measured head related transfer function (HRTF) for the different incident angles  $\eta$  in the median plane



We shall clearly show differences in the three temporal factors,  $\tau_e$ ,  $\tau_1$ , and  $\phi_1$ , extracted from the ACF as a function of the incident angle of elevation. According to the model described in Section 5.1, localization in the median plane  $L_{\text{median}}$  is grouped with monaural, temporal sensations because it appears to be based on factors extracted from the autocorrelation function:

$$L_{\text{median}} = S_L = f_L(t_e, t_1, f_1) \quad (7.2)$$

Like other perceptual attributes based on ACF factors, we expect sound localization in the median plane to exhibit “left hemisphere specialization.”

The amplitudes of the transfer functions for sound incident from the median plane to the ear entrances as measured by Mehrgardt and Mellert (1977) were transformed into equivalent autocorrelation functions (ACFs). The following steps obtain the ACFs from these transfer functions:

1. Data sets were obtained from the figures from the Mehrgardt and Mellert paper using an optical image reader (scanner), with 300 data points each.
2. Amplitude as a function of frequency in the logarithmic scale was obtained.
3. Each amplitude in decibel scale was converted to a corresponding real number, and the ACF was calculated by inverse Fourier transform after passage through an A-weighted filter.

Examples of the NACF are shown in Fig. 7.2. There is a certain degree of correlation between both  $\tau_n$  and  $\tau_{n+i}$ ,  $\phi_n$  and  $\phi_{n+i}$ , where  $\tau_n$  and  $\phi_n$  are the delay time and amplitude, respectively, of the  $n$ -th peak of the NACF. Thus,  $\tau_1$  and  $\phi_1$  can be representatives for sets of  $\tau_n$  and  $\phi_n$ . Examples of plotting the amplitude of the ACF on a logarithmic scale as a function of the delay time are shown in Fig. 7.3. A straight line can fit the envelope of the decay of the ACF in the logarithmic scale, and  $\tau_e$  was easily obtained from the delay at which the envelope drops below  $-10$  dB. The value  $\Phi(0)$  is not considered as a cue for sound localization here. Three factors extracted from the NACF are shown in Fig. 7.4. The value of  $\tau_1$  for incident angles from  $0^\circ$  to  $45^\circ$  is almost the same, but  $\tau_e$  for incident angle of  $45^\circ$  ( $\tau_e = 3.1$  ms) is much larger than that at  $0^\circ$  ( $\tau_e = 2.1$  ms). The value of  $\tau_1$  for incident angle of  $180^\circ$  is different from those for the above two angles, however, the  $\phi_1$  is relatively small.

Obviously, the angle in the median plane can be distinguished by the three monaural temporal factors  $\tau_1$ ,  $\phi_1$ , and  $\tau_e$ . These factors, therefore, may play an important role in the perception of localization in the median plane.

## 7.2 Apparent Source Width (ASW)

We shall show that apparent source width (ASW) may be described in terms of factors extracted from the IACF.

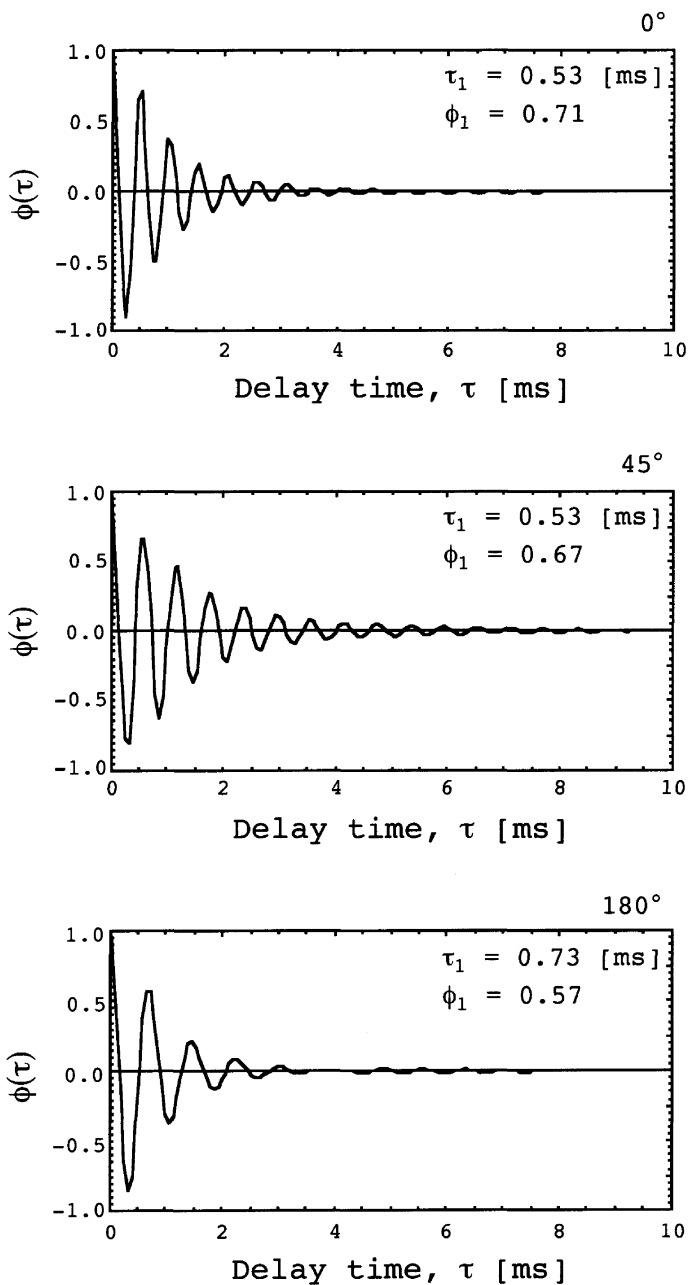


Fig. 7.2 Examples of the NACF of different incident angles in the median plane, 0°, 45°, and 180°

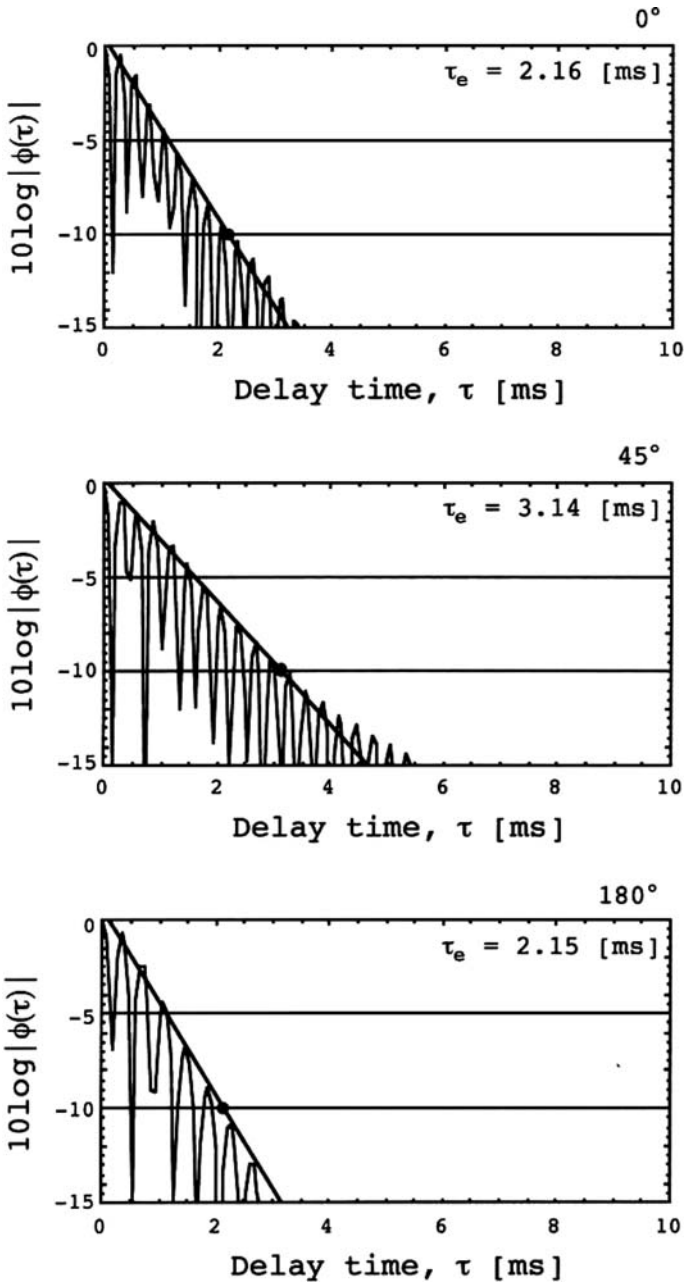
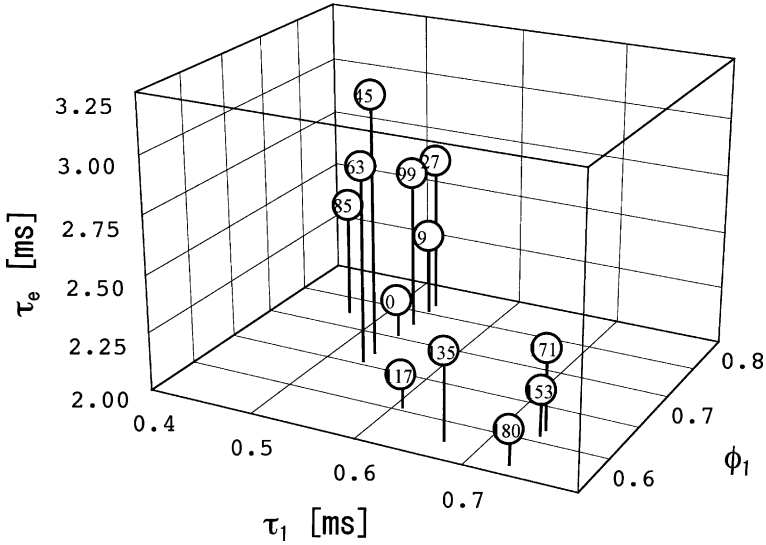


Fig. 7.3 Examples of the effective duration extracted from the NACF envelope (in logarithm) for different incident angles in the median plane, 0°, 45°, and 180°



**Fig. 7.4** Three-dimensional illustration plotted for three factors,  $\phi_1$ ,  $\tau_1$ , and  $\tau_e$ , which may distinguish different incident angles in the median plane. Numbers in the circles indicate incident angle in degrees

### 7.2.1 Apparent Width of Bandpass Noise

For sound fields passing sounds with predominately low frequencies, the interaural correlation function IACF has no sharp peaks in the interaural delay range of less than 1 ms, and the width of the major IACF peak,  $W_{\text{IACC}}$ , becomes “wider.” The  $W_{\text{IACC}}$  of band-pass noise may calculated theoretically using the following equation (Ando, 1998),

$$W_{\text{IACC}}^{(\delta)} \approx \frac{4}{\Delta\omega} \cos^{-1} \left( 1 - \frac{\delta}{\text{IACC}} \right) \quad (7.3)$$

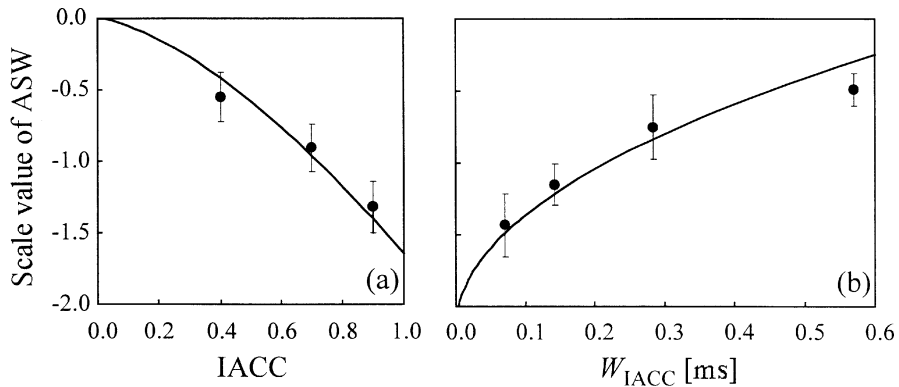
where  $\Delta\omega_c = 2\pi(f_1 + f_2)$ , and  $f_1$  and  $f_2$  are the lower and upper frequencies of an ideal filter. For the sake of simplicity,  $\delta$  is defined as 10% of the maximal value of the interaural correlation function IACF, i.e.  $\delta = 0.1(\text{IACC})$ .

The scale value of the ASW was obtained by the PCT with 10 subjects (Sato and Ando, 1996). In order to control the value of  $W_{\text{IACC}}$ , the center frequency of the 1/3-octave band-pass noise was changed as 250, 500, 1, and 2 kHz. The value of IACC was adjusted by controlling the sound pressure ratio between reflections ( $\xi = \pm 54^\circ$ ) and the direct sound ( $\xi = 0^\circ$ ). To avoid effects of the listening level on the ASW (Keet, 1968), the total SPL 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  $W_{IACC}$  contribute to  $S_R$  independently ( $p < 0.01$ ), so that

$$S = S_R = f_R(IACC) + f_R(W_{IACC}) \approx \alpha(IACC)^{3/2} + \beta(W_{IACC})^{1/2} \quad (7.4)$$

where coefficients  $\alpha \approx -1.64$  and  $\beta \approx 2.44$  are obtained by regressions of the scale values with 10 subjects as shown in Fig. 7.5. This holds under the condition of  $\tau_{IACC} = 0$ . Obviously, as shown in Fig. 7.6, the calculated scale values by Equation (7.4) and measured scale values are in good agreement ( $r = 0.97$ ,  $p < 0.01$ ).

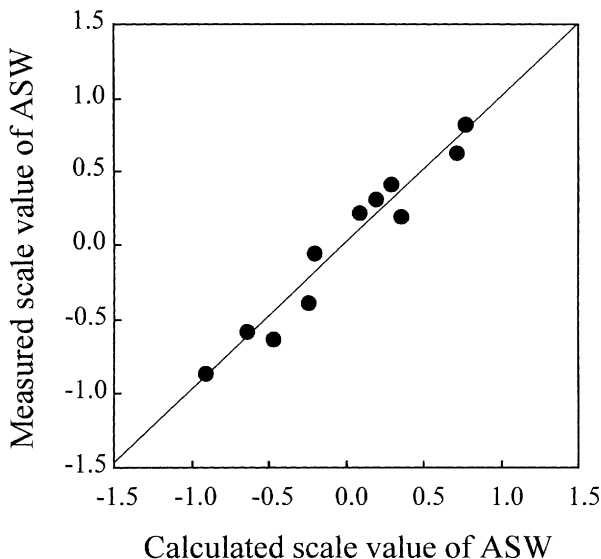


**Fig. 7.5** Scale values of apparent source width (ASW) for the 1/3-octave band-pass noises with 95% reliability as a function of (a) the IACC and (b) the  $W_{IACC}$ . The regression curves are expressed by Equation (7.4) with  $\alpha = 1.64$  and  $\beta = 2.44$

For each individual listener, the scale value can be calculated by Equation (7.4) in a similar manner. Coefficients  $\alpha$  and  $\beta$  in the equation for each listener are obtained by the multiple regression analysis and are indicated in Table 7.1. Figure 7.7 shows the relationship between the measured scale values and the calculated scale values with the constants for each of 10 subjects. The different symbols indicate the scale values of different subjects. The correlation coefficient between the measured and calculated  $S(ASW)$  is 0.91 ( $p < 0.01$ ).

### 7.2.2 Apparent Width of Multiband Noise

Keet (1968) showed that the apparent source width ASW of sounds depends on the amplitude of coherence between signals fed to the two loudspeakers and their listening level LL. On the basis of our central auditory signal processing model, however, source width may be described in terms of spatial factors extracted from the interaural correlation function IACF (Sato and Ando, 2002). It is assumed that this depends most directly on  $W_{IACC}$ , rather than on the spectrum of the source signal per se, and also as well on the interaural correlation magnitude IACC and



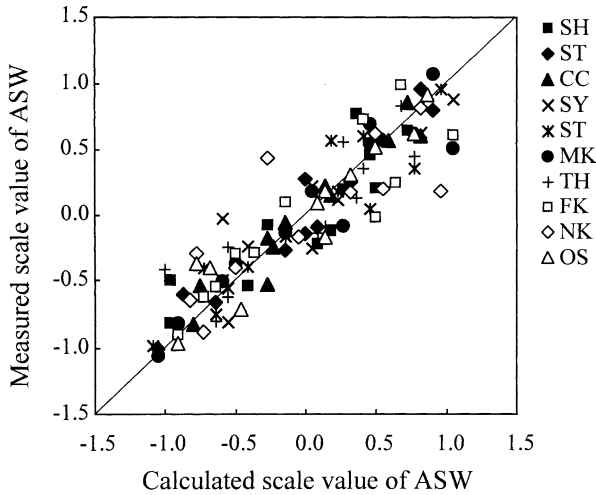
**Fig. 7.6** Relationship between the measured scale values of apparent source width (ASW) and the scale values of ASW calculated by Equation (7.4) with  $\alpha = 1.64$  and  $\beta = 2.44$ . Correlation coefficient  $r = 0.97$  ( $p < 0.01$ )

**Table 7.1** Coefficients  $\alpha$  and  $\beta$  in Equation (7.4) for calculating ASW of each individual and the correlation coefficient between the measured and calculated scale values of the ASW

Individual	$\alpha$	$\beta$	Correlation coefficient
S.H.	-1.21	2.58	0.88
T.S.	-1.50	3.18	0.97
C.C.	-1.05	2.82	0.97
S.Y.	-0.94	2.92	0.91
M.K.	-2.21	2.09	0.92
S.T.	-2.57	1.94	0.94
T.H.	-2.04	1.32	0.87
F.K.	-0.99	3.27	0.89
N.K.	-1.79	2.14	0.80
O.S.	-2.09	2.14	0.94
Average	-1.64	2.44	0.97

binaural listening level LL. It has been reported that even if the IACC is constant, the ASW increases as low-frequency components increase (Morimoto and Maekawa, 1988; Hidaka et al., 1995). The fact that wider source widths are perceived for sound sources with predominately low-frequency components might have a correlate in the behavior of  $W_{\text{IACC}}$ .

This study examines the apparent source widths of complex noise signals, which consist of several band-pass noises whose center passband frequencies are the harmonics of a fundamental frequency. The question here is whether pitches associated



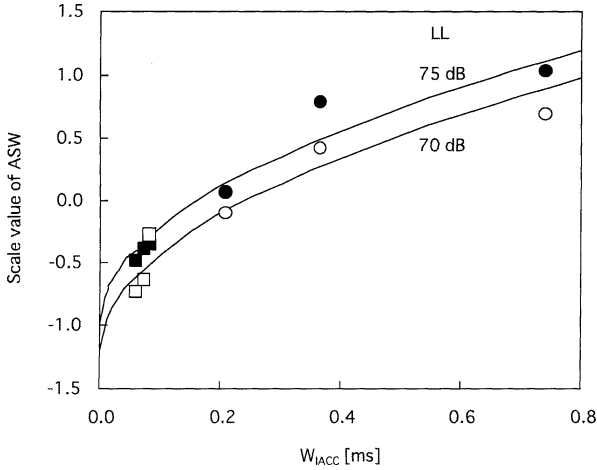
**Fig. 7.7** Relationship between the measured individual scale values of apparent source width (ASW) and the scale values of ASW calculated by Equation (7.4) with each individual value of  $\alpha$  and  $\beta$  as listed in Table 7.1,  $r = 0.90$  ( $p < 0.01$ ). Different symbols indicate data obtained by different individual subjects

with the fundamental frequencies of these noise complexes in some way influence or predict the apparent source widths ASWs of these sounds (Sato and Ando, 2002). Scale values of the ASW for these complex noise signals are compared with those of comparable noise signals with single passbands.

Single band-pass noises with center frequencies of 200, 400, and 800 Hz and multiband complex noises with center fundamental frequencies of 200, 400, and 800 Hz were used as source signals. Each complex noise stimulus consisted of three passbands, and the center frequencies of the lowest passbands were fixed at 1,600 Hz. When the fundamental frequency ( $F_0$ ) was 200 Hz, the three passbands were 1,600, 1,800, and 2,000 Hz. Similarly, for  $F_0 = 400$  Hz these were 1,600, 2,000, 2,400 Hz, and for  $F_0 = 800$  they were 1,600, 2,400, 3,200 Hz. The amplitudes of all passbands were adjusted to be the same by measuring  $\Phi(0)$ . Bandwidths of the noise passbands in all signals were 80 Hz with a cutoff slope of 2,068 dB/octave.

A single frontal loudspeaker for direct sound ( $\xi = 0^\circ$ ) and two symmetrical loudspeakers ( $\xi = \pm 54^\circ$ ) that added reflections were used simulate different source widths in an anechoic chamber (Section 2.3). To produce incoherent sound signals, time delays of reflections  $\Delta t_1$  and  $\Delta t_2$  were fixed at 20 and 40 ms, respectively. To reconfirm the effects of listening level LL on apparent source width ASW, the sound pressure level at the listener's head position was also changed from 70 to 75 dB. The values of the IACC of all sound fields were adjusted to  $0.90 \pm 0.01$  by controlling the identical amplitude of the reflections ( $A_1 = A_2$ ). The IACF was measured with two 1/2-in. condenser-type microphones, placed at the ear entrances of a dummy head. The output from the microphones was passed through an A-weighting network and was digitized at a sampling frequency of 44.1 kHz.

Paired-comparisons tests (PCT) using 12 sound fields ( $6 \times 2$ ) were performed on five subjects with normal hearing ability in order to obtain scale values for perceived source width ASW. Subjects were seated in an anechoic chamber and asked to judge which of the two paired stimuli they perceived to be wider. The duration of each sound stimulus was 3 s, the rise and fall times were 50 ms, and the silent interval between stimuli was 1 s. Each stimulus pair was separated by an interval of 4 s and the pairs were presented in random order. Twenty-five responses (five subjects  $\times$  five repeats) to each stimulus were obtained. Scale values of ASWs were obtained by applying the law of comparative judgment. The relationship between the scale values of source widths ASWs and  $W_{IACC}$ s of the source signals is shown in Fig. 7.8. There are significant differences between the scale values of apparent widths of noises with single and multiple passbands ( $p < 0.01$ ) even when they evoke the same pitch. However, the change in the factor  $W_{IACC}$  explains these differences. The results of the analysis of variance for scale values  $S$  for apparent source width ASW revealed that the explanatory factors  $W_{IACC}$  and LL are significant ( $p < 0.01$ ) and independent, so that

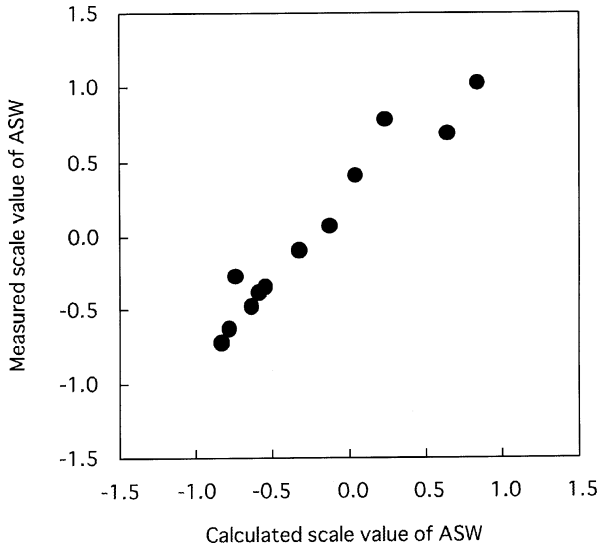


**Fig. 7.8** Average scale values of apparent source width (ASW) as a function of interaural correlation magnitude (IACC)  $W_{IACC}$  and as a parameter of listening level (LL). ●, band-pass noise, LL = 75 dB; ○, band-pass noise, LL = 70 dB; ■, complex noise, LL = 75 dB; □, complex noise; LL = 70 dB. The regression curve is expressed by Equation (7.5) with  $a = 2.40$  and  $b = 0.005$

$$S = S_R = f_R(W_{IACC}) + f_R(LL) \approx a(W_{IACC})^{1/2} + b(LL)^{3/2} \quad (7.5)$$

where  $a$  and  $b$  are coefficients. The powers,  $1/2$  and  $3/2$  for the terms of  $W_{IACC}$  and LL in Equation (7.5), respectively, were determined to obtain the best correlation between the scale values measured and its calculated value. The curves in Fig. 7.9 confirm the scale values calculated using Equation (7.5) with the coefficient values  $a \approx 2.40$  and  $b \approx 0.005$  that were obtained. The correlation coefficient between measured and calculated scale values was 0.97 ( $p < 0.01$ ).





**Fig. 7.9** Relationship between the measured scale values of ASW and the scale values of ASW calculated by Equation (7.5) with  $a = 2.40$  and  $b = 0.005$ . Correlation coefficient  $r = 0.97$  ( $p < 0.01$ )

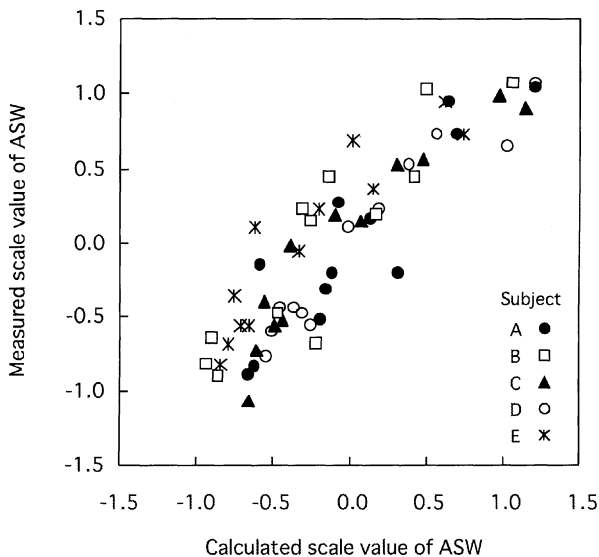
It is noteworthy that the above-mentioned scale value of ASW for 1/3-octave band-pass noise is expressed in terms of the 1/2 power of  $W_{IACC}$  and that the coefficient for  $W_{IACC}$  ( $\beta \approx 2.44$ ) is close to that of this study. The factor  $W_{IACC}$  is determined by the frequency component of the source signal, thus the pitch or the fundamental frequency represented by the temporal factor  $\tau_1$  is not necessary to describe or predict apparent source width. Results of ANOVA for scale values of ASW indicated that the explanatory factor LL is also significant ( $p < 0.01$ ). The scale values of ASW increase with an increase in binaural listening level LL, similar to Keet (1968).

Individual differences may be also expressed by the weighting coefficients in Equation (7.5), which are obtained by multiple regression analysis. For each individual, Fig. 7.10 shows the relationship between measured scale values of ASW and those calculated from Equation (7.5) using coefficients  $a$  and  $b$  (Table 7.2). The different symbols correspond to the different subjects. The correlation coefficient between the measured and the calculated scale values is 0.90 ( $p < 0.01$ ).

Because the weighting coefficients of  $(W_{IACC})^{1/2}$  in Equations (7.4) and (7.5) are apparently similar, one can construct a common formula

$$S = S_R = f(IACC) + f(W_{IACC}) + f(LL) \approx \alpha(IACC)^{3/2} + \beta(W_{IACC})^{1/2} + \gamma(LL)^{3/2} \tag{7.6}$$

where  $\alpha \approx -1.64$ ,  $\beta \approx 2.42$ ,  $\gamma \approx 0.005$ . It is worth noting that units of scale value of subjective preference even in different subjects and using different source signals appeared to be almost constant (Section 3.3).



**Fig. 7.10** Relationship between the measured individual scale values of apparent source width (ASW) and the scale values of ASW calculated by Equation (7.5) for each individual. Correlation coefficient  $r = 0.90$  ( $p < 0.01$ ). Different symbols indicate data obtained by different individual subjects

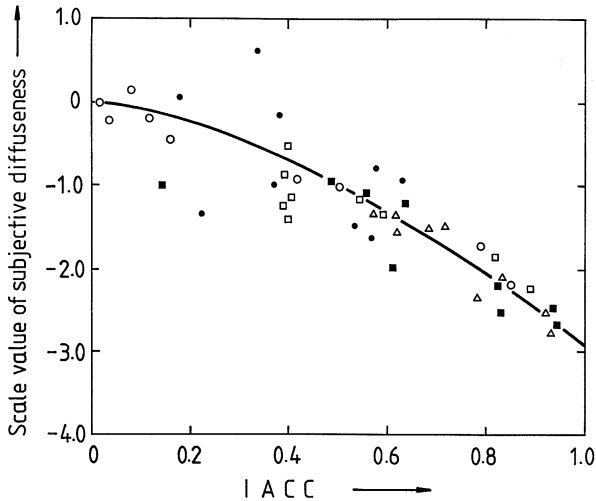
**Table 7.2** Coefficients  $a$  and  $b$  in Equation (7.5) for estimating the apparent source width (ASW) for each individual listener in Fig. 7.10 and the correlation coefficients between measured and estimated ASWs

Individual	$a$	$b$	Correlation coefficient
A	2.2	0.008	0.90
B	2.2	0.010	0.90
C	2.6	0.003	0.94
D	2.6	0.003	0.96
E	2.3	0.002	0.92
Average	2.4	0.005	0.97

### 7.3 Subjective Diffuseness

Scale values for the subjective diffuseness of sounds are described by the representative spatial factor, the interaural crosscorrelation magnitude IACC.

In order to obtain scale values for subjective diffuseness, paired comparisons were conducted using 1/3-octave band-pass Gaussian noise and by varying the horizontal angle of two symmetric reflections (Ando and Kurihara, 1986; Singh et al., 1994). Listeners judged which of two sound fields were perceived to be more diffuse. A remarkable finding is that scale values  $S$  of subjective diffuseness are



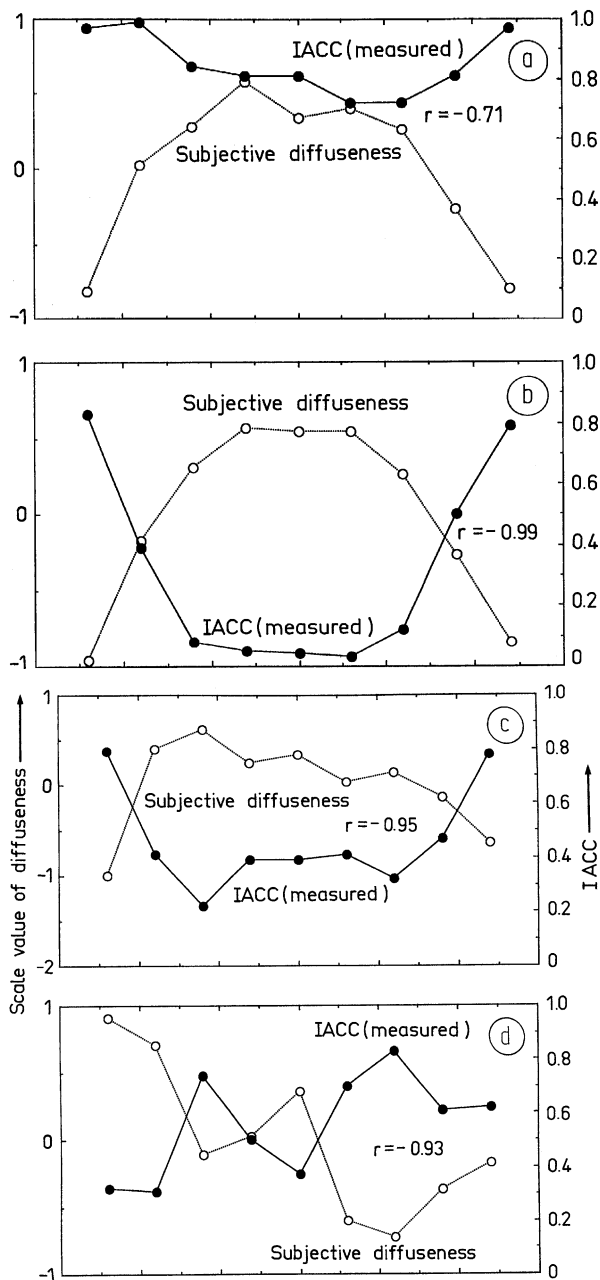
**Fig. 7.11** Scale values of subjective diffuseness as a function of the IACC (calculated). Different symbols indicate different frequencies of the 1/3-octave band-pass noise:  $\Delta$ , 250 Hz;  $\circ$ , 500 Hz;  $\square$ , 1 kHz;  $\bullet$ , 2 kHz;  $\blacksquare$ , 4 kHz. (—): Regression line by Equation (7.6)

inversely proportional to interaural correlation magnitude IACC and may therefore be reformulated in terms of the 3/2 power of the IACC in a manner similar to that for other subjective preference values (see Section 3.1.4), i.e.,

$$S = S_R \approx -\alpha(\text{IACC})^\beta \tag{7.7}$$

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

The results of scale values obtained through paired comparisons together with values calculated using Equation (7.7) are shown as a function of the IACC in Fig. 7.11. There is great variation in the data in the range of the IACC < 0.5, however, no essential difference may be found in the results for different frequencies between 250 Hz and 4 kHz. The scale values of subjective diffuseness, which depend on horizontal angle, are shown in Fig. 7.12, for 1/3-octave band-pass noises with the center frequencies of 250 Hz, 500 Hz, 1 kHz, 2 kHz, and 4 kHz. The scale values for each individual listener are shown in Fig. 7.13. Clearly, the most effective horizontal angles of reflections depend on the frequency range (Fig. 7.14). These are about  $\pm 90^\circ$  for the low-frequency range of less than 500 Hz, around  $\pm 55^\circ$  for the 1 kHz range (the most important angle for music), and smaller than  $18^\circ$  for the 2 and 4 kHz bands. Such directional reflections for each frequency range can be controlled by using a fractal structure for the wall surface, for an example see (Ando, 1998).



**Fig. 7.12** Scale values of subjective diffuseness and the IACC as a function of the horizontal angle of incidence to a listener, with 1/3-octave-band noise of center frequencies. (a) 250 Hz. (b) 500 Hz. (c) 1 kHz. (d) 2 kHz. (e) 4 kHz

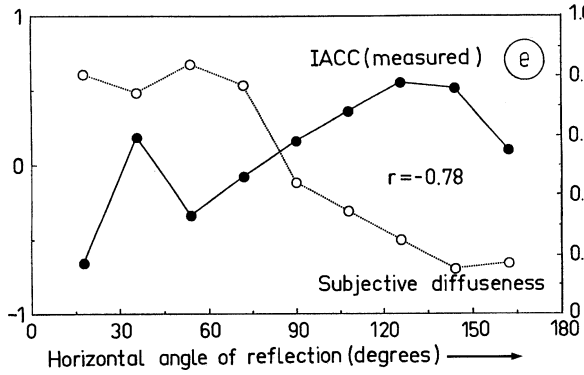


Fig. 7.12 (continued)

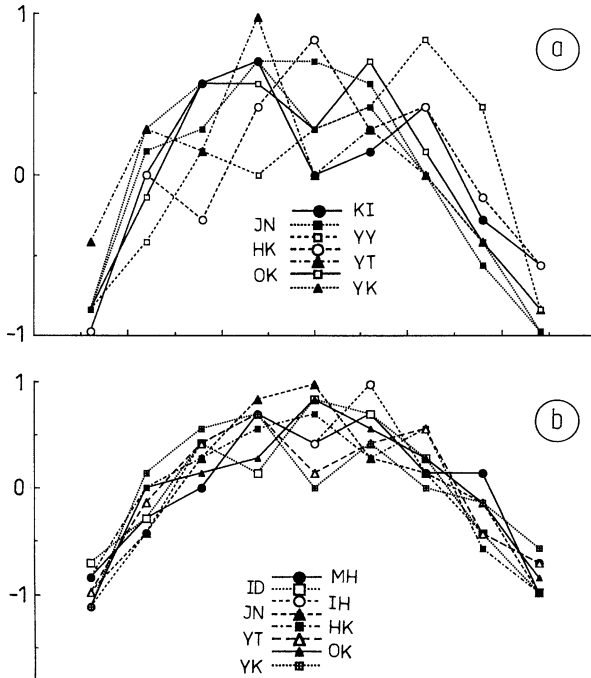


Fig. 7.13 Scale values of subjective diffuseness for each individual as a function of the horizontal angle of incidence to a listener, with 1/3-octave-band noise of center frequencies. (a) 250 Hz. (b) 500 Hz. (c) 1 kHz. (d) 2 kHz. (e) 4 kHz. Different symbols indicate data obtained by different individual subjects with their initials

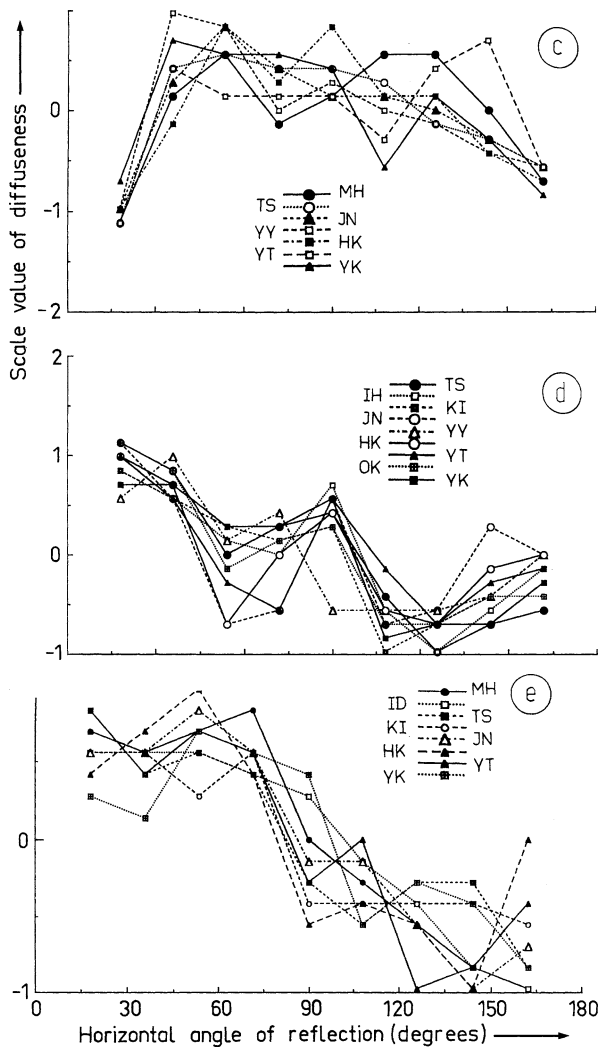
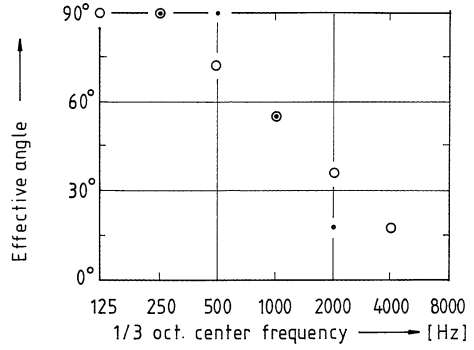


Fig. 7.13 (continued)

**Fig. 7.14** The optimal horizontal angles of reflections to a listener for each frequency range for the purpose of decreasing the IACC and thus increasing subjective diffuseness. ○: Angles obtained by the calculated IACC; ∅: angles obtained by the observed IACC



# Chapter 8

## Applications (I) – Music and Concert Hall Acoustics

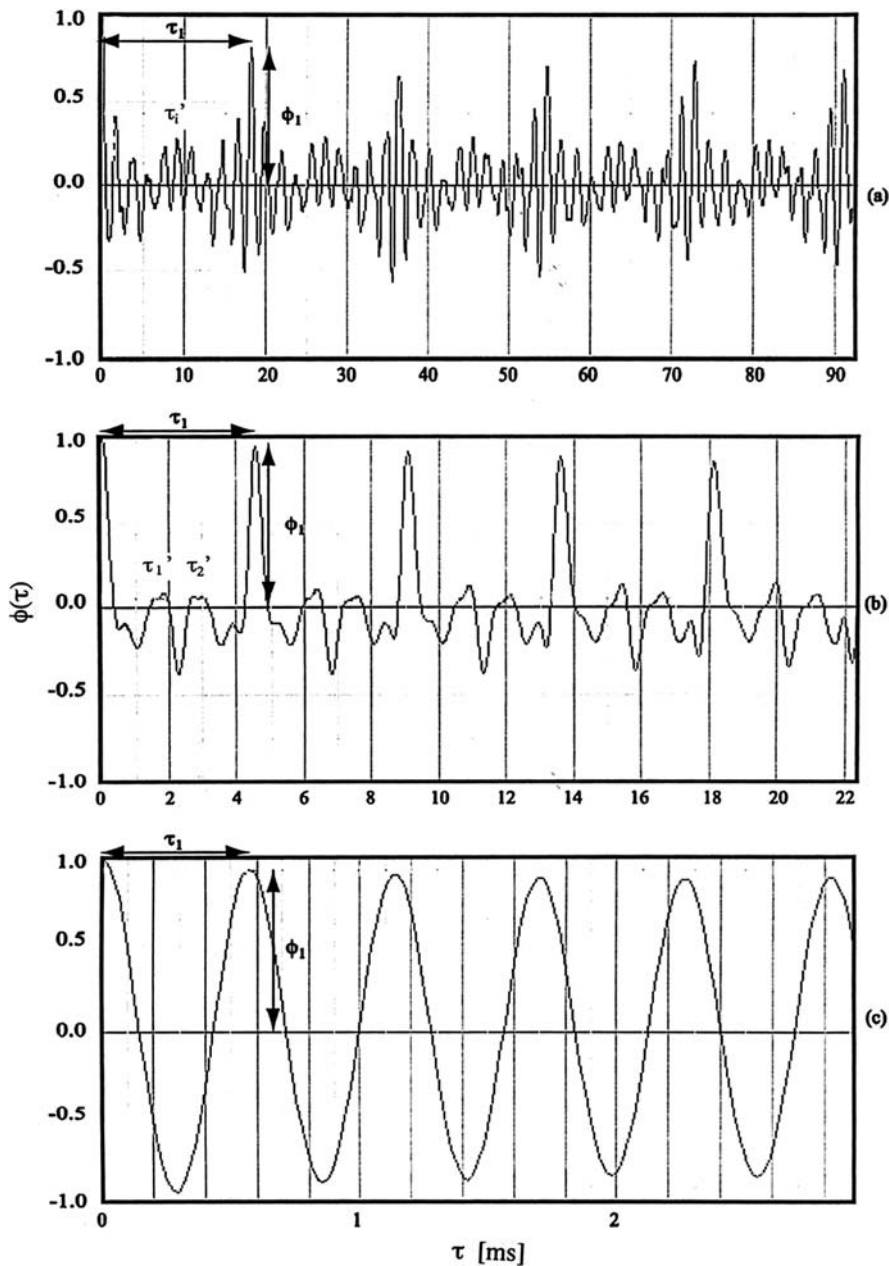
This chapter offers some applications of the central auditory signal processing model for temporal and spatial sensations as well as for subjective preferences. The measurement of pitches of notes sounded by a piano are discussed in Section 8.1. Examples of adaptive acoustic design of a public concert hall using global listener preference data are presented in Section 8.2. A seat selection system for enhancing individual listening experiences in a concert hall is discussed in Section 8.3. The preferred temporal conditions for music performance by cellists are discussed as it relates to acoustic design of the stage in Section 8.4.

### 8.1 Pitches of Piano Notes

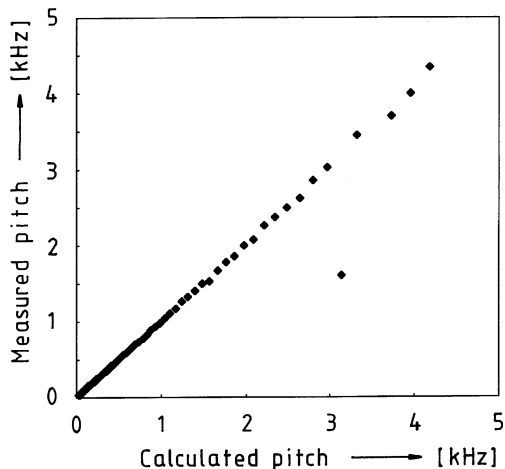
It is well known that the source signal of a piano is a complex tone with mostly low-frequency harmonics. According to the method described in Section 6.2, source signals of pianos were analyzed and compared with those calculated using ratios of neighboring notes in an equally tempered chromatic scale, i.e., semitone steps of  $2^{1/12}$ .

For this study, source signals were picked up by a single microphone placed at the center position above a grand piano with its top lid opened at the usual angle for performance (Inoue and Ando, unpublished). The piano was equipped with an automatic performance system, was tuned before measurement, and produced source signals that could be reliably reproduced. Examples of the ACF analyzed for notes A1 (55 Hz), A3 (220 Hz), and A6 (1760 Hz) are shown in Fig. 8.1. It is clear that the delay times of the first peak  $\tau_1$  extracted from the ACF correspond well to the measured fundamental frequencies of notes A1 (55.2 Hz), A3 (219.7 Hz), and A6 (1785.7 Hz). Its amplitude  $\phi_1$  is large enough (more than 0.8) and in this condition, a clear pitch is perceived. Calculated and measured pitches for all of 88 notes are shown in Figs. 8.1 and 8.2, and these values are listed in Table 8.1.





**Fig. 8.1** Examples of the ACF analyzed for source signals from a piano. (a) Note A1 of the pitch of 55 Hz. (b) Note A3 of the pitch of 220 Hz. (c) Note A6 of the pitch of 1760 Hz. The missing fundamental phenomena may be observed by  $\tau_1$  extracted from the ACF, which corresponds to 55 Hz and 220 Hz. However, the pitch of 1760 Hz is observed by  $\tau_1$  at just the fundamental frequency



**Fig. 8.2** Relationship between calculated and measured pitches obtained by the value of  $\tau_1$  extracted from the ACF of source signals from a “tuned” piano

Most of the measured values are in good agreement with calculated ones. It is interesting, however, that measured pitches below G3 (207.6 Hz) were a little bit higher than calculated ones. Below this pitch, the amplitudes of the fundamental frequency component in the spectrum analyzed were small and not significant, and thus the low pitch that one hears from these piano notes is primarily a missing fundamental phenomenon. For pitches below 55 Hz, no appreciable energy at the fundamental frequency in the measured spectrum was observed.

It is worth noting that the upper frequency limit of the ACF model for the pitches of missing fundamentals is about 1,200 Hz (Section 6.2.3). Above the frequency of this pitch, discrepancies between calculated and measured pitches grew large, reaching about 2%. At G7 (3135.9 Hz), in particular, a large discrepancy was observed, where a measured value of 1612.9 Hz corresponded roughly to half the calculated one. The octave error could conceivably have been caused by a mistake in tuning adjustment or, perhaps more likely, in the autocorrelation analysis (picking the second major peak rather than the first one). In some neural autocorrelation models (Cariani, 2004), this problem is largely avoided by analyzing the ACF for regular patterns of major interspike interval peaks rather than choosing the highest peak. The method uses a dense set of interval sieves that quantify the pattern strengths of all possible periodicities, which allows the model to estimate the relative strengths of multiple, competing pitches that may be heard in a given note or chord.

**Table 8.1** Calculated pitches and the values measured by the ACF ( $\tau_1$ ) of the 88-note signals for a piano that was said to be “tuned”

Note	Calculated pitch (Hz)	Measured pitch (Hz)	Difference (Hz)
A0	27.5	29.9	-2.4
B0	29.1	31.5	-2.4
H0	30.8	33.5	-2.7
C1	32.7	35.1	-2.4
Cis1	34.6	34.8	-0.2
D1	36.7	36.9	-0.2
Es1	38.8	39.2	-0.4
E1	41.2	41.4	-0.2
F1	43.6	43.6	0.0
Fis1	46.2	46.5	-0.3
G1	48.9	49.2	-0.3
Gis1	51.9	52.3	-0.4
A1	55.0	55.2	-0.2
B1	58.2	58.8	-0.6
H1	61.7	62.1	-0.4
C2	65.4	65.7	-0.3
Cis2	69.2	69.4	-0.2
D2	73.4	74.0	-0.6
Es2	77.7	78.1	-0.4
E2	82.4	82.6	-0.2
F2	87.3	87.7	-0.4
Fis2	92.4	92.5	-0.1
G2	97.9	98.0	-0.1
Gis2	103.8	104.1	-0.3
A2	110.0	111.1	-1.1
B2	116.5	117.6	-1.1
H2	123.4	124.2	-0.8
C3	130.8	131.5	-0.7
Cis3	138.5	138.8	-0.3
D3	146.8	149.2	-2.4
Es3	155.5	156.2	-0.7
E3	164.8	165.2	-0.4
F3	174.6	175.4	-0.8
Fis3	184.9	185.1	-0.2
G3	195.9	196.0	-0.1
Gis3	207.6	208.3	-0.7
A3	220.0	219.7	+0.3
B3	233.0	232.5	+0.5
H3	246.9	246.9	0.0
C4	261.6	263.1	-1.5
Cis4	277.1	277.7	-0.6
D4	293.6	294.1	-0.5
Es4	311.1	312.5	-1.4
E4	329.6	327.8	+1.8

**Table 8.1** (continued)

Note	Calculated pitch (Hz)	Measured pitch (Hz)	Difference (Hz)
F4	349.2	350.8	-1.6
Fis4	369.9	370.3	-0.4
G4	391.9	393.7	-1.8
Gis4	415.3	416.6	-1.3
A4	440.0	444.4	-4.4
B4	466.1	465.1	+1.0
H4	493.8	495.0	-1.2
C5	523.2	526.3	-3.1
Cis5	554.3	558.6	-4.3
D5	587.3	591.7	-4.4
Es5	622.2	625.0	-2.8
E5	659.2	666.6	-7.4
F5	698.4	704.2	-5.8
Fis5	739.9	740.7	-0.8
G5	783.9	787.4	-3.5
Gis5	830.6	826.4	+4.2
A5	880.0	892.8	-12.8
B5	932.3	943.3	-11.0
H5	987.7	980.3	+7.4
C6	1046.5	1041.6	+4.9
Cis6	1108.7	1111.1	-2.4
D6	1174.6	1176.4	-1.8
Es6	1244.5	1265.8	-21.3
E6	1318.5	1333.3	-14.8
F6	1396.9	1408.4	-11.5
Fis6	1479.9	1492.5	-12.6
G6	1567.9	1538.4	+29.5
Gis6	1661.2	1666.6	-5.4
A6	1760.0	1785.7	-25.7
B6	1864.6	1851.8	+12.8
H6	1975.5	2000.0	-24.5
C7	2093.0	2083.3	+9.7
Cis7	2217.4	2272.7	-55.3
D7	2349.3	2380.9	-31.6
Es7	2489.0	2500.0	-11.0
E7	2637.0	2631.5	+5.5
F7	2793.8	2857.1	-63.3
Fis7	2959.9	3030.3	-70.4
G7	3135.9	1612.9	+1523.0*
Gis7	3322.4	3448.2	-125.8
A7	3520.0	3703.7	-183.7
B7	3729.3	3703.7	+25.6
H7	3951.0	4000.0	-49.0
C8	4186.0	4347.8	-161.8

\*A little less than one octave mistuned.

## 8.2 Design Studies of Concert Halls as Public Spaces

Ideally, to optimize listening experiences, specific music programs should be performed in concert halls that have the most compatible acoustics. Inversely, concert halls should be designed and modified with particular kinds of sound programs in mind. Sound programs can be classified in terms of their minimum effective duration  $(\tau_e)_{\min}$ , and concert halls can be designed with dimensions and absorption coefficient of walls that enhance temporal and spatial sensations given particular effective durations. Using subjective preferences for the four orthogonal factors of the sound field that are common to most listeners, one can add individual factor preferences together to derive an overall estimated preference value for each seat (i.e., by the “principle of superposition” expressed by Equation (3.8)). Comparison of ensembles of seat preference values for different configurations for a concert hall allows a designer to choose the best scheme. In this section, we discuss the optimum space form for a hall in terms of two spatial factors, the IACC and LL.

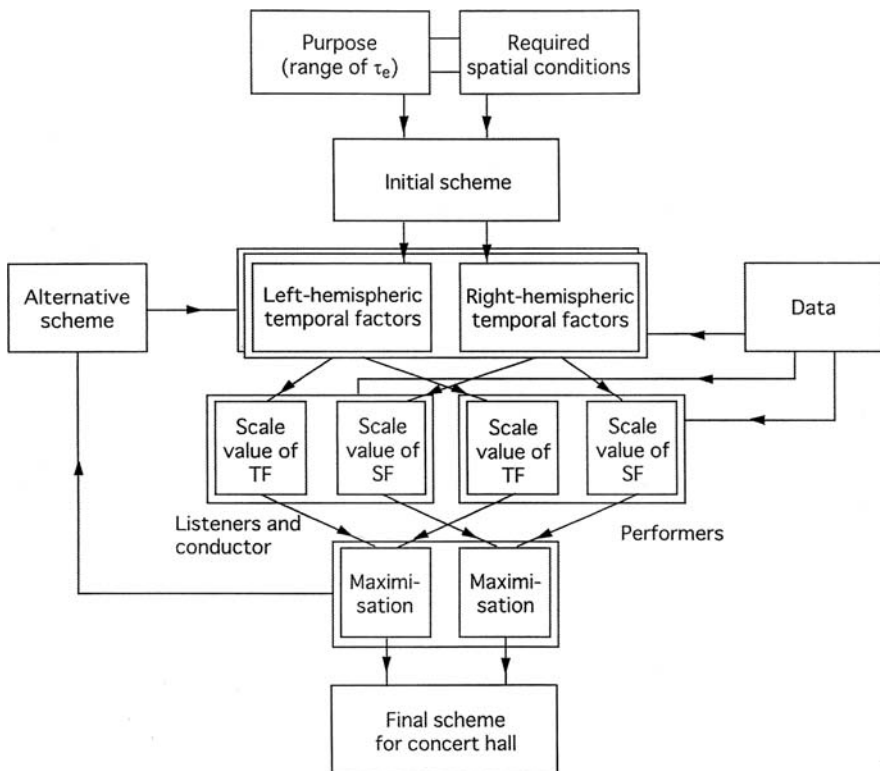
### 8.2.1 Genetic Algorithms (GAs) for Shape Optimization

The theory of subjective preference described in Section 3.3 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 (3.8). Genetic algorithms (GAs) 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 (see Table 5.1). Genetic algorithms (Holland, 1975), a form of evolutionary computing, have been applied to the design of concert halls (Sato et al., 2002, 2004). In architectural optimization using genetic algorithms (GAS), 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 acous-

tical 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.

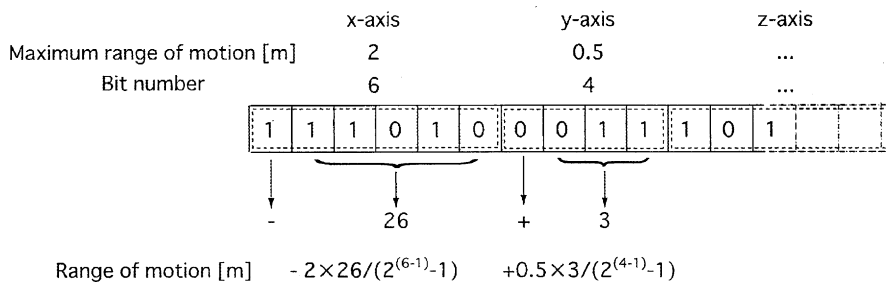
Here the GA system was used to generate alternative schemes for evolutionary selection (seen on the left-hand side of Fig. 8.3), and these alternative schemes can be evaluated in terms of the preference scale values in relation to the spatial factors associated with the right hemisphere that they would 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.



**Fig. 8.3** Procedure for designing the sound field in a concert hall that best satisfies the subjective preferences of listeners, conductor, and performers

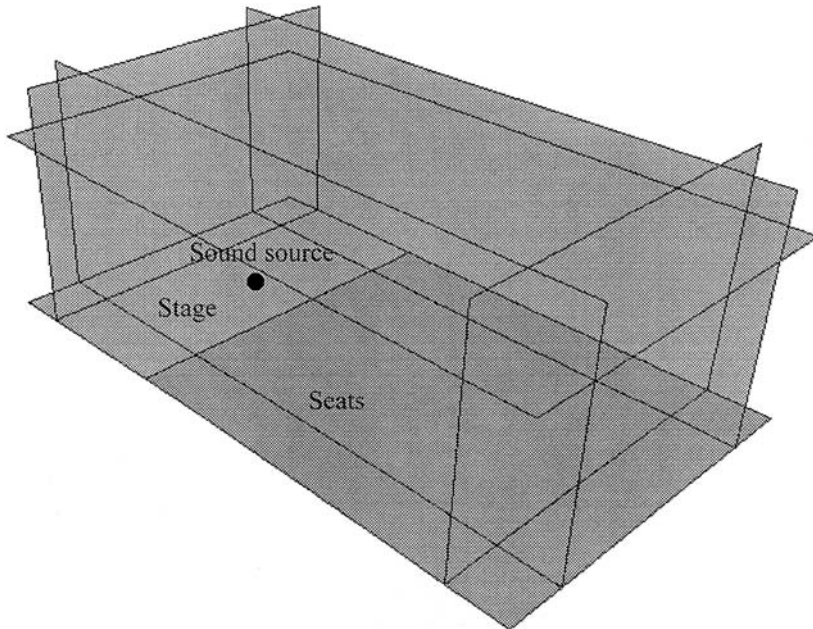
The behavior of the scale value in relation to each orthogonal factor  $S_i$  is given by Equations (3.8) through (3.11). Here, the parameters  $x_i$  and coefficients  $\alpha_i$  are listed in Table 3.4. In this calculation, linear scale values of subjective preference  $S_1$  and  $S_4$  are employed as fitness function due to the LL and IACC, respectively, because the geometric 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 omnidirectional source was assumed to be at the center of the stage, 1.5 m above the stage floor. The receiving points that correspond with 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 Fig. 8.4. 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 include 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.



**Fig. 8.4** An example of the binary strings used in encoding of the chromosome to represent modifications to the room shape

First, the proportions of the shoebox hall were optimized (model 1). The initial geometry is shown in Fig. 8.5. 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 of motion for each sidewall and the ceiling was  $\pm 5$  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



**Fig. 8.5** The initial scheme of the concert hall used as a base of model 1. The range of sidewall and the ceiling varied  $\pm 5$  m from the initial scheme

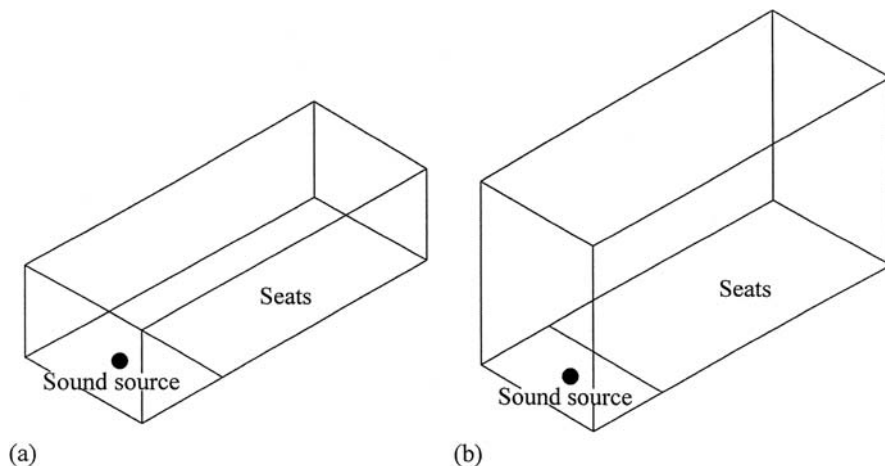
sidewalls were included in the averages ( $S_1$  and  $S_4$ ). In this calculation, the most preferred listening level,  $[LL]_p$ , was chosen at the front seat near the stage.

Results of optimization of the hall for  $S_1$  and  $S_4$  are shown in Fig. 8.6a and b, respectively. The width and length were almost the same in the two results, but the respective heights indicated were 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 (Fig. 8.6a). 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 (Fig. 8.6b).

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 (Fig. 8.6a). 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 (Fig. 8.6b). The reflection from the sidewalls were useful decreasing the IACC, however, this was not true of the reflections from the flat ceiling.

Next, to obtain even better sound fields, a slightly more complicated form (model 2, Fig. 8.7) was adopted as the starting point for floor plan optimization. The hall in





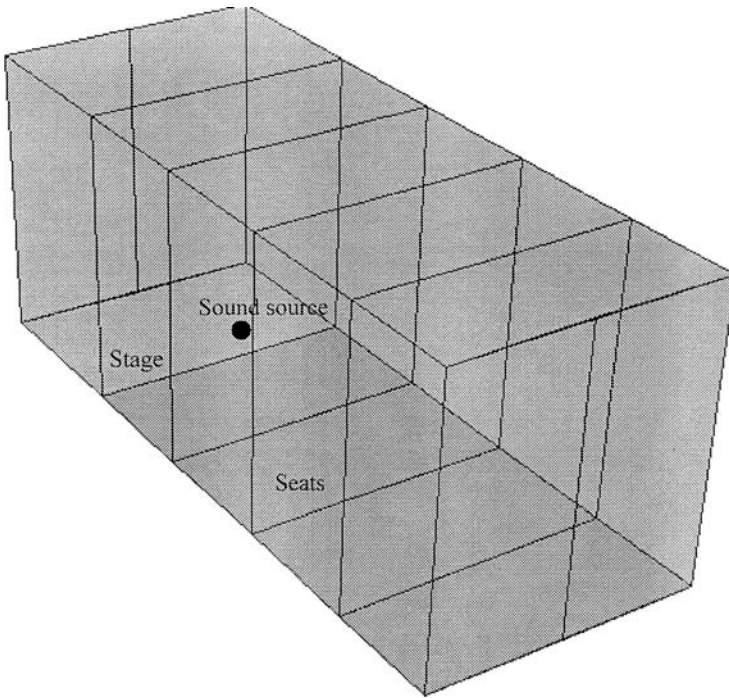
**Fig. 8.6** Results for model 1. (a) Geometry optimized for preferred listening level  $\overline{S}_1$ . (b) Geometry optimized for preferred diffuseness  $\overline{S}_4$  by minimizing the IACC

**Table 8.2** Comparison of proportions for the optimized forms by the LL ( $\overline{S}_1$ ) and the IACC ( $\overline{S}_4$ ), and those for the Grosser Musikvereinsaal

	Length/width	Height/width
Optimized for $\overline{S}_1$	2.50	0.71
Optimized for $\overline{S}_4$	2.57	1.43
Grosser Musikvereinsaal	2.55	0.93

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 distributed throughout the seating area on a  $2 \times 4$  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  $\pm 2$  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.

The result of optimizing the hall for  $S_1$  (the preference factor associated with listening level) is shown in Fig. 8.8 and contour lines of equal  $S_1$  values are shown in Fig. 8.9. 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  $S_4$  (the preference scale values associated with the IACC) is shown in Fig. 8.10, and contour lines of equal



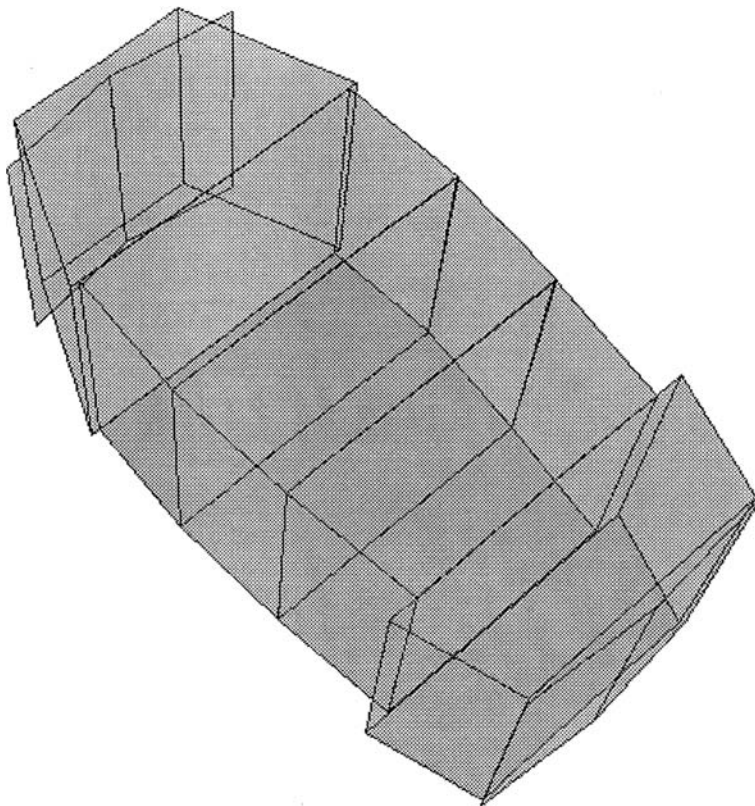
**Fig. 8.7** Initial scheme of the concert hall as a base of model 2. The rear wall of the stage and the rear wall of the audience area were divided into two. Sidewalls were divided into four

$S_4$  values are shown in Fig. 8.11. 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. Thus it was proved that a “leaf-shape” plan of concert hall is one of the most optimal schemes, and this basic plan was applied in the design of the Kirishima International Concert Hall, which opened in 1994.

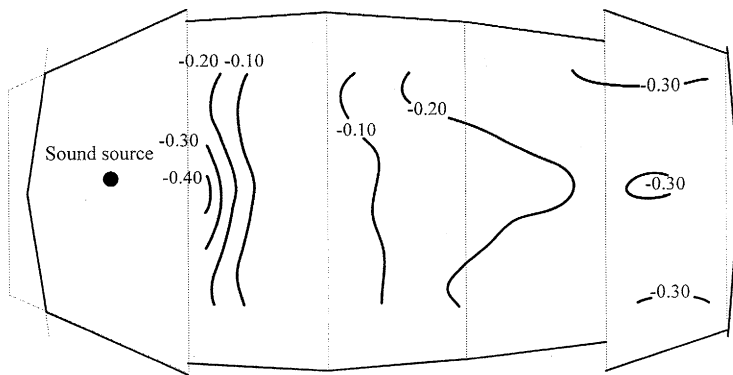
As for the conflicting requirements for  $S_1$  and  $S_4$ , the maximization of  $S_4$  (IACC) may take a higher priority than  $S_1$  (listening level LL) because the preference increases with a decreasing IACC for all subjects tested (Singh et al., 1994), while there are a large individual differences in the preferred listening level LL (Sakai et al., 1997). Thus, listeners can choose the seat with respect to their preferred level LL when they buy tickets for concert (Section 8.3).

### **8.2.2 Two Actual Designs: Kirishima and Tsuyama**

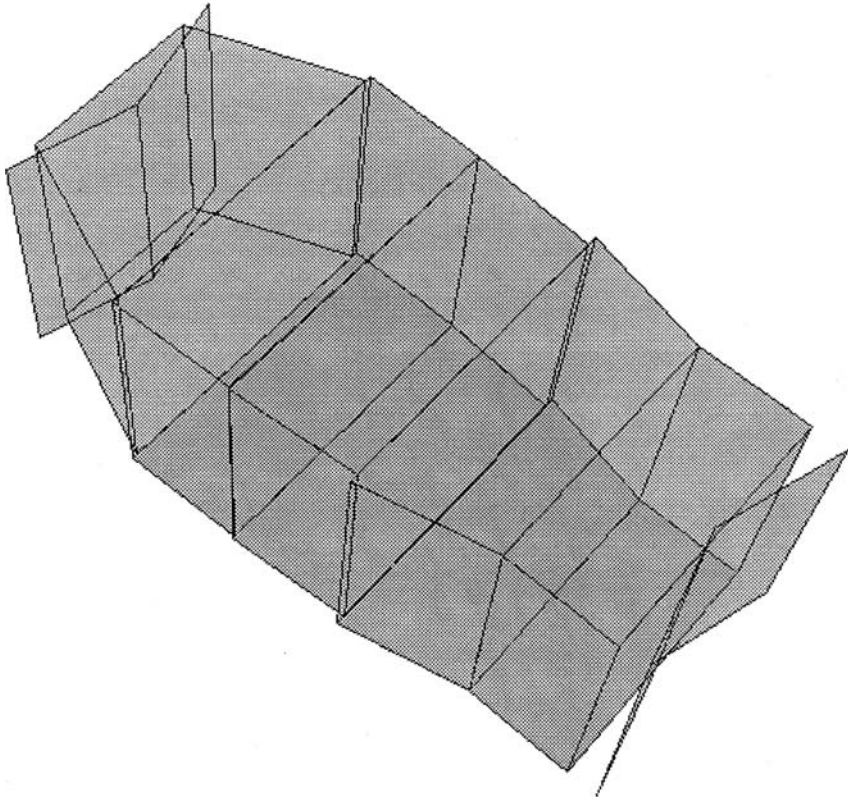
This preference-based optimization of architectural acoustics was applied to the design of the Kirishima International Concert Hall in cooperation with the architect Fumihiko Maki in 1992, as shown in Fig. 8.12 (Maki 1997; Ando et al., 1997; Nakajima and Ando, 1997). Acoustic design elements were as follows (Ando, 1998,



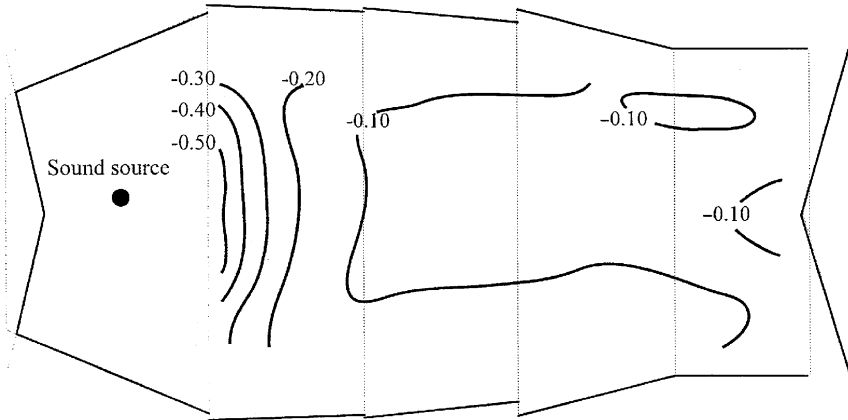
**Fig. 8.8** Resulting shape for model 2 optimized with respect to scale preferences for listening level,  $\bar{S}_1$



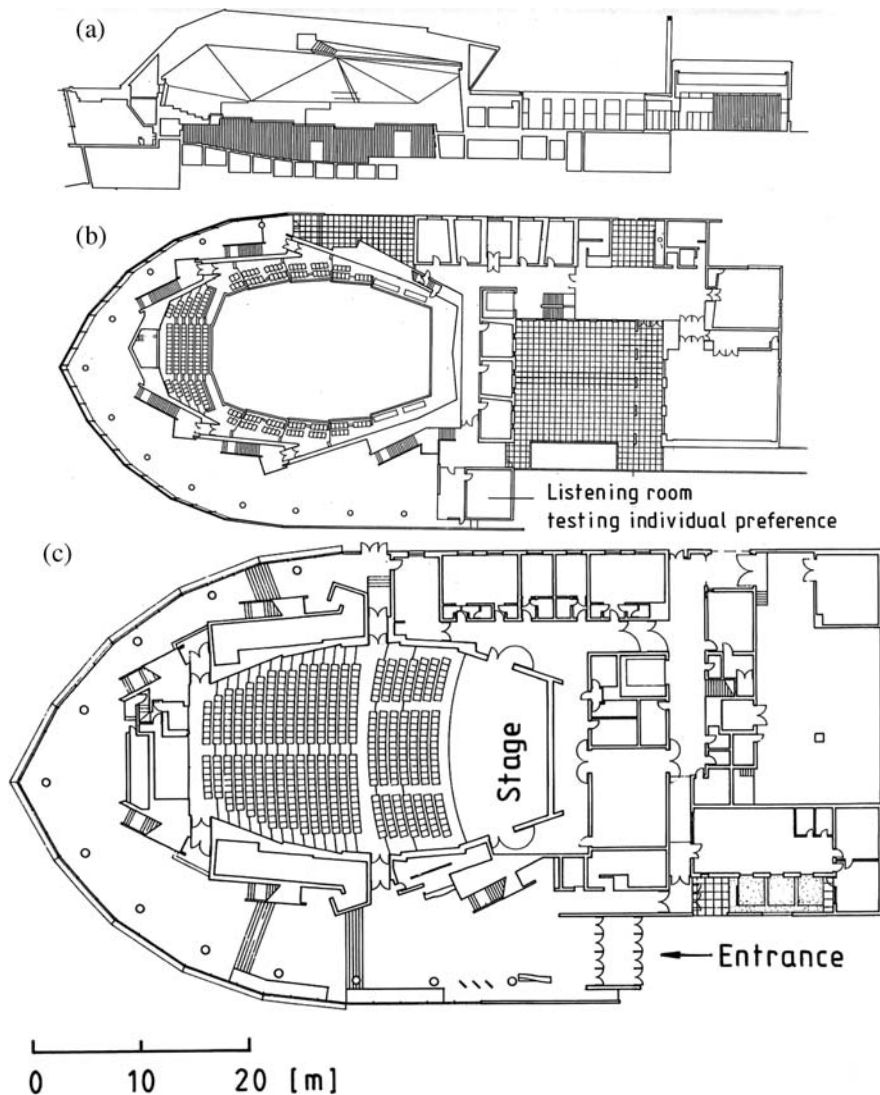
**Fig. 8.9** Contour lines of equal  $\bar{S}_1$  listening level preference values calculated for the geometry shown in Fig. 8.8



**Fig. 8.10** Resulting of shape for model 2 optimized with respect to scale preferences for sound diffuseness and envelopment that are associated with minimal IACCs, i.e.  $\bar{S}_4$



**Fig. 8.11** Contour lines of equal  $\bar{S}_4$  sound diffuseness preference values calculated for the geometry shown in Fig. 8.10



**Fig. 8.12** Scheme of the Kirishima International Concert Hall (Miyama Concertu) designed by architect Maki (1997). (a) Longitudinal section. (b) Plane of balcony level. (c) Plane of audience level

2007b): (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 (Photo 8.1).

Another example is the Tsuyama Music Cultural Hall (Suzumura et al., 2000) with a shape similar to that of the Kirishima Hall. An additional design, introduced 52 columns (each 30 cm diameter) as a design element in the hall



**Photo 8.1** Kirishima International Concert Hall (Miyama Conceru)



**Photo 8.2** Tsuyama Music Cultural Hall with 52 columns each of 30 cm diameter around the walls

(Photo 8.2). The columns provide surfaces for scattering reflected sound waves for the higher-frequency range above 1 kHz. This avoids image shifts of sound sources on the stage, and brings about a smaller value of IACC for areas near the columns and at the seating-center area close to the stage. Because columns

weakened the specular reflection from the sidewalls by scattering effects (Fuji, Hotehama, Kato, Shimokura, Okamoto, Suzumura and Ando, 2004), sound pressure levels are decreased near the sidewalls and  $\Delta t_1$  is prolonged throughout the hall.

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 50–90 ms, because the subsequent reverberation time  $T_{\text{sub}}$  measured was almost constant, about 1.7 s when the audience is present. This canopy plays an important role in decreasing the IACC at the seating audience area close to the stage (Nakajima et al., 1992).

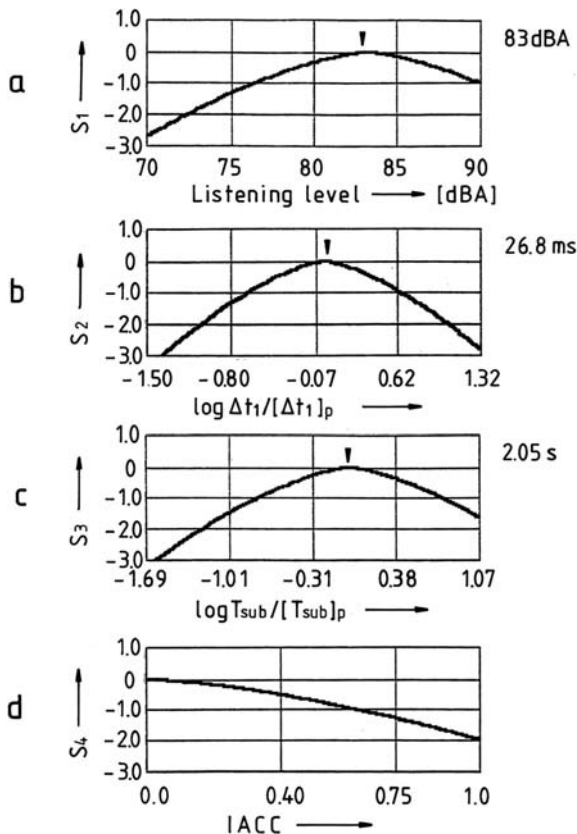
## 8.3 Individualized Seat Selection Systems for Enhancing Aural Experience

### 8.3.1 A Seat Selection System

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 based on the system described in Section 2.3 with multiple loudspeakers. The system used arrows for testing subjective preference of sound fields for listeners at the same time. Because the four orthogonal factors of the sound field influence the preference judgments almost independently, as was discussed in Chapter 3, a single orthogonal factor is varied, whereas 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).

### 8.3.2 Individual Subjective Preference

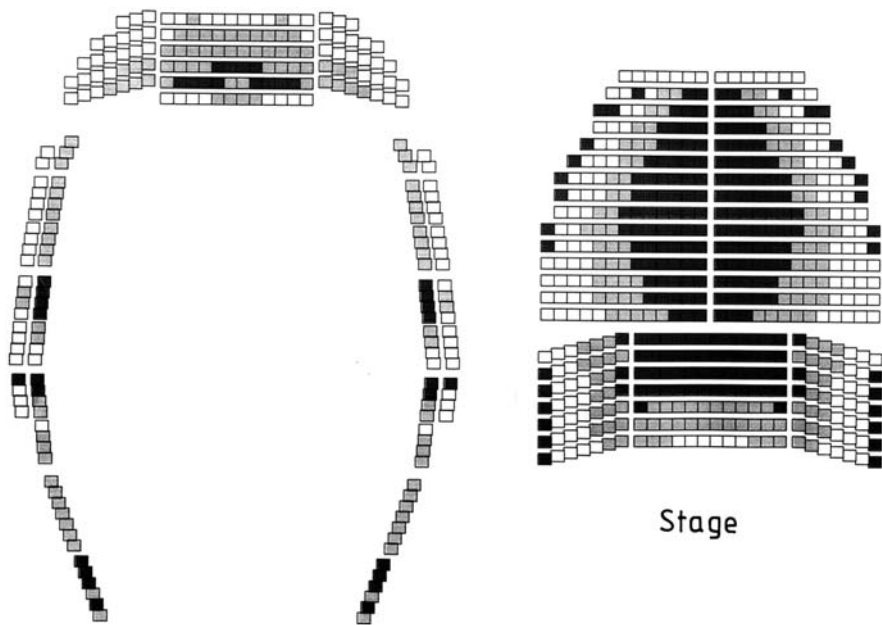
This set of experiments investigated the effects of taking individual listening preferences into account when choosing seat locations in a concert hall. The music source was the orchestral "Water Music" by Handel, which has an effective duration  $\tau_e$  of 62 ms. The total number of listeners participating for individual paired-comparison tests was 106 (Ando and Singh, 1996; Ando, 1998). Test results for listener BL, who was a typical listener, are shown in Fig. 8.13. The preferred scale values for this listener were close to the average for subjects previously collected: the most preferred listening level  $[LL]_p$  was 83 dBA, most preferred first reflection time  $[\Delta t_1]_p$  was 26.8 ms (close to 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



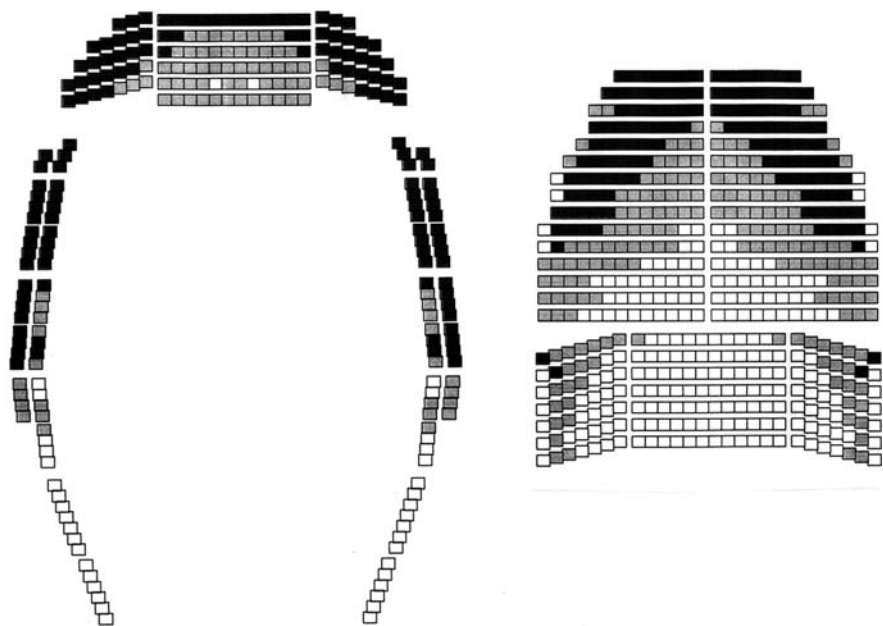
**Fig. 8.13** Examples of the scale value of subjective preference obtained by PCT for each of the four orthogonal factors of the sound field (subject B.L.). **(a)** S<sub>1</sub>: The most preferred listening level was 83 dBA, the individual weighting coefficient in Equation (3.7):  $\alpha_1 = 0.06$ . **(b)** S<sub>2</sub>: The preferred initial time delay gap between the direct sound and first reflection was 26.8 ms, the individual weighting coefficient in Equation (3.7):  $\alpha_2 = 1.86$ , where  $[\Delta t_1]_p$  calculated by Equation (3.7) with  $\tau_e = 62$  ms for the music used ( $A = 4$ ) is 24.8 ms. **(c)** S<sub>3</sub>: The preferred subsequent reverberation time was 2.05 s, the individual weighting coefficient in Equation (3.7):  $\alpha_3 = 1.46$ , where  $[T_{sub}]_p$ , calculated by Equation (3.4) with  $\tau_e = 62$  ms for the music used is 1.43 s. **(d)** S<sub>4</sub>: The preferred interaural crosscorrelation magnitude (IACC) was 0. Individual weighting coefficient in Equation (3.7):  $\alpha_4 = 1.96$

center area of seats was preferred for listener BL, similar to estimates that were calculated at the design stage (Fig. 8.14). With regard to the IACC, for all listeners 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 (Fig. 8.15). Listener KK indicated a preferred listening level exceeding 90 dBA. For this listener, the front seating areas close to the stage were preferable (Fig. 8.16). On the contrary, for listener DP, whose preferred listening level was rather low (76.0 dBA) and preferred initial delay time

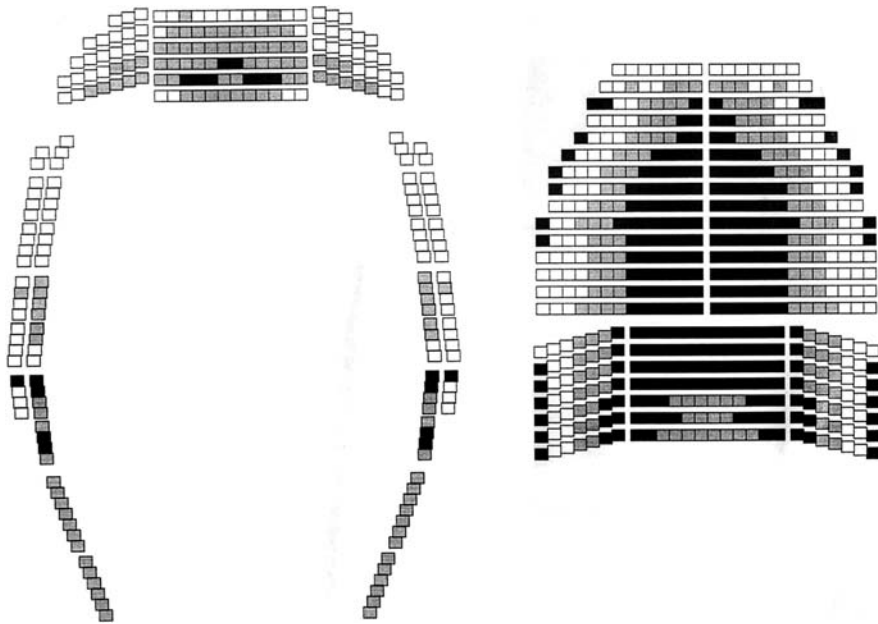




**Fig. 8.14** Preferred seating area calculated for subject B.L. The seats are classified in three parts according to the overall level of listener satisfaction estimated by Equation (3.6), which sums together scale preference factors  $S_1$  through  $S_4$ . Black portion of seats indicate preferred areas, about one third of all seats in this concert hall, for subject B.L.



**Fig. 8.15** Preferred seat area calculated for subject K.H.



**Fig. 8.16** Preferred seat area calculated for subject K.K.

short (15.0 ms), the preferred seats are in the rear part of hall (Fig. 8.17). The preferred initial time delay gap for listener AC exceeds 100.0 ms, but was not critical. Thus, all initial delay times are acceptable, but the IACC is critical. Therefore, the preferred seats were all located in the rear part of the hall (Fig. 8.18).

### 8.3.3 Distributions of Listener Preferences

In order to optimize the listening experience for the most people in a typical audience, assuming that they can find the best seats for their listening tastes, it is essential to take into account the distribution of listening preferences. Cumulative frequencies of the listening preferences of 106 listeners are shown in Figs. 8.19–8.21. As about 60% of listeners preferred the range of 80–84.9 dBA in listening to music. However, some listeners preferred higher sound levels above 90 dBA (Fig. 8.19), widening the total range of the preferred levels to beyond a 20 dB range. Roughly 45% of listeners preferred the initial delay times 20–39 ms (Fig. 8.20), near the predicted value of 24.8 ms (Equation 4.9). Here some listeners preferred 0–9 ms, while others preferred delays of more than 80 ms. About 45% of listeners preferred reverberation times of 1.0–1.9 s (Fig. 8.21) close to the predicted preferred value of 1.43 s. However, as with the other factors, other listeners indicated preferences away from this norm: less than 0.9 s or more than 4.0 s.

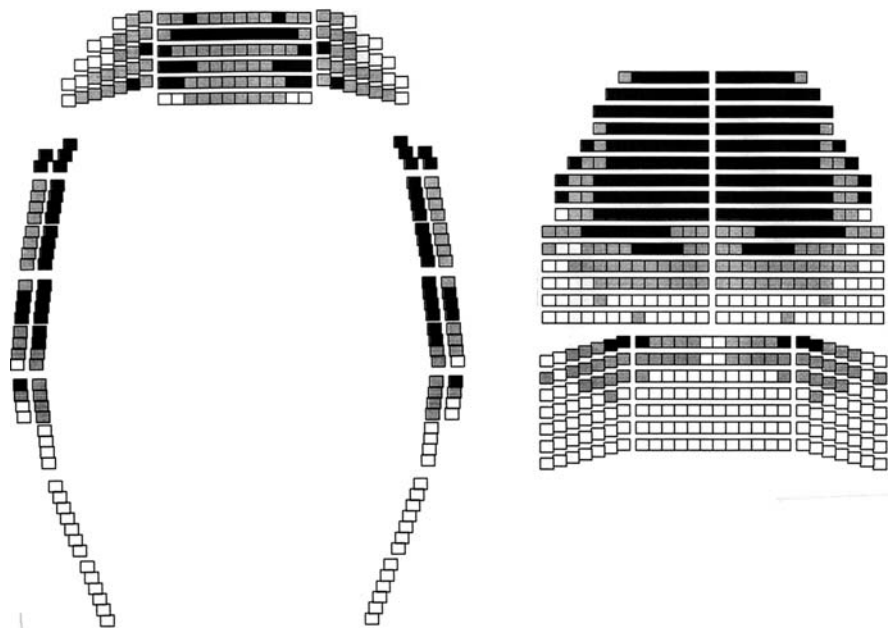


Fig. 8.17 Preferred seat area calculated for subject D.P.

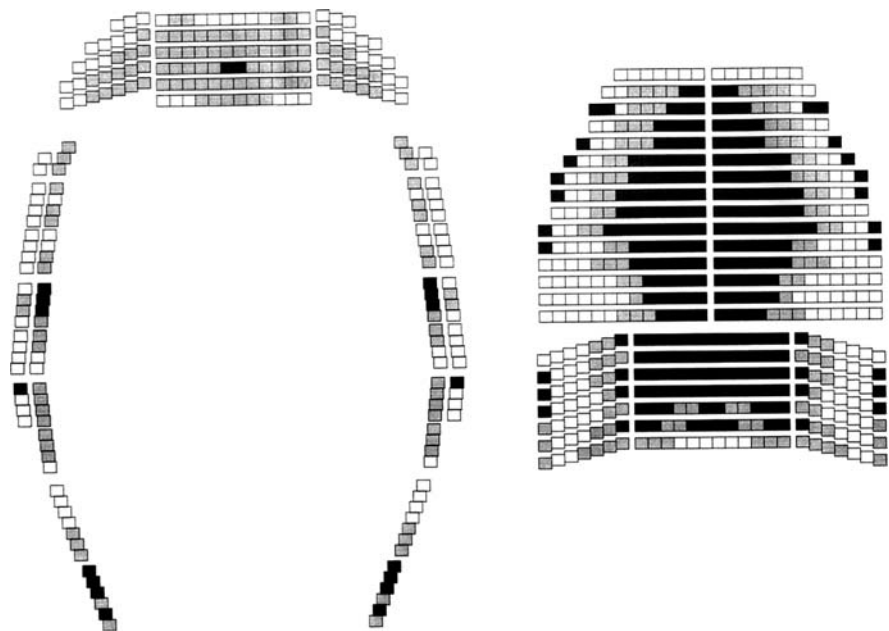
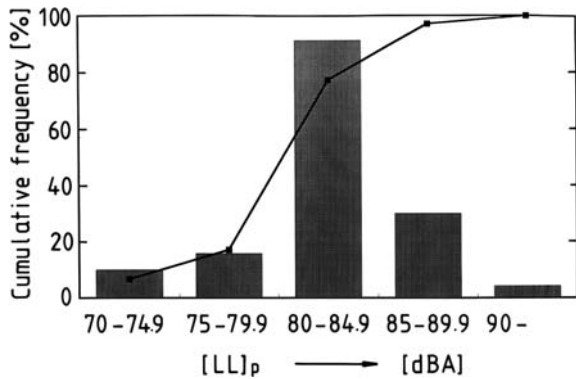
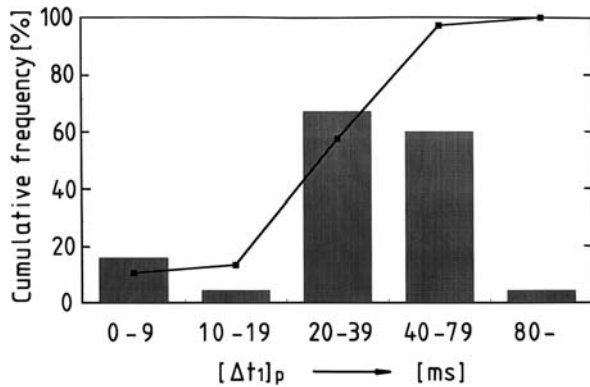


Fig. 8.18 Preferred seat area calculated for subject C.A.

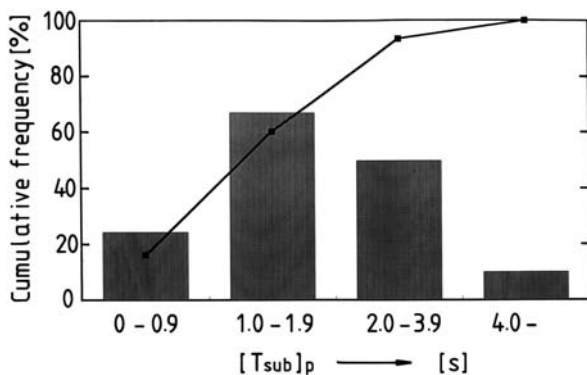
**Fig. 8.19** Cumulative frequency of preferred listening level  $[LL]_p$  (106 subjects). About 60% of subjects preferred the range 80–84.9 dBA



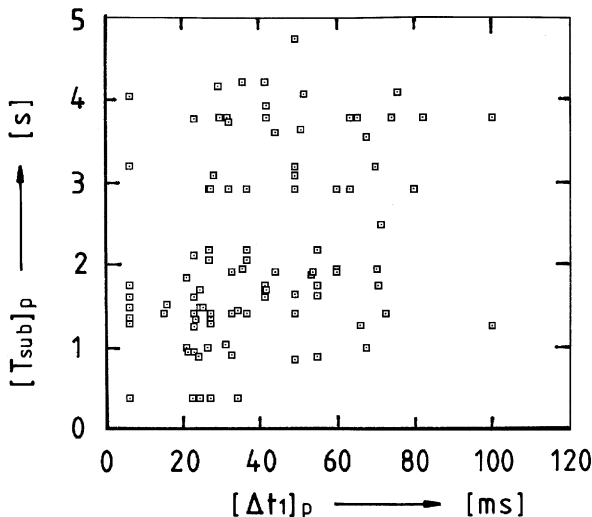
**Fig. 8.20** Cumulative frequency of the preferred initial time delay gap between the direct sound and the first reflection  $[\Delta t_1]_p$  (106 subjects). About 45% of subjects preferred the range 20–39 ms. Calculated value of  $[\Delta t_1]_p$  by Equation (3.3) is 24.8 ms



**Fig. 8.21** Cumulative frequency of the preferred subsequent reverberation time  $[T_{sub}]_p$  (106 subjects). About 45% of subjects preferred the range 1.0–1.9 s. Calculated value of  $[T_{sub}]_p$  by Equation (3.5) is 1.43 s

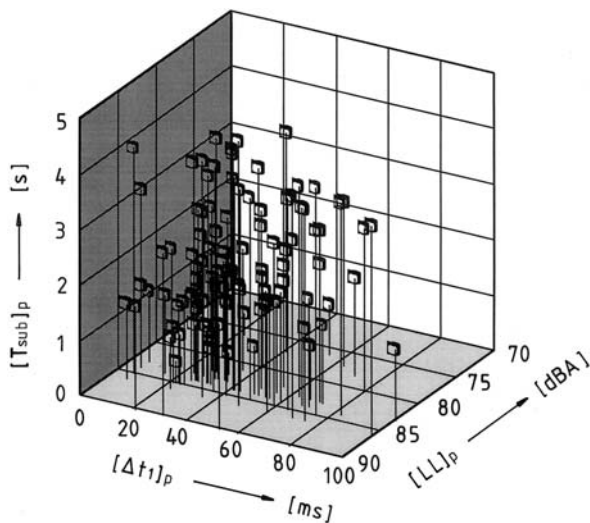


**Fig. 8.22** Relationship between preferred values of  $[\Delta t_1]_p$  and  $[T_{sub}]_p$  for each subject. No significant correlation between them was observed



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, little correlation exists ( $r = 0.06$ ) between preference values of  $[\Delta t_1]_p$  and  $[T_{sub}]_p$  (Fig. 8.22). The same is true for correlations between  $[T_{sub}]_p$  and  $[LL]_p$  and between  $[LL]_p$  and  $[\Delta t_1]_p$ , where  $r < 0.11$ . Figure 8.23 shows the three-dimensional plots of the preferred values of

**Fig. 8.23** Three-dimensional illustration of preferred factors,  $[LL]_p$ ,  $[\Delta t_1]_p$ , and  $[T_{sub}]_p$  of the sound field for each individual subject. Preferred condition for the factor of IACC is excluded, because no fundamental individual differences could be observed. Preferred conditions are distributed in a certain range of each factor, such that subjects could not be grouped into any natural classes



$[LL]_p$ ,  $[\Delta t_1]_p$ , and  $[T_{\text{sub}}]_p$  excluding the consensus factor of IACC. Looking at a continuous distribution of preferred values, no specific groupings of individuals or natural classes emerged from the data.

In calculations using Equation (3.8), there is no correlation between weighting coefficients  $\alpha_i$  and  $\alpha_j$ ,  $i \neq j$ , ( $i$  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's preference can critically depend on some factors, while being insensitive to other factors. This results in individual characteristics that may be distinct from other listeners. This is an indication of individual difference or taste.

## 8.4 Subjective Preferences of Cello Soloists for First Reflection Time, $\Delta t_1$

We evaluated the subjective preferences of five cello soloists to provide knowledge useful in designing the stage enclosure in a concert hall. This study investigated their subjective preferences, with regard to ease of performance, for first reflection time. The scale value of preference for the delay time of the single reflection was obtained using paired comparison tests, and the results were compared with those for alto-recorder players and listeners. Scale preference values for cellists 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  $(\tau_e)_{\text{min}}$  of the running ACF of the music played by each cellist.

In order to produce superior concerts, we need to design the sound fields not only in the audience area for the listeners but also in the stage area for the performers. The primary issue is that the stage enclosure should be designed to provide a sound field in which performers can play easily.

Marshall et al. (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 the performers. On the other hand, the preferred delay time of the 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, it is more accurately expressed by the minimum value of the effective duration  $(\tau_e)_{\text{min}}$  of the running ACF of the source signal (Ando et al., 1989; Mouri et al., 2000). Nakayama (1984) showed that the amplitude of the reflection and the duration of the long-time ACF of the source signal in a similar manner could determine the preferred delay time of a single reflection for alto-recorder soloists (for a modification, see Ando, 1998). We investigated whether the most preferred condition of the single reflection for an individual singer may be described

by the  $(\tau_e)_{\min}$  and a modified amplitude of reflection by the overestimate and bone-conduction effect (Noson et al., 2000, 2002).

The current study examines whether or not the preferred delay time of a single reflection for the individual soloist can be calculated by the minimum value of the effective duration of the running ACF of the music signal played by the cellist (Sato et al., 2000). Five cellists participated in the experiment. The same music motifs (motifs I and II) used in the experiments conducted by Nakayama (1984) were applied here. The tempo of motif I was faster than that of motif II as shown in Fig. 8.24. 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  $50 \pm 1.0$  cm. The music tempo was maintained with the help of a visual and silent metronome. Each music motif was played 3 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  $\tau_e$  by the extrapolation at  $-10$  dB as described in Section 5.2. Examples of effective durations of the running ACF for music motif I played by subjects B and E are shown in Fig. 8.25. The minimum value of the effective duration  $(\tau_e)_{\min}$  of the running ACF for each cellist and each session are listed in Table 8.3. 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

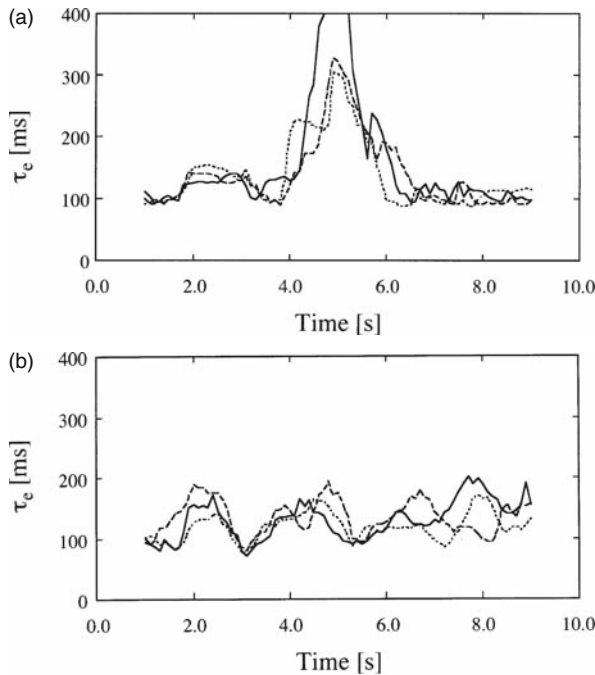
#### Music Motif I



#### Music Motif II



**Fig. 8.24** Music scores of motifs I and II composed by Tsuneko Okamoto applied for experiment with cellists (Ando, 1998)



**Fig. 8.25** Examples of the measured effective duration of the running ACF with the 100-ms interval as a function of time. Each music motif was played 3 times by each cellist. (—): First session; ----: second session; ( . . . . ): third session. (a) Music motif I for subject B,  $(\tau_e)_{\min} = 50 \pm 2$  ms. (b) Music motif I for subject E,  $(\tau_e)_{\min} = 37 \pm 1$  ms

ranges of  $(\tau_e)_{\min}$  are within  $\pm 5$  ms. Individual differences in the effective durations of the running ACF may depend on the performer's style.

The single reflection from the back wall in the stage enclosure was simulated in an anechoic chamber by a loudspeaker  $80 \pm 1.0$  cm measured from the head of the cellist. The sound signal was picked up by a half-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 amplitude of reflection  $A_1$ , relative to that of the direct sound measured at the entrance of the performer's left ear, was kept constant at  $-15$  dB or  $-21$  dB when the cellist played the musical note "a" (442 Hz).

The preferred delay time of the single reflection was assumed to depend on the  $(\tau_e)_{\min}$  of the running ACF of source signal. The PCT was conducted for five sound fields, in which the delay time of reflection was adjusted for every cellist according to results of  $(\tau_e)_{\min}$  listed in Table 8.3. The subjects were asked which of the 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 3 times interchanging the order of the pairs. It took about 20 min for each cellist and for each music motif. Fifteen responses (five subjects  $\times$  three repeats) to each



**Table 8.3** Minimum values of running  $\tau_e$  of ACF for music motif played by each cellist

Cellist	Session	Music motif I (ms)	Music motif II (ms)
A	First	35	90
	Second	41	96
	Third	41	89
B	First	52	92
	Second	49	87
	Third	49	89
C	First	37	89
	Second	38	86
	Third	36	93
D	First	57	87
	Second	56	85
	Third	54	86
E	First	37	71
	Second	38	74
	Third	36	79
Averaged		46	84

sound field were obtained and were confirmed by consistency tests. The scale values of preference for each cellist were obtained (Ando and Singh, 1996; Ando, 1998).

Figure 8.26 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 8.4. Global and individual results (except for that of subject E) for music motif II were longer than those for music motif I.

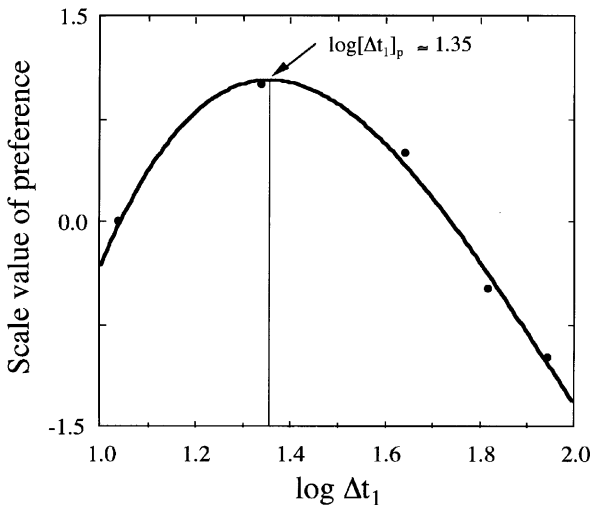
The most-preferred delay time of the single reflection also is described by the duration  $\tau'_p$  of the ACF as similar to that of listeners (see Section 3.2.2), which is expressed by

$$[\Delta t_1]_p = \tau'_p \approx [\log_{10}(1/k') - c' \log_{10} A'](\tau_e) \min \quad (8.1)$$

where the values  $k'$  and  $c'$  are constants that depend on a musical instrument. 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 overestimation of the reflection by a performer. This is called "missing reflection" of a performer.

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 were respectively  $0.1$  and  $1$ . After setting  $k' = 1/2$ , we obtained the coefficient  $c'$  for each individual as listed in Table 8.5. The average value of the coefficient  $c$  for the five cellists obtained was about  $1.0$ . The relation between the most-preferred delay time  $[\Delta t_1]_p$  obtained by preference judgment and

**Fig. 8.26** An example of the regression curve for the preferred delay time (subject D, music motif I, -15 dB),  $\log[\Delta t_1]_p \approx 1.35$ , that is,  $[\Delta t_1]_p \approx 22.6$  (ms)



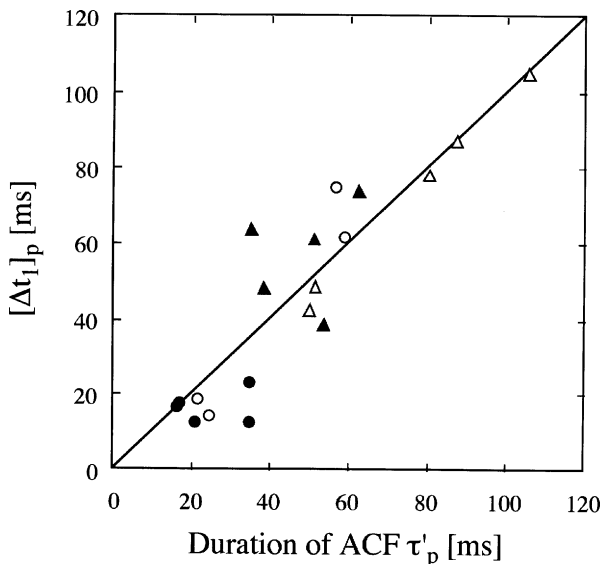
**Table 8.4** Judged and calculated preferred delay times of a single reflection for cello soloists. Calculated values of  $[\Delta t_1]_p$  are obtained by Equation (8.1) using the amplitude of the reflection  $A_1$  and  $(\tau_e)_{\min}$  for music signal performed by each cellist

A (dB)	A' (dB) (= A + 10)	A'	Cellist	Judged $[\Delta t_1]_p$ (ms)		Calculated $[\Delta t_1]_p$ (ms)	
				Motif I	Motif II	Motif I	Motif II
-15	-5	0.56	A	16.2	47.9	16.3	38.5
			B	<12.0	73.8	35.2	62.7
			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
-21	-11	0.28	A	18.1	48.4	21.8	51.5
			B	61.2	105.0	59.3	105.6
			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

the duration  $\tau'_p$  of the ACF calculated by Equation (8.1) using  $(\tau_e)_{\min}$  is shown in Fig. 8.27. 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 Fig. 8.28. Different symbols indicate the scale values obtained in different test series. Each symbol has 25 data sets (five subjects  $\times$  five sound fields)

**Table 8.5** Coefficients  $c'$  in Equation (8.1) for calculation of the preferred delay times of the reflection for individual cellists and for the global average (the coefficient  $k'$  is fixed at  $1/2$ )

	Cellist					Averaged (global)
	A	B	C	D	E	
Coefficient $c'$	0.47	1.61	1.10	1.30	0.67	$\approx 1$



**Fig. 8.27** Relationship between the most preferred delay time  $[\Delta t_1]_p$  and the duration  $\tau'_p$  of the ACF calculated by Equation (9.1). Correlation coefficient,  $r = 0.91$  ( $p < 0.01$ ). ●, music motif I,  $-15$  dB; ○, music motif I,  $-21$  dB; ▲, music motif II,  $-15$  dB; △, music motif II,  $-21$  dB

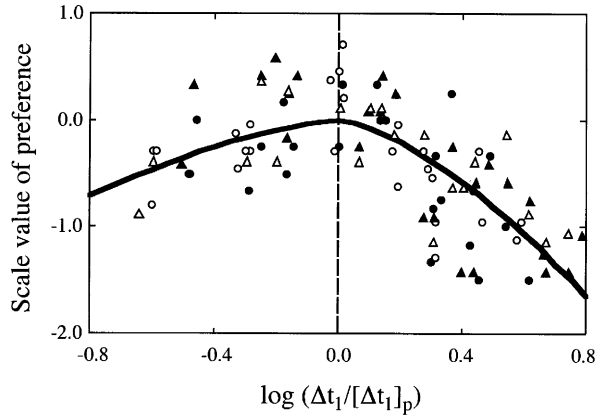
except for the amplitude of  $-15$  dB for music motif I (for which there were 20 data sets because consistency tests did not indicate a significant ability to discriminate preference in the results of subject C). Although the scale values were obtained in different test series, tendencies are consistent with each other. The regression curve is expressed by

$$S = S_L - \alpha|x|^\beta \tag{8.2}$$

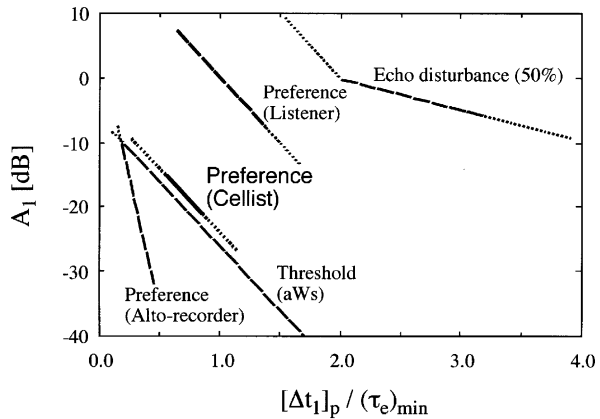
where  $x = \log \Delta t_1 / [\Delta t_1]_p$ , the power of  $x$  may be always fixed by  $\beta = 3/2$ , and the weighting coefficient  $\alpha$  is 2.3 for  $x \geq 0$  and 1.0 for  $x < 0$ .

Figure 8.29 shows the relative amplitude of a single reflection to that of the direct sound for the preference of cello soloists as a function of the delay time of a single reflection normalized by the minimum value of the effective duration  $(\tau_e)_{\min}$  of the running ACF, as well as several subjective responses as a function of the delay time of a single reflection normalized by the value of the effective duration  $\tau_e$  of the

**Fig. 8.28** Scale values of preference for each of five cellists as a function of the delay time of a single reflection normalized by its most preferred delay time calculated by Equation (9.3).  $\bullet$ , music motif I, -15 dB;  $\circ$ , music motif I, -21 dB;  $\blacktriangle$ , music motif II, -15 dB;  $\triangle$ , music motif II, -21 dB. The regression curve is expressed by Equation (8.2)



**Fig. 8.29** Relative amplitude of the single reflection for the preference of cello soloists as a function of the delay time of a single reflection normalized by the value of  $(\tau_e)_{\min}$ . Also, the amplitudes of several subjective responses as a function of the delay time of the single reflection normalized by the value of  $(\tau_e)_{\min}$  of the source signal are plotted



long-time ACF of the source signal. All these values can be calculated by Equation (8.1) with constants  $k$  and  $c$  for each subjective response (see Table 3.2). The alto-recorder soloist's preference is also plotted in this figure. The values for cellists are close to the threshold of perception (absolute Wahrnehmbarkeits-schwelle, aWs) for listeners. These reconfirm the phenomenon of “missing reflection” for performers.

As an application, adjusting the height of the reflectors above the stage can control the delay time of a reflection. As listed in Table 8.6, the optimum distance between the performer and the 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. Here it is assumed that the distance between the instruments and the ear of the performer is 60 cm for a cello soloist and 20 cm for an alto-recorder soloist. The height of the reflector above the stage can be adjusted if the minimum value of the effective duration  $(\tau_e)_{\min}$  of the running ACF of the music to be played is measured before the concert. For practical convenience, this

**Table 8.6** Optimum distances between the performer and the reflector calculated from Equation (8.1) in relation to the value of  $(\tau_e)_{\min}$  for the music signal played

$(\tau_e)_{\min}$ of the music signal (ms)	Distance of the reflector (m)						Alto-recorder soloist
	Cello soloist						
	A	B	C	D	E	Averaged	
30	3	10	6	8	4	6	2
50	6	21	13	16	8	13	4
70	9	(33)	21	(26)	13	20	6
90	13	(46)	(30)	(36)	18	(29)	8

Note: The value of  $\tau_e$  for alto-recorder soloist was obtained for a long-time ACF ( $2T = 32$  s).

adjustment may be made in the real sound field with the reverberation. In this situation, the total amplitude of the reflections might replace the amplitude of the single reflection.

The most-preferred delay time of a single reflection for each cellist can be calculated by Equation (8.1) 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 for global cellists with regard to the delay time of the single reflection can be expressed by such a simple formula, normalizing the delay time by the most-preferred delay time observed for different music motifs.

## 8.5 Concert Hall as Musical Instrument

When performers on the stage play a musical instrument, the concert hall acts as a second instrument. Let us now discuss what kind of musical expressions the second instrument can produce. In Chapters 6 and 7, we have described temporal sensations related to the temporal factor of the sound field and spatial sensations in relation to the spatial factor. Musical expressions in composition and performance consist of not only temporal expressions but also spatial expressions (Tables 8.7 and 8.8) in a given concert hall (Ando, 2007a).

### 8.5.1 *Composing with the Hall in Mind: Matching Music and Reverberation*

In music composition, a composer can consider the reverberation time of the spaces in which the music is likely to be performed. For example, Mozart-style motifs at fast tempos might be composed for a guest entry hall in a Court with a reverberation time of about 1.5 s, but not intended for a church with a long reverberation time

**Table 8.7** Methods of blending the temporal characteristics of sounds and sound fields in concert halls by controlling effective duration  $(\tau_e)_{\min}$  value of source signals. Note that effective durations of impulsive sounds and broadband noise tend towards 0, whereas those of sustained periodic sounds equal the duration of the entire note

Composition and performance	To make decrease of the $(\tau_e)_{\min}$ value	To make increase of the $(\tau_e)_{\min}$ value
<b>Composition</b>		
Note	Whole note	. . . 64th note
Tempo	Fast	Slow
Frequency component	High	Low
Selection of instruments	Vocal, speech (consonants)	Pipe organ
<b>Performance</b>		
Piano performance	Staccato	Legato, super legato, full pedal
String instruments and vocal	Vibrato	Less vibrato
Vocal	Fushi-Mawashi (intonation) <sup>1</sup> Grace note	

<sup>1</sup> Kato et al. (2007b).

**Table 8.8** Musical composition and performance of temporal and spatial expressions in a given concert hall

	Temporal expression	Spatial expression
Composition	Selection of tempo, note (whole note, . . . 64th), and selection of musical instrument, blending ACF $(\tau_e)_{\min}$ of music sources and the temporal factor of a given concert hall. To enhance reverberance of a hall, introduction of tremolos and rests are effective.	(1) Selection of frequency component for the spatial sensation (ASW). (2) Selection of dynamics ( <i>ppp</i> – <i>fff</i> ) for subjective diffuseness and/or envelopment.
Performance	(1) Selection of music programs performed in a given concert hall. (2) Selection of performing position on the stage for ease of performance of music program. (3) Selection of style such as vibrato extent, staccato–super legato for ACF $(\tau_e)_{\min}$ of music source according to the temporal factor in a given concert hall.	(1) Selection of performing position on the stage getting small values of IACC at most seating positions. (2) Control of source strength as an interpretation of music expressing subjective diffuseness, envelopment or “embracement” with full sound-surround of listeners.

of over 4s or a lecture room with a short reverberation time of less than 1 s. The scientific mathematical expressions needed for matching source signals with appropriate temporal factors of sound fields are described in Sections 3.2.2 and 3.2.3. For example, when the acceptable reverberation time is defined by the scale value  $S = -0.1$  in reference to  $S = 0$  at the most preferred reverberation time  $[T_{sub}]_p$ , then the reverberation time should be restricted to a range of  $T_{sub} = 0.5$  to  $1.3 [T_{sub}]_p$ .

If the room has a shorter reverberation time than the preferred value  $[T_{sub}]_p$ , then the music can be adapted to consist of a rapid movement without any repetitive features, so that the value of  $(\tau_e)_{min}$  would then tend to be smaller. For example, introduction of tremolos and rests may enhance the reverberation of the concert hall. On the other extreme, if a music source is a slow tempo and with a long duration of tones – like pipe organ music – then performance space with a long reverberation time would be well blended with this musical composition (Fig. 8.30).

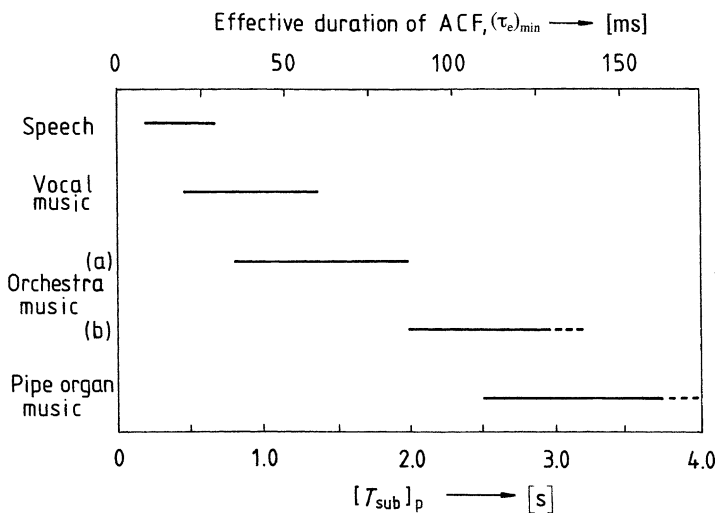


Fig. 8.30 Recommended reverberation times for several music/instrument sound programs

### 8.5.2 Expanding the Musical Image: Spatial Expression and Apparent Source Width

It has been shown that the apparent source width (ASW) may be described by both spatial factors  $W_{IACC}$  and  $IACC$  as shown in Fig. 7.5 (Sato and Ando, 1999). If the value of the  $W_{IACC}$  is large due to the low-frequency component of the music signal particularly below 200 Hz, then the sound source will be perceived as “wide.” For example, *The Moldau*, composed by Bedrich Smetana (1824–1884), consists of high frequencies in the early portion of the piece, producing a “narrow” image of two

upper valleys, both small sources of water. Then, as the piece develops, it becomes an expression of the joining of both streams into one and consists of increasing low-frequency spectral component producing a wide image of the downstream flows into a broad valley.

Another example is Piano Sonata No. 14 in C sharp minor, Op. 27, No. 2 (“Moonlight”) by Ludwig van Beethoven, in which there are heavy low-frequency components (large  $W_{IACC}$ ) composed to represent an ambient vision of a big blue sky with a big body of water probably and spacious ground, while accompanying this with high-frequency components (small  $W_{IACC}$ ) representing a focused vision of the moon.

### ***8.5.3 Enveloping Music: Spatial Expression and Musical Dynamics***

If the acoustic design for IACC in most of the seating area in a concert hall is small enough, then the strength of stage sound sources at volumes “*ppp*” to “*fff*” can be used to control subjective diffuseness as well as envelopment (Damaske, 1967/68). Musical dynamics and the resulting amplitude of reflections prevailing at the threshold level will decrease the perceived IACC, resulted in subjectively diffuse sound, while intentionally produced low dynamic levels can be used by the composer to produce a narrowing of the sound field environment.

## **8.6 Performing in a Hall: Blending Musical Performances with Sound Fields**

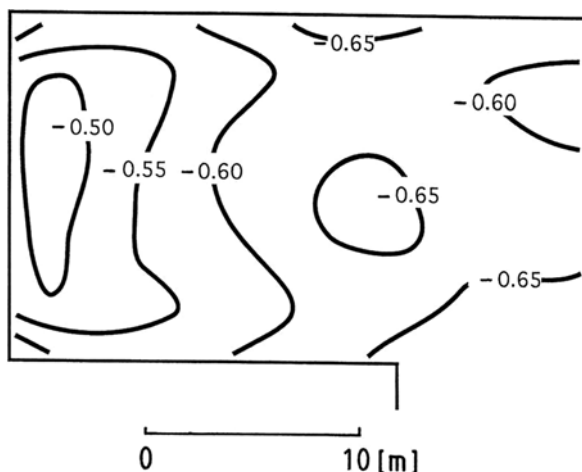
### ***8.6.1 Choosing a Performing Position on the Stage***

A performer can optimize the experienced quality of his or her musical performance by judiciously choosing where on the stage the music will be played. Soloists should select a position on the stage that yields an initial delay time of reflection,  $\Delta t_1$  that is most favorable to the music that they will perform. This in turn is mainly related to the range of the  $(\tau_e)_{\min}$  value (Ando, 1998; Sato et al., 2000; Noson et al., 2002). The soloist may select a location on stage at the time of rehearsal, adopted both to ease of performance and thus to enhanced listener satisfaction. It is further recommended that a soloist situate near the middle or rear of the stage, rather than close to the front, in order to produce a small value of IACC most seating positions in the audience area.

The performing position that minimizes the IACC of the sound field for the listener’s seats is demonstrated here by means of an example. The values of IACC are calculated with music motif B [Arnold;  $(\tau_e)_{\min} = 35$  ms] at 112 listener positions in a Békésy Courtyard (Békésy, 1934). For simplicity, the directivity of the sound source is assumed to be uniform in this calculation. The height of the sound source



**Fig. 8.31** Contour lines of equal averaged IACC calculated to find optimum performing position. The value was averaged for 112 listening positions. The most effective location of performance is found in the area IACC < 0.5



is 80 cm from the ground level, and the height of the listeners' ears is 110 cm. The contour lines of equal average values of the IACC for 112 listener positions are calculated to find the optimum performing position and are then plotted in Fig. 8.31. The effective positions for performance may be found in the area minimizing IACC for all listening positions within the area of IACC < 0.5. A more effective procedure is that this positioning of a given concert space could be suggested by a “sound coordinator” of each concert hall (Section 8.6.3).

### 8.6.2 Performance Adjustments that Optimize Temporal Factors

Given a particular performing space, performers can adapt their playing style to match the acoustics of the space, and thereby enhance the perceived quality of the sound that they produce. One seeks to match the effective duration of the music performed with the reverberatory characteristics of the hall. Thus a pianist confronted with a smaller hall with short reverberation times can decrease the effective duration of the performed music by introducing staccato instead of legato, *supper legato* and full pedal (Taguti and Ando, 1997). One example is a performance by Glenn Gould, who performed by staccato, *The Art of Fugue*, *Contrapunctus II & IX*, that was composed by Johann Sebastian Bach (1685–1750). One can also control the minimum value of  $(\tau_e)_{\min}$  in vocal performance, which determines the preferred temporal condition for vocal music. This strategy has been discussed as a means for blending the sound source and a given concert hall (Kato and Ando, 2002; Kato et al., 2004, Fujii et al., 2006, 2007). When vibrato is introduced during singing, for example, it can decrease the value of  $(\tau_e)_{\min}$ , blending the sound field with a shorter reverberation time.

Because a characteristic of vibrato depends on the individual performer, the individual performer can be strongly urged to attain a skill for controlling the  $(\tau_e)_{\min}$  value by use of a real-time ACF analyzer of music signals (Kato et al., 2006). In the

same manner, any performer may also play with a certain amount of vibrato or its equivalent for the particular instrument.

Conversely, if the reverberation time is long and the dimension of a given hall is large, then the  $(\tau_e)_{\min}$  value should be controlled to be long. For example, the pianist can produce a long  $(\tau_e)_{\min}$  value by legato, super legato, and full sustain pedal instead of staccato. And, the singer and the violinist can produce a long  $(\tau_e)_{\min}$  value by decreasing the extent of vibrato.

Table 8.7 summarizes suggestive methods of controlling the  $(\tau_e)_{\min}$  value of source signals.

In July 2004, the author requested Tsuyoshi Tsutsumi to present an invited special paper on blending cello music and the sound field in a concert hall. He immediately accepted this invitation and said that it is a “musical lifeline.” Tsutsumi (2006) has shown how cello music can be blended with the specific case of the sound field at the Kirishima International Music Hall. On the Internet, his musical performance is available at the Web site of the *Journal of Temporal Design in Architecture and the Environment* [Vol. 6, No. 3 (2006) 78–81; <http://www.jtdweb.org/>].

### ***8.6.3 Towards Future Integration of Composition, Performance and Hall Acoustics***

The scientific approach made here suggests that further dimensions of musical temporal and spatial expressions in composition can be based on a concert hall’s acoustics (Table 8.8). In blending music sources performed on the stage with the temporal factor of the sound field in a given concert hall, we may take the effective duration of the ACF of the source signal into account, both for practical considerations (satisfaction of audience) and for artistic purposes (expressivity of the composer and performer). For spatial expressions, the strength of music source enhancing spatial sensations due to the perceived IACC and the frequency component due to  $W_{IACC}$  could be carefully included in the production of each musical note.

After selecting a suitable performing position on the stage, music sources and the sound field in a concert hall may be fully blended by control of the temporal and the spatial expressions. For the temporal expression of the performer, methods of blending the temporal factor of the sound field have been proposed based on the effective duration of ACF of the sound sources. For spatial expression, the minimum source energy exceeding the threshold is needed to realize subjective diffuseness or envelopment for listeners.

It is hoped that the survey presented here might encourage musicians in furtherance of music composition and performance – using simultaneously the two primary instruments: the musical instrument and the given concert hall enclosure.

As a temporal design, if the conductor or music director is aware of the acoustics of a concert hall, they can plan a program of music that will sound best in that hall in terms of the temporal factors involved. This mainly depends on the minimum value of effective duration of the ACF of source signals,  $(\tau_e)_{\min}$  as discussed in Chapter 3.

It has been found, for example, that music with rapid sound movements or vibrato can decrease the value of  $(\tau_e)_{\min}$ , which best fits a concert hall with a short initial time delay gap  $\Delta t_1$ , and a short subsequent reverberation time  $T_{\text{sub}}$ . Music with a slow tempo usually sounds best in a hall with relatively long values for the factor  $\Delta t_1$  and  $T_{\text{sub}}$ .

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 that matches the acoustics of a given concert hall. Each major concert hall would have on hand such an expert in architectural and music acoustics, who could work with its music director to plan concert programs. The sound coordinator could suggest, for instance, appropriate music programs to be performed in the hall, optimal stage positions for performers, and possible modulations of the performing style. The professional role of this expert would be similar to the specialist of an art museum, with qualifications involving formal training at the professional school or college level in architectural and music acoustics. The existence of such positions would further new artistic creations that would utilize simultaneously the two primary instruments for the effective presentation of sounds: the musical instrument and the concert hall.

Beyond this, in order to realize truly excellent constructed environments, one should always explicitly consider both temporal and spatial values in the design process at its outset (Ando et al., 1996; Ando, 2004). A general theory of temporal and spatial design of environments has been proposed (Ando, 2009). Further development of temporal design ideas in architecture and the environment can be accessed in the Journal of Temporal Design (JTD, <http://www.jtdweb.org/>), which has been published since 2001.

## Chapter 9

# Applications (II) – Speech Reception in Sound Fields

Our correlation-based auditory signal processing model can be applied to the perception of a single syllable in a sound field. Reception of speech signals reproduced in the frontal direction for a direct sound with a single echo is well described in terms of temporal factors that are extracted from the running ACF of the sound signal. Effects of noise disturbances from different horizontal angles on error rates of single syllable identification were investigated. Results show that syllable non-identification (NI) rates can be predicted from temporal factors extracted from the ACF and from spatial factors extracted from the IACF. Perceptual dissimilarity of sounds due to different source locations was assessed in psychophysical experiments and accurately estimated using temporal and spatial factors.

### 9.1 Effects of Temporal Factors on Speech Reception

This section describes a method of calculating the speech intelligibility (SI) of each syllable in terms of the distance between the direct sound as a template and the sound field with a single echo. In the calculation of the distance, three temporal factors as well as the sound energy  $F(0)$  extracted from the ACF are taken into account. The speech transmission index (STI) for the global speech intelligibility of sound fields is well known (Houtgast et al., 1980). We have also proposed a method for calculating the speech intelligibility of single syllables (Ando et al., 1999, 2000).

A perceptual distance between a template source signal and a sound-field signal is introduced. This distance quantifies the effect of room reverberations in altering the perceptual representation of the signal, such that identification is degraded. The greater the perceptual distance between the two signals, the more different they are perceived to be. The greater the perceptual difference between the signal at its source and the signal received by the listener, the lower the likelihood that the listener will recognize and identify it correctly.

A distance between the template source signal and the sound-field signal is introduced. Let  $S_K^T$  be the characteristics of an isolated template-syllable K, and let  $S_X^{SF}$  be that of another syllable X in the sound field; symbol T refers to the template, and SF is the sound field. Let  $C_K^T$  and  $C_X^{SF}$  be characteristics of the

isolated template  $S_K^T$  of syllable K and another syllable X in a sound field  $S_X^{SF}$  being processed in the auditory-brain system. Then, the distance between  $S_K^T$  and  $S_X^{SF}$  is given by

$$d_k = D(S_X^{SF}, S_K^T) = |C_X^{SF}, C_K^T| \quad (9.1)$$

where

$$\begin{aligned} D_{\tau_e}(X, K) &= \left[ \sum_{i=1}^I \left| \log(\tau_e^i \left( \frac{SF}{X} \right)) - \log(\tau_e^i \left( \frac{T}{K} \right)) \right| \right] / I \\ D_{\phi_1}(X, K) &= \left[ \sum_{i=1}^I \left| \log(\phi_1^i \left( \frac{SF}{X} \right)) - \log(\phi_1^i \left( \frac{T}{K} \right)) \right| \right] / I \\ D_{\tau_1}(X, K) &= \left[ \sum_{i=1}^I \left| \log(\tau_1^i \left( \frac{SF}{X} \right)) - \log(\tau_1^i \left( \frac{T}{K} \right)) \right| \right] / I \\ D_{\Phi(0)}(X, K) &= \left[ \sum_{i=1}^I \left| \log(\Phi(0)^i \left( \frac{SF}{X} \right)) / \Phi(0)^{\max} \left( \frac{SF}{X} \right) \right. \right. \\ &\quad \left. \left. - \log(\Phi(0)^i \left( \frac{T}{K} \right)) / \Phi(0)^{\max} \left( \frac{T}{K} \right) \right| \right] / I \end{aligned} \quad (9.2)$$

and K and X represent the syllable number of template (T) and the syllable in a sound field (SF), respectively. Also, i is the frame number of the running ACF, and I is the total frame number.

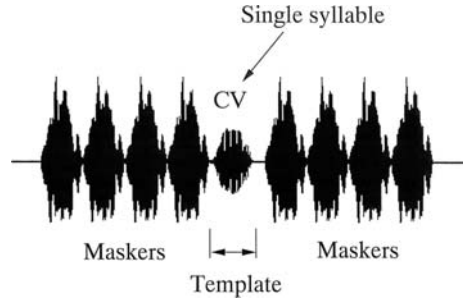
Let N be the total number of single syllables, then the percentile SI of the syllable K for one of four factors may be obtained as

$$SI_K(\psi) = 100N \exp \left( -\frac{d_K}{d_1} \dots \frac{d_K}{d_{K-1}} \frac{d_K}{d_{K+1}} \dots \frac{d_K}{d_N} \right) \quad (9.3)$$

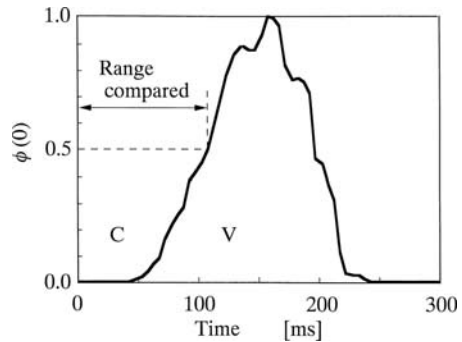
where  $\psi = \tau_e, \tau_1, \phi_1, \text{ or } \Phi(0)$ .

Let us now demonstrate an example for estimating the intelligibility of single syllables in the sound field, which consists of a direct sound and a single echo. The amplitude of the echo was the same as that of the direct sound, and the delay time of the single echo  $\Delta t_1$  was varied in the range between 0 and 480 ms. The test signal consisted of Japanese single syllables with maskers (Fig. 9.1). The masker was used to control the percentage of correctly identified signals over a wide range, thereby avoiding saturation effects when identification would otherwise approach 100%. The direct sound without maskers was used as a template. As shown in Fig. 9.2, the important initial parts of the normalized running ACF for the range of was analyzed.

**Fig. 9.1** An example of a single syllable with artificial nonmeaningful forward and backward maskers. The direct sound without maskers was used as a template. CV is a single syllable consisting of a consonant and a vowel



**Fig. 9.2** Relative sound energy of a single syllable as a function of time. The important former half part of the running ACF  $\phi(0) < 0.5$  of both a template and a test syllable was analyzed. C and V are parts of a consonant and a vowel, respectively



$$\Phi(0) > 0.5 \tag{9.4}$$

In the analyses, two distances for the direct sound and the echo were calculated with the four factors. One was the distance between the template and the direct sound with the maskers. The other was the distance between the template and the echo with the maskers. The shorter of two distances was selected in the calculation of syllable intelligibility. Using the distances given by Equation (9.2), we calculated the intelligibility of all syllables due to each factor by using Equation (9.3). The SI in total due to the four factors (see Section 5.2) is combined linearly, so that

$$SI = aSI(\tau_e) + bSI(\tau_1) + cSI(\phi_1) + dSI(\Phi(0)) \tag{9.5}$$

where a, b, c, and d are weighting coefficients signifying contributions of the four factors, which are determined so as to maximize SI. If the sound-pressure level is fixed at a constant and spatial factors are invariable, then the fourth term is eliminated and SI is the left hemisphere specialization.

Japanese syllables classified by the consonant and vowel categories of Table 9.1 were used because listeners recognize these phonetic elements in syllables with minimal confusion (Korenaga, 1997). The frontal loudspeaker in an anechoic chamber

**Table 9.1** Categorization of Japanese single syllables

## (a) Unvoiced consonant

		Consonant				
Vowel		K	S	T	H	P
Not contracted (Category A)	A	KA	SA	TA	HA	PA
	I	KI	SI	TI	HI	PI
	U	KU	SU	TU	HU	PU
	E	KE	SE	TE	HE	PE
	O	KO	SO	TO	HO	PO
Contracted (Category B)	YA	KYA	SYA	TYA	HYA	PYA
	YU	KYU	SHU	TYU	HYU	PYU
	YO	KYO	SHO	TYO	HYO	PYO

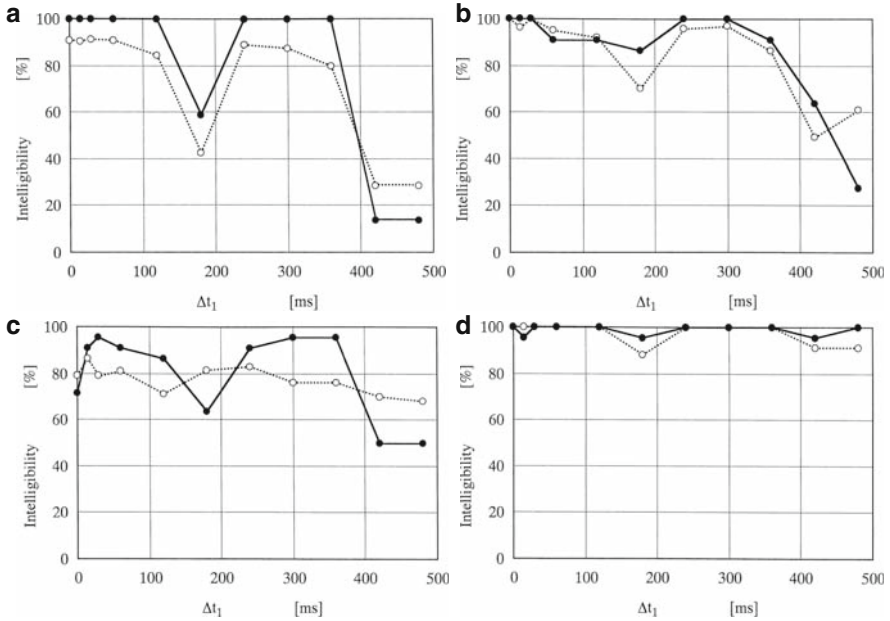
## (b) Voiced consonant

		Consonant								
Vowel		N	M	Y	R	W	G	Z	D	B
Not contracted (Category C)	A	NA	MA	YA	RA	WA	GA	ZA	DA	BA
	I	NI	MI	–	RI	–	GI	ZI	–	BI
	U	NU	MU	YU	RU	–	GU	ZU	–	BU
	E	NE	ME	–	RE	–	GE	ZE	DE	BE
	O	NO	MO	YO	RO	–	GO	ZO	DO	BO
Contracted (Category D)	YA	NYA	MYA	–	RYA	–	GYA	ZYA	–	BYA
	YU	NYU	MYU	–	RYU	–	GYU	ZYU	–	BYU
	YO	NYU	MYO	–	RYU	–	GYO	ZYO	–	BYO

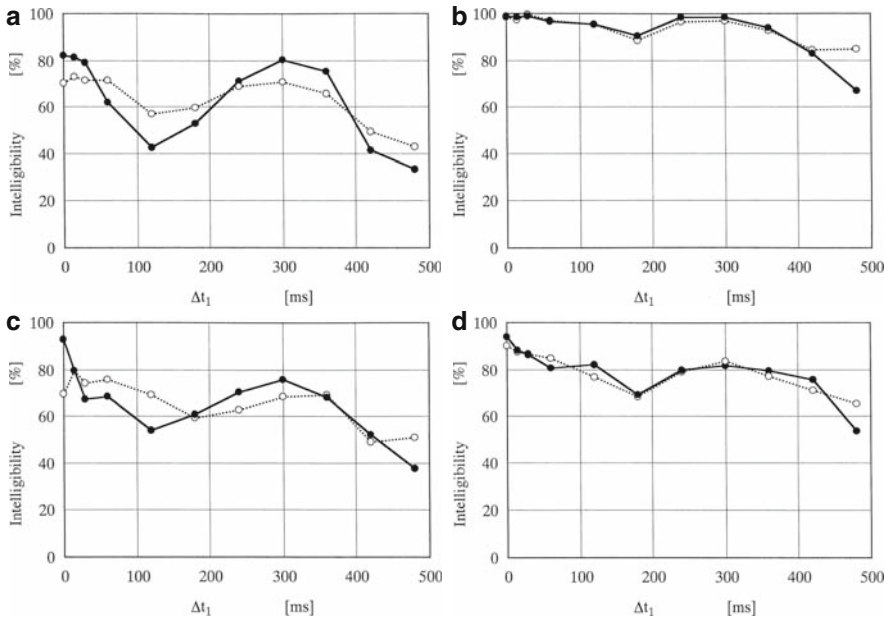
reproduced both of the direct sound and the echo as well as the noise masker. The running ACF with the integration interval  $2T = 30$  ms with the running step of 10 ms was analyzed. Twenty-one subjects participated in the experiments, so that about a 5% ( $= 1/21$ ) error might result if a single subject judged wrong.

Results of both calculated and tested intelligibility for a single syllable belonging to each category as a function of the delay time of echo are demonstrated in Fig. 9.3a–d. Averaged results for each category are shown in Fig. 9.4a–d, and averaged values for all syllables are shown in Fig. 9.5. Because we can hear the single syllables twice after the delay time of echo in the condition  $\Delta t_1 > 100$  ms, the local maxima appeared around 300 ms. In the larger delay range of the reflection of more than 300 ms, intelligibility is decreased due to effects of the masker. Even so, calculated values are in good agreement with the tested values. This implies that the four factors extracted from the ACF of the source signals and the sound-field signals are effective in identifying speech. Contributions of the four factor to the SI obtained by the multiple regressions are indicated in Table 9.2. It is found that the most significant factor of the four is the effective duration of the ACF,  $\tau_e$ .

It is concluded, therefore, that:



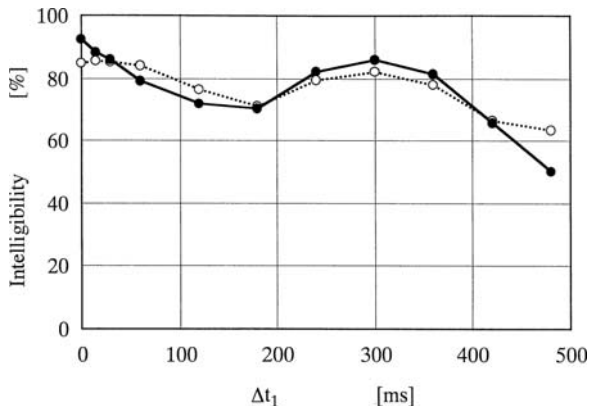
**Fig. 9.3** Examples of calculated and tested intelligibilities for the single syllable belonging to each category. (a) /ha/. (b) /be/. (c) /hya/. (d) /nyu/. Twenty-one subjects participated in the experiments



**Fig. 9.4** Averaged results of calculated and tested intelligibilities for each category. (a) Category A. (b) Category B. (c) Category C. (d) Category D



**Fig. 9.5** Mean values of calculated and tested intelligibilities for all syllables used



**Table 9.2** Contribution of each temporal factor extracted from the running ACF on speech intelligibility. Values were normalized by their maxima of four coefficients obtained from the multiple regression analysess

Consonant	Vowel	$\tau_e$	$\tau_1$	$\phi_1$	$\Phi(0)$
Unvoiced	A	0.50	0.60	0.36	1.00
	U	0.07	0.08	0.25	1.00
	E	1.00	0.66	0.23	0.46
	O	0.62	1.00	0.13	0.25
	YA	1.00	0.17	0.55	0.56
	YU	1.00	0.19	0.30	0.47
	YO	1.00	0.30	0.41	0.88
Voiced	A	0.74	0.92	0.89	1.00
	I	0.88	0.38	0.01	1.00
	U	1.00	0.80	0.30	0.29
	E	0.19	0.42	1.00	0.01
	O	1.00	0.25	0.09	0.80
	YA	1.00	0.17	0.55	0.56
	YU	1.00	0.19	0.30	0.47
YO	0.25	0.21	1.00	0.51	
					1.00

1. The speech identification in the sound field may be described by the four factors extracted from the running ACF of the target signal and the sound field with the signal.
2. The most significant factor is the effective duration of the ACF,  $\tau_e$ , which is intimately related to  $[\Delta t_1]_p$  and whose cortical response correlates are mainly associated with the left hemisphere (Table 5.1).

In this investigation another factor  $W_{\phi(0)}$  was not considered because only more recently, in 2007, was it identified as a factor related to timbre (Section 6.6). The factor  $W_{\phi(0)}$  is related to  $W_{IACC}$ , because both of them are determined by a signal's frequency composition. Once the factor  $W_{\phi(0)}$  is also taken into consideration, the present results may be explained more precisely.

## 9.2 Effects of Spatial Factors on Speech Reception

We are interested in the effect of sound fields on the interactions of sounds, and in particular how reverberant environments degrade speech sounds. In these experiments, a loudspeaker located in front of the listener presented single syllables, while continuous white noise as a disturbance was produced from another loudspeaker located at different horizontal angles. Three temporal factors and the sound energy were extracted from the ACF of the speech signal, and three spatial factors were extracted from the IACF. Results show that two factors had significant effects on syllable identification: the effective duration,  $(\tau_e)_{\min}$ , in the temporal factors extracted from the running ACF, and the  $W_{IACC}$  in the spatial factors extracted from the IACF.

In the previous section, we discussed how temporal factors extracted from the running ACF are related to speech intelligibility in sound fields with single echos. The auditory model was used to attempt to account for the identification of single syllables in noise disturbances from different directions (Ando and Yamasaki, unpublished). It is assumed that the specialization of the human cerebral hemisphere may relate to the highly independent contributions of spatial and temporal factors on speech identification. It may be the case that “cocktail party effects” might well be explained by such specialization of the human brain, because speech is mainly processed in the left hemisphere, while spatial information is independently processed in the right hemisphere at the same time. Based on such a model, we have described temporal and spatial sensations in Chapters 6 and 7, respectively. According to the model shown in Fig. 5.1 three temporal factors associated with the left hemisphere together with the sound energy were extracted from the ACF of the sound signal arriving at one of ear entrances. In addition, three spatial factors associated with the right hemisphere were extracted from the IACF of sound signals arriving at the two ear entrances. The running ACF and the running IACF with the integration interval  $2T = 30$  ms were analyzed using running steps of 10 ms.

For identification of the speech signals, psychological distances between characteristics of single syllables are calculated by Equation (9.2). The distance is a function of four factors extracted from the ACF, and these are mainly associated with neuronal responses from the left cerebral hemisphere. In addition, to find effects of off-direction noise, three spatial factors are extracted from the IACF, which are associated with the right cerebral hemisphere (Fig. 5.1). The distances due to the spatial factors,  $D_{IACC}$ ,  $D_{\tau IACC}$ , and  $D_{W_{IACC}}$ , respectively, are given by

$$\begin{aligned}
D_{IACC}(X,K) &= \left[ \sum_{i=1}^I |IACC^i \left( \begin{smallmatrix} SF \\ X \end{smallmatrix} \right) - IACC^i \left( \begin{smallmatrix} T \\ K \end{smallmatrix} \right)| \right] / I \\
D_{\tau_{IACC}}(X,K) &= \left[ \sum_{i=1}^I |\tau_{IACC}^i \left( \begin{smallmatrix} SF \\ X \end{smallmatrix} \right) - \tau_{IACC}^i \left( \begin{smallmatrix} T \\ K \end{smallmatrix} \right)| \right] / I \\
D_{W_{IACC}}(X,K) &= \left[ \sum_{i=1}^I |W_{IACC}^i \left( \begin{smallmatrix} SF \\ X \end{smallmatrix} \right) - W_{IACC}^i \left( \begin{smallmatrix} T \\ K \end{smallmatrix} \right)| \right] / I
\end{aligned} \tag{9.6}$$

In general, shorter distances between the template syllable and the syllables accompanied by noise signify higher intelligibilities. According to multiple regression analysis, the non-identification (NI) rate of syllables that were not matched with the template, has been directly calculated, so that

$$\begin{aligned}
NI(S_0, S_X) = S_L + S_R &= [aD\tau_e + bD\tau_1 + cD\phi_1]_L \\
&+ [dD\Phi(0) + eD_{IACC} + fD_{\tau_{IACC}} + gDW_{IACC}]_R
\end{aligned} \tag{9.7}$$

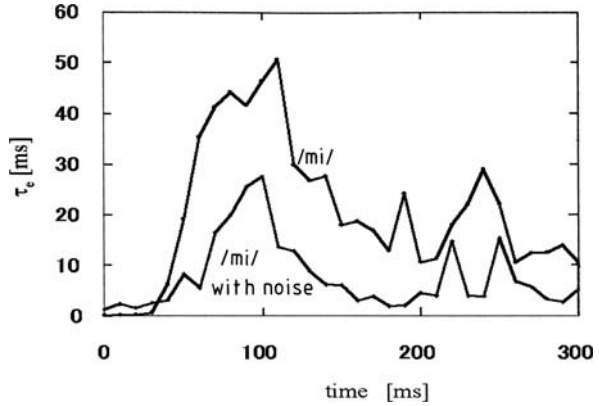
where  $S_L = [aD\tau_e + bD\tau_1 + cD\phi_1]_L$ ,  $S_R = [dD\Phi(0) + eD_{IACC} + fD_{\tau_{IACC}} + gDW_{IACC}]_R$ , and  $\Phi_p(0)$  is measured in dBA. The seven factors are classified into the left and right hemispheres by the model (Fig. 5.1). Note that the listening level or  $\Phi(0)$  is associated with the right hemisphere (Table 5.1). Weighting coefficients a through g in Equation (9.7) were determined by maximizing NI with experimental data.

Fourteen single syllables, /pa/ /pu/ /te/ /zo/ /bo/ /yo/ /mi/ /ne/ /kya/ /kyo/ /pya/ /gya/ /nya/ /zya/, with 4-s intervals between syllables, were presented to each subject by the frontal loudspeaker ( $\xi = 0^\circ$ , the distance to the center of the subject's head,  $d = 70 \text{ cm} \pm 1 \text{ cm}$ ) in an anechoic chamber. The white noise used as a disturbance was continuously produced by one of the loudspeakers located at different horizontal angles:  $\xi = 30^\circ, 60^\circ, 90^\circ, 120^\circ$ , or  $180^\circ$  ( $d = 70 \text{ cm}$ ). The sound-pressure level measured in terms of  $\Phi_p(0)$  of both speech signals and the continuous white noise were fixed at 65.0 dBA at the peak level. Ten subjects participated in the experiment, who were asked to identify what syllable was heard.

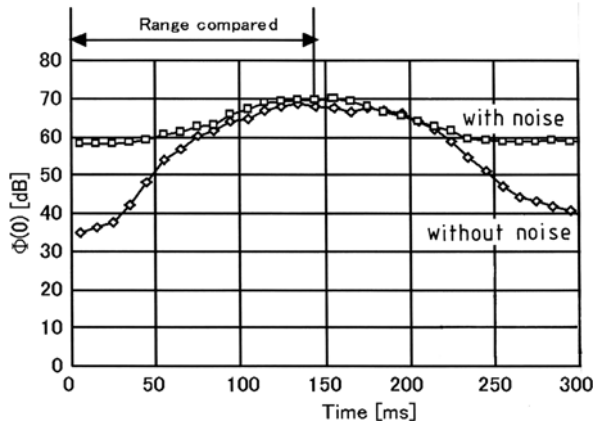
For example, values of  $\tau_e$  extracted from the running ACF for the signal /mi/ with and without the noise ( $\xi = 90^\circ$ ) as a function of time are shown in Fig. 9.6. The important initial half parts of the speech signal indicating  $\Phi(0) < 0.5$  as shown in Fig. 9.7 of both template and test syllables with the noise were applied in computation by Equations (9.6) and (9.7).

Results of the non-identification NI rate for some single syllables as a function of the horizontal angle  $\xi$  of the noise disturbance are shown in Fig. 9.8. Almost similar tendencies NI of these syllables were found. When the noise arrived from  $30^\circ$ , the NI indicated the maxima in the horizontal angle range tested, and when the noise was presented from  $120^\circ$ , it was the minima. The same was true for the averaged NI rate as shown in Fig. 9.9.

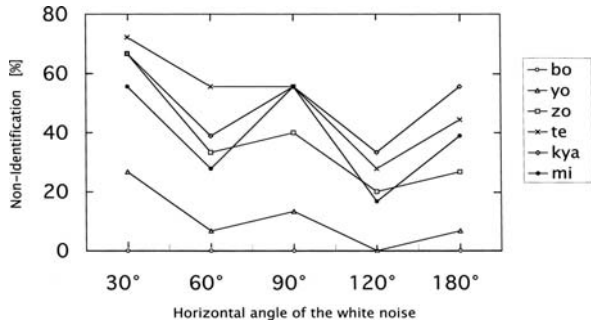
**Fig. 9.6** Values of effective duration  $\tau_e$  extracted from the running ACF for the frontal signal /mi/ only, and the /mi/ with the white noise from  $\xi = 90^\circ$

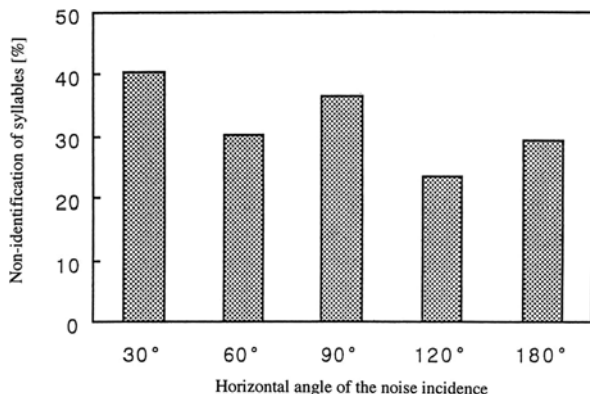


**Fig. 9.7** For making comparison, initial pieces analyzed of a frontal single syllable with and without the white noise from  $\xi = 90^\circ$



**Fig. 9.8** Examples of the percentage of non-identification (NI) for single syllables as a function of the horizontal angle of the white noise from different horizontal angles  $\xi$ . At the horizontal angle  $\xi = 120^\circ$ , the percentage of NI was minimum for the single syllables





**Fig. 9.9** Averaged percentile of nonidentified syllables with all single syllables tested obtained by the listening test for different angles  $\xi$  of white-noise incidence as a disturbance

Because the direct speech sound arrived from the frontal direction to the listener, the value of  $\tau_{IACC}$  is always close to zero being invariant. Thus, this factor was eliminated from the analysis by Equation (9.7) (Table 9.3). The minima of the psychological distance were always found for the noise disturbance from 120°, so that the NIs were minima. On the other hand, when the noise disturbance arrived from 30°, the distance due to  $\tau_e$  for all of the syllables commonly indicated the maxima in six factors.

**Table 9.3** Psychological distance calculated due to each of six factors

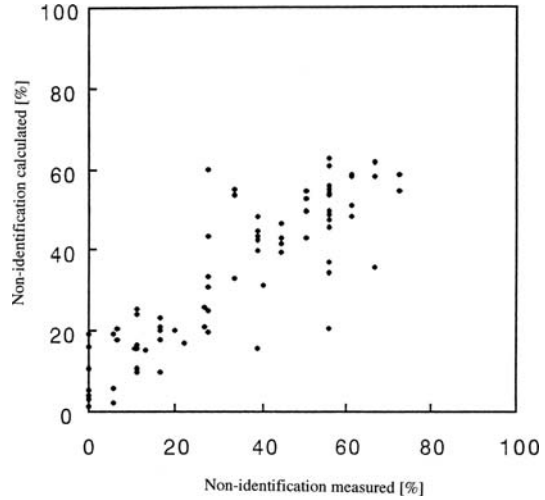
Horizontal angle of noise	$D\Phi(0)$	$D\tau_e$	$D\tau_1$	$D\phi_1$	$D_{IACC}$	$D_{W_{IACC}}$
30°	0.064	0.420	0.164	0.442	0.248	0.052
60°	0.056	0.351	0.247	0.355	0.266	0.049
90°	0.063	0.348	0.162	0.401	0.292	0.049
120°	0.058	0.279	0.157	0.376	0.270	0.043
180°	0.074	0.383	0.171	0.494	0.247	0.071

The weighting coefficients in Equation (9.7) for the six factors are listed in Table 9.4 . According to the weighting coefficients obtained here, the factors  $\tau_e$  and  $W_{IACC}$  contributed significantly to the NI. For each single syllable, the relationship between the calculated values by Equation (9.7) and the measured values are shown in Fig. 9.10. Obviously, the linear relationship was achieved ( $r = 0.86, p < 0.01$ ).

**Table 9.4** Weighting coefficients determined

	$\Phi(0)$	$\tau_e$	$\tau_1$	$\phi_1$	IACC	$W_{IACC}$
Coefficient	0.053	0.335	0.028	0.136	0.086	0.384

**Fig. 9.10** Relationship between the calculated percentile of nonidentified syllables and that obtained by listening tests ( $r = 0.86$ ,  $p < 0.01$ )



A remarkable finding was that the most significant factor in the previous section was the effective duration of the ACF,  $\tau_e$ . In this study also, the most effective and significant temporal factor was the  $\tau_e$  value in the temporal factors. In order to obtain effects of the different direction of the noise disturbance, the spatial factors may be taken into consideration. Conclusions are as follows:

1. The syllable identification (NI) may be calculated by both temporal factors extracted from the ACF and spatial factors extracted from the IACF.
2. Particularly in the condition of this experiment, the value of  $\tau_e$  as the temporal factor is the most significant as is similar to previous results (Ando et al., 1999); in addition, the  $W_{IACC}$  in the spatial factor contributes significantly to the speech identification.

### 9.3 Effects of Sound Fields on Perceptual Dissimilarity

Perceptual dissimilarity is the perceived difference between the same sounds when they are either produced or heard at different locations. In this section, we discuss dissimilarity for the real sound field of an existing hall in relation to all of the temporal and spatial factors extracted from the ACF and the IACF, respectively. To incorporate dissimilarity into the model, two temporal factors ( $\Delta t_1$  and  $T_{sub}$ ) of the sound field are added to these temporal and spatial factors. At a given fixed seating position in a real room, the overall psychological distance between sound fields changing source locations on the stage can be obtained by keeping other sensory effects in a room constant. We shall show that the scale value of dissimilarity may be described by the linear combination of all the temporal and spatial factors of the sound field.

Previously, Yamaguchi (1972) carried out an experiment to empirically measure the perceptual dissimilarity between the same sounds heard from different seats in an existing concert hall. He demonstrated that two significant factors associated with listening position affected subjective similarity: the sound-pressure level and the reverberation characteristics. Edward (1974) also tested dissimilarity by studying differences between different halls and reported that the early-echo pattern, the reverberation time, and the volume level were the significant factors. Cocchi et al. (1990) and Sato et al. (1997) confirmed the effectiveness of the theory of subjective preference through investigations in existing concert halls. Sato et al. (2002) reconfirmed the effectiveness of the theory in an existing opera house as well.

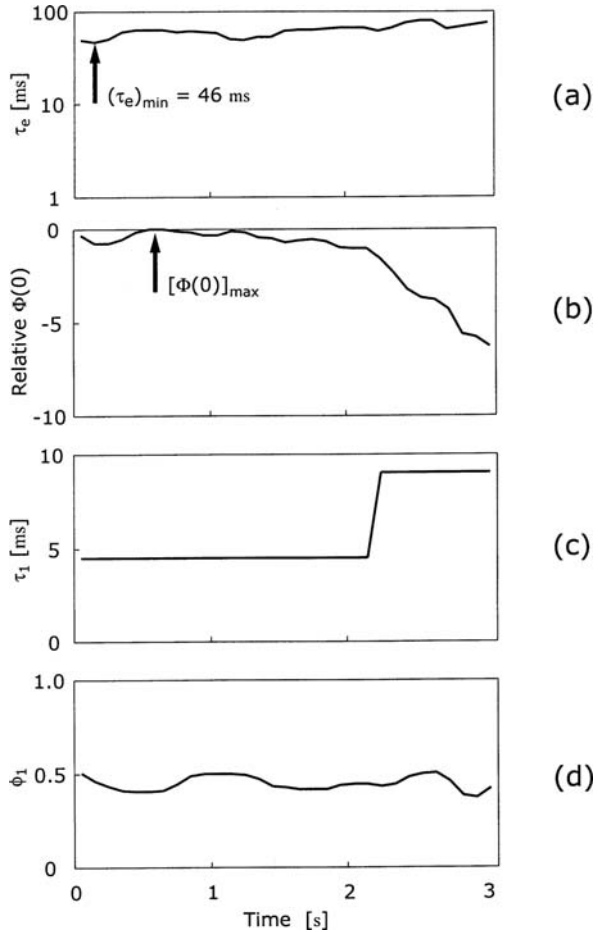
In the current study, a reverberation-free, 4-s fragment of recorded orchestral music (*Water Music*, Suite No. 2 – Alla Hornpipe, composed by Handel) was used as a source signal (Hotehama et al., 2002). The music source was characterized in terms of the running ACF of the source signal after passing through an A-weighted network. An ACF analysis was carried out with an integration interval  $2T = 2.0$  s and running step-size of 100 ms, and factors  $\Phi(0)$ ,  $\tau_e$ ,  $\tau_1$ , and  $\phi_1$  were extracted. As shown in Fig. 9.11, the minimum value of the effective duration of the source signal,  $(\tau_e)_{\min} = 46$  ms. This value is obtained at the most active piece of music and thus strongly related to preferred values of the temporal factors ( $\Delta t_1$  and  $T_{\text{sub}}$ ) of the sound field as discussed in Chapter 3 (Ando et al. 1989; Mouri et al., 2000).

Dissimilarity judgments were performed in a multiple-purpose hall, the 400-seat ORBIS Hall in Kobe, Japan, which is shown in Fig. 9.12. Six loudspeakers with identical characteristics were placed on the stage. Twenty student subjects participated in the experiment. They were divided into four groups and seated at specific positions in seating locations A, B, C, and D. To avoid the effects of other environmental conditions, each dissimilarity judgment was conducted at a fixed seat, and sound sources only were switched between the six source locations. Subjects judged difference as an overall impression between the paired stimuli. They were asked to rate the pairs on a subjective linear scale that had two extreme ends: “no different,” and “extremely different.” The judgment was made for 15 pairs of the six sound fields at each listener’s location. The interval between paired stimuli was 1 s. Each pair of sound fields was separated by a silent interval of 5 s, and the pairs were arranged in random order. Each session was repeated 5 times. To obtain the scale value of dissimilarity between sound fields, the original data of dissimilarity judgment were categorized into seven categories by the method of successive categories (Torgerson, 1958). The scale value of dissimilarity for each pair of source locations at the seat positions was obtained with all listeners as listed in Table 9.5.

To measure the acoustical factors extracted from the ACF and IACF, the music signal used in the dissimilarity judgments was reproduced from each loudspeaker. The signal was recorded at each listening position, through two microphones placed at the two ear entrances of a real person facing the center of the stage. To obtain binaural impulse responses, a maximum-length sequence (MLS) signal was reproduced from each loudspeaker (Alrutz, 1981).

Considering the fact that dissimilarity as well as subjective preference judgments may be made at the most “active and informative” running music piece, which indi-

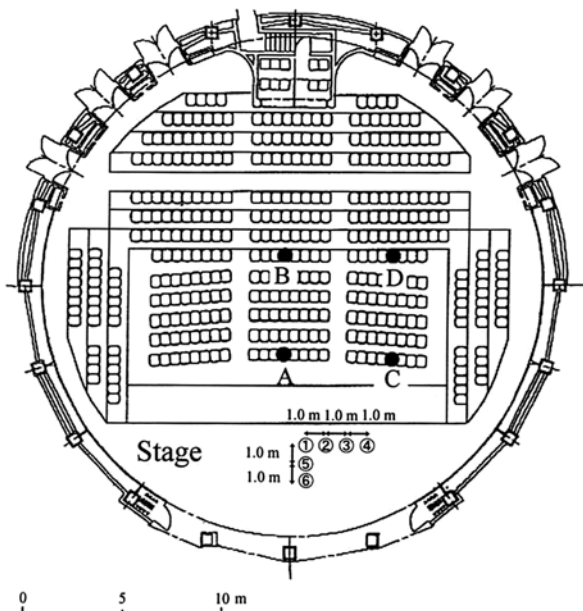
**Fig. 9.11** Factors of the running ACF of the source signal analyzed by  $2T = 2.0$  s with 100 ms of the running interval. (a)  $\tau_e$ . (b) Relative  $\Phi(0)$ , obtained as relative to the maximum value at  $\tau = 0.5$  s. (c)  $\tau_1$ . (d)  $\phi_1$



icates the minimum value of  $(\tau_e)_{\min}$  extracted from the running ACF, values of  $\tau_1$  and  $\phi_1$  at the particular piece also were extracted. Values of LL, IACC,  $\tau_{IACC}$ , and  $W_{IACC}$  from the running IACF were also computed.

After obtaining the binaural impulse responses, values of  $\Delta t_1$  and  $T_{\text{sub}}$  were calculated. The value of  $\Delta t_1$  was defined by the time difference between the arrival time of the direct sound and that of the reflection, which is the maximum energy in the impulse responses. From the two measured values of  $\Delta t_1$  obtained at both ears, the one with the largest amplitude of the first reflection was selected as the  $\Delta t_1$  (Ando and Gottlob, 1979). The averaged value of  $T_{\text{sub}}$  of the 500-Hz and 1-kHz octave band center frequencies were applied here, because these frequency ranges are the dominant of the source signal. The measured temporal and spatial factors obtained by running the ACF, IACF, and binaural impulse response analysis are shown in Fig. 9.13. The factors extracted from the IACF were also chosen from

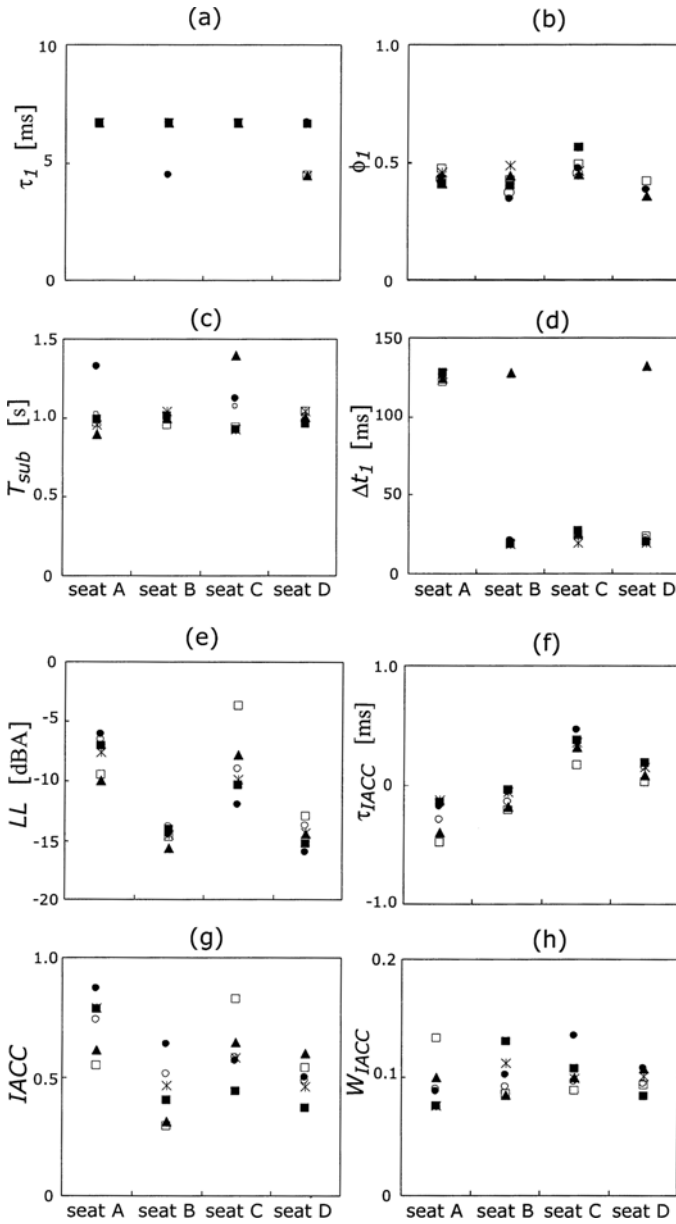




**Fig. 9.12** Plan of the ORBIS Hall in which dissimilarity judgment was made. A–D: Locations of listeners. ①–⑥: Source locations changed in the PCT conducted at fixed seating position during the judgment

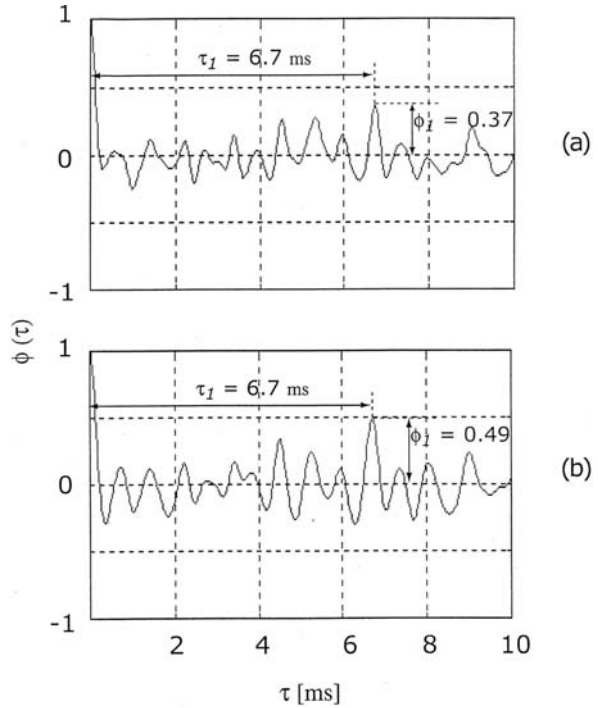
**Table 9.5** Scale values of dissimilarity judgments for each pair of source locations at seating positions of A, B, C, and D

Pair of source locations	Seat position			
	Position A	Position B	Position C	Position D
1–2	1.4	0.8	0.8	0.9
1–3	1.9	1.7	1.8	1.6
1–4	2.3	2.0	2.5	2.4
1–5	0.7	0.9	0.8	0.7
1–6	1.3	1.0	1.4	1.0
2–3	0.8	0.6	1.2	1.3
2–4	1.2	1.2	2.2	2.0
2–5	1.6	1.2	0.7	0.9
2–6	1.8	1.4	0.8	0.6
3–4	0.5	0.4	1.6	1.2
3–5	2.1	1.9	1.8	1.7
3–6	2.1	2.0	1.4	1.7
4–5	2.4	2.0	2.2	2.1
4–6	2.2	2.1	2.0	2.0
5–6	0.4	0.4	0.9	0.8



**Fig. 9.13** Measured physical factors at each listener’s locations measured. The location of sources and listeners are shown in Fig. 9.12. (a)  $\tau_1$ . (b)  $\phi_I$ . (c)  $T_{sub}$ . (d)  $\Delta t_1$ . (e) LL. (f)  $\tau_{IACC}$ . (g) IACC. (h)  $W_{IACC}$ . ●, values measured for source location ①; ○, values measured for source location ②; ▲, values measured for source location ③; □, values measured for source location ④; ■, values measured for source location ⑤; and \*, values measured for source location ⑥

**Fig. 9.14** Examples of the ACF analyzed. The locations of sources and listeners are shown in fig. 9.12. (a) Source location ② at seat position B. (b) Source location ⑥ at seat position B



the short time interval centered on the time, when  $(\tau_e)_{\min}$  of the source signal was obtained. Figure 9.14 illustrates examples of the running ACF of source locations 2 and 6 at seat position B. A difference can be observed in the measured ACF due to the different transmission characteristics of the sound field.

Next, in order to find a relationship between the scale value and physical factors obtained by the measurement, a multiple regression analysis was made. The perceptual distance between the sound fields of *a* and *b* with respect to each factor was estimated in the following manner.

### 9.3.1 Perceptual Distance due to Temporal Factors

$$D_{\tau_1} = |\log(\tau_1)^a - \log(\tau_1)^b| \quad (9.8)$$

$$D_{\phi_1} = |\log(\phi_1)^a - \log(\phi_1)^b| \quad (9.9)$$

$$D_{\Delta t_1} = \left| \log \left( \frac{\Delta t_1}{[\Delta t_1]_p} \right)^a - \log \left( \frac{\Delta t_1}{[\Delta t_1]_p} \right)^b \right| \quad (9.10)$$

$$D_{T_{sub}} = \left| \log \left( \frac{T_{sub}}{[T_{sub}]_p} \right)^a - \log \left( \frac{T_{sub}}{[T_{sub}]_p} \right)^b \right| \tag{9.11}$$

where  $D_{\Delta t_1}$  and  $D_{T_{sub}}$  are the distances due to the normalized values with the most preferred  $[\Delta t_1]_p$  and  $[T_{sub}]_p$ , respectively. These preferred values are calculated by Equations (3.4) and (3.6) using  $(\tau_e)_{min}$  instead of  $(\tau_e)$ . The distances of temporal factors  $D_{\tau_1}$ ,  $D_{\phi_1}$ ,  $D_{\Delta t_1}$ , and  $D_{T_{sub}}$  were calculated using logarithmic values.

### 9.3.2 Perceptual Distance due to Spatial Factors

$$D_{LL} = |(LL)^a - (LL)^b| \tag{9.12}$$

$$D_{IACC} = |(IACC)^a - (IACC)^b| \tag{9.13}$$

$$D_{\tau_{IACC}} = |(\tau_{IACC})^a - (\tau_{IACC})^b| \tag{9.14}$$

$$D_{W_{IACC}} = |(W_{IACC})^a - (W_{IACC})^b| \tag{9.15}$$

In the multiple regression analysis, the distance of dissimilarity for multiple physical factors is combined linearly, so that the total distance is given by

$$D = D_L + D_R = a'D_{LL} + b'D_{\tau_1} + c'D_{\phi_1} + d'D_{IACC} + e'D_{\tau_{IACC}} + f'D_{W_{IACC}} + g'D_{\Delta t_1} + h'D_{T_{sub}} \tag{9.16}$$

where  $D_L = b'D_{\tau_1} + c'D_{\phi_1} + g'D_{\Delta t_1} + h'D_{T_{sub}}$ ,  $D_R = a'D_{LL} + d'D_{IACC} + e'D_{\tau_{IACC}} + f'D_{W_{IACC}}$  and  $a'$ ,  $b'$ ,  $c'$ ,  $d'$ ,  $e'$ ,  $f'$ ,  $g'$ , and  $h'$  are coefficients, which may be obtained by a stepwise regression method.

Prior to the multiple regression analysis, correlation coefficients between factors were figured out as listed in Table 9.6. Concerning the value of  $W_{IACC}$ , it is a significant factor for determining the ASW, if source signals with different fre-

**Table 9.6** Correlation coefficients between physical factors obtained by the acoustic measurements

	$D_{LL}$	$D_{\tau_1}$	$D_{\phi_1}$	$D_{IACC}$	$D_{\tau_{IACC}}$	$D_{W_{IACC}}$	$D_{\Delta t_1}$	$D_{T_{sub}}$
$D_{LL}$	1.00	-0.26*	-0.30*	0.41**	0.56**	0.21	-0.10	0.28*
$D_{\tau_1}$		1.00	0.42**	0.08	-0.18	-0.23	0.13	-0.34**
$D_{\phi_1}$			1.00	0.38**	-0.28*	-0.04	0.23	-0.29*
$D_{IACC}$				1.00	0.54**	0.26	0.15	0.03
$D_{\tau_{IACC}}$					1.00	0.59**	-0.05	0.04
$D_{W_{IACC}}$						1.00	-0.02	-0.11
$D_{\Delta t_1}$							1.00	-0.25
$D_{T_{sub}}$								1.00

\*\*p < 0.01; \*p < 0.05.

quency ranges are applied (Ando et al., 1999). However, it was eliminated from the analysis, due to the fact that the single source signal was used in this experiment. The same is true for the factor  $W_{\phi(0)}$ , fortunately. Results of the table show that  $D_{W_{IACC}}$ ,  $D_{LL}$  and  $D_{IACC}$  highly correlated with  $D_{\tau_{IACC}}$  (correlation coefficients with were 0.59, 0.56, and 0.54, respectively). Thus,  $\tau_1$ ,  $\phi_1$ ,  $\tau_{IACC}$ ,  $\Delta t_1$ , and  $T_{sub}$  were selected as a representative of these factors. The resulting distance of dissimilarity  $D$  is given by,

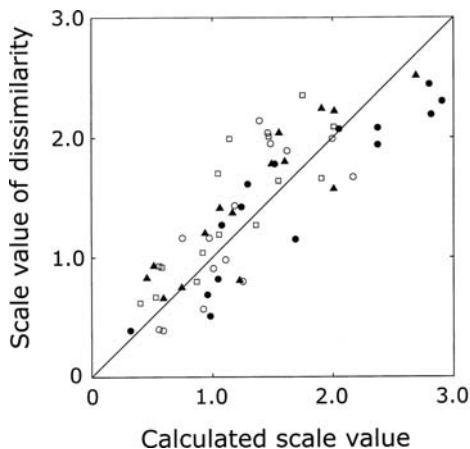
$$D \approx D_L + D_R = aD_{\tau_1} + bD_{\phi_1} + cD_{\tau_{IACC}} + dD_{\Delta t_1} + eD_{T_{sub}} \tag{9.17}$$

where  $D_L = aD_{\tau_1} + bD_{\phi_1} + dD_{\Delta t_1} + eD_{T_{sub}}$ ,  $D_R = cD_{\tau_{IACC}}$ , and coefficients obtained are  $a \approx 1.91$ ,  $b \approx 3.37$ ,  $c \approx 7.59$ ,  $d \approx 0.37$ , and  $e \approx 3.90$  (Table 9.7).

Figure 9.15 shows the relationship between measured scale values of dissimilarity obtained at each seat position and calculated values of dissimilarity. The correlation coefficients between them at each seat position were 0.92 ( $p < 0.01$ ) at seat position A, 0.79 ( $p < 0.01$ ) at seat position B, 0.90 ( $p < 0.01$ ) at seat position C, and 0.84 ( $p < 0.01$ ) at seat position D. The total correlation coefficient between scale

**Table 9.7** Partial regression coefficients for significant factors obtained by multiple regression analysis with normalized partial regression coefficients

	$D_{\tau_1}$	$D_{\phi_1}$	$D_{\tau_{IACC}}$	$D_{\Delta t_1}$	$D_{T_{sub}}$
Normalized partial coefficients	0.10	0.15	0.69	0.08	0.17
p value	<0.02	<0.01	<0.01	<0.01	<0.05



**Fig. 9.15** Relationships between calculated scale values by Equation (9.17) and scale values of dissimilarity judgments at each seat position ( $r = 0.84$ ;  $p < 0.01$ ). The locations of listeners are shown in Fig. 9.12. ●, values obtained at seat position A ( $r = 0.92$ ;  $p < 0.01$ ); ○, values obtained at seat position B ( $r = 0.79$ ;  $p < 0.01$ ); ▲, values obtained at seat position C ( $r = 0.90$ ;  $p < 0.01$ ); □, values obtained at seat position D ( $r = 0.84$ ;  $p < 0.01$ )

values of dissimilarity and calculated values of dissimilarity for all seats was 0.84 ( $p < 0.01$ ).

In summary, the significant factors that influenced dissimilarity judgment in the existing hall were:

1. Temporal factors  $\tau_1$  and  $\phi_1$  extracted from the ACF at the minimum effective duration  $(\tau_e)_{\min}$  of the signal. These factors correspond to the percepts of pitch and pitch salience.
2. The spatial factor  $\tau_{\text{IACC}}$  extracted from the IACF at  $(\tau_e)_{\min}$  that corresponds to the perception of spatial diffuseness and envelopment.
3. Temporal factors of the sound field,  $\Delta t_1$  and  $T_{\text{sub}}$  i.e. the times of early reflections and later reverberations, respectively.

# Chapter 10

## Applications (III) – Noise Measurement

The central auditory signal processing model lends itself to a wide range of applications. In this chapter, we will first discuss a method of measuring identification and subjective evaluation of environmental noise. Then, examples of noise measurement are discussed in terms of both the temporal factors extracted from the ACF of the source signal and spatial factors extracted from the IACF.

### 10.1 Method of Noise Measurement

A method of measuring the noise for its identification and subjective evaluations of the noise based on the model of the central auditory system are described in this section (Ando, 2001a; Ando and Pompoli, 2002). As is discussed in Chapters 5 through 7, temporal sensations are described by the three temporal factors extracted from the ACF, and spatial sensations are described by the four spatial factors from the IACF. Since the temporal and spatial factors may be dominantly processed in the left and right hemisphere, respectively, the two main factors may independently contribute to judgments of any subjective responses of noise. Thus, any of subjective evaluation  $S$  of a noise signal may be expressed essentially by Equation (6.6), so that,

$$S = S_L + S_R = f_L(\tau_1, \phi_1, \tau_e) + f_R(LL, IACC, W_{IACC}, \tau_{IACC}) \quad (10.1)$$

where  $S_L = f_L(\tau_1, \phi_1, \tau_e)$  and  $S_R = f_R(LL, IACC, W_{IACC}, \tau_{IACC})$ ; both temporal factors and spatial factors are defined in Chapter 5 and formulated in Section 6.1. As indicated in Table 10.1a, temporal sensations and spatial sensations and identification of noise sources are described by these measured temporal and spatial factors, respectively. When  $W_{IACC}$  is taken into account, then the factor  $W_{\phi(0)}$  may be eliminated, due to a large correlation between them. Computer software for the identification of noise sources and noise measurement has been discussed based on Equation (10.1) (Sakurai et al., 2001; Sakai et al., 2001; Sakai et al., 2002; Fujii et al., 2004a). It is worth noting that the previous envelope-based method for the sound level meter (SLM) often does not well describe subjective responses (Table 10.1b).

**Table 10.1** Comparison between the previous envelope-based method measured by use of the sound-level meter and the correlation method proposed here

(a) Physical factors to be measured

Temporal and spatial factors	Envelope method	Correlations method
(1) SPL	Envelope SPL (dBA)	$\Phi_{ll}(0), \Phi_{rr}(0), LL$ , or binaural SPL (dBA)
(2) Running temporal window (2T)	Fast, slow, and peak hold	Adaptive; it is about $30(\tau_e)_{min}$
(3) $\tau_1$	Impossible	Measurable
(4) $\phi_1$	Impossible	Measurable
(5) $\tau_e$ and $(\tau_e)_{min}$	Impossible	Measurable
(6) IACC	Impossible	Measurable
(7) $W_{IACC}$	Impossible	Measurable
(8) $\tau_{IACC}$	Impossible	Measurable

(b) Temporal and spatial sensations to be related to identify and evaluate the noise source

Subjective responses	Envelope method	Correlations method
(1) Loudness	Partially possible	$f_L(\tau_e, \tau_1, \phi_1)$ at constant LL
(2) Pitch or missing fundamental	Impossible	$f_L(\tau_1, \phi_1)$
(3) Duration sensations		$f_L(\tau_e, \tau_1, \phi_1)$
(4) Timbre	Impossible	$f_L(W_{\phi(0)})$
(5) Localization in horizontal plane	Impossible	$f_R(\Phi_{ll}(0), \Phi_{rr}(0), \tau_{IACC}, IACC, W_{IACC})$
(6) ASW	Impossible	$f_R(LL, W_{IACC}, \tau_{IACC}, IACC)$
(7) Subjective diffuseness	Impossible	$f_R(\Phi_{ll}(0), \Phi_{rr}(0), IACC, W_{IACC})$
(8) Annoyance	Impossible	$f_L(\tau_e, \tau_1, \phi_1) + f_R(\Phi_{ll}(0), \Phi_{rr}(0), \tau_{IACC}, IACC, W_{IACC})$
(9) Identification of noise sources	Impossible	$f_L(\tau_e, \tau_1, \phi_1, W_{\phi(0)}) + f_R(\Phi_{ll}(0), \Phi_{rr}(0), \tau_{IACC}, IACC, W_{IACC})$

## 10.2 Aircraft Noise

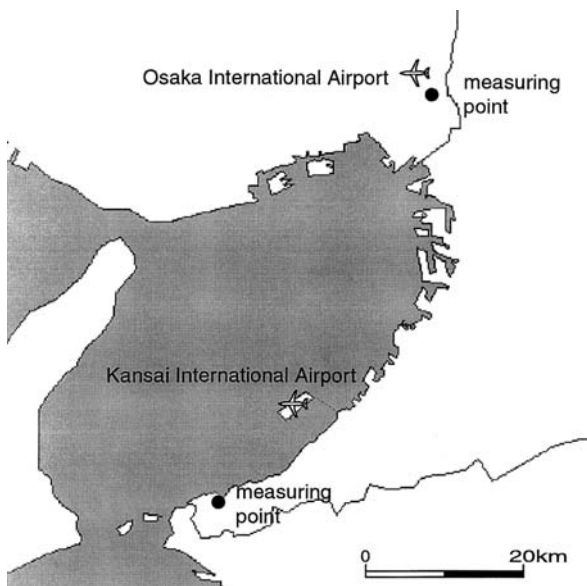
The acoustic properties of aircraft noise were investigated by means of temporal and spatial factors in a real field. As is described in Chapter 5, from the ACF analysis (1) the sound energy  $\Phi(0)$ , (2) the effective duration of ACF,  $(\tau_e)_{min}$ , (3) the delay time of the first peak,  $\tau_1$ , and (4) its amplitude  $\phi_1$  were extracted. From the IACF analysis, three spatial factors are extracted, (1) the magnitude of the interaural crosscorrelation, IACC, (2) the interaural delay time at IACC,  $\tau_{IACC}$ , and (3) the width of the maximum peak of the IACF,  $W_{IACC}$ , as well as the binaural listening level, LL.

This section describes the acoustic properties of aircraft noise in terms of both temporal and spatial factors (Fujii et al., 2001). Aircraft noise causes serious problems such as hearing loss and also has serious effects on the growth of unborn



babies, infants, and children as discussed later. Much effort has been made in noise research and noise reduction technologies and methods to reduce noise levels. However, it is likely that perceived acoustic properties have not been established sufficiently. In particular, the relationship between physical properties and psychological effects is still unclear. For example, even if a sound exists that has an SPL below standards such as the equivalent sound level ( $L_{eq}$ ), the effective perceived noise level (EPNL), and the noise and number index (NNI), it can be perceived as much more noisy than expected in a given situation. Such an annoyance may be related to both temporal and spatial factors associated with the left and right cerebral hemispheres, respectively.

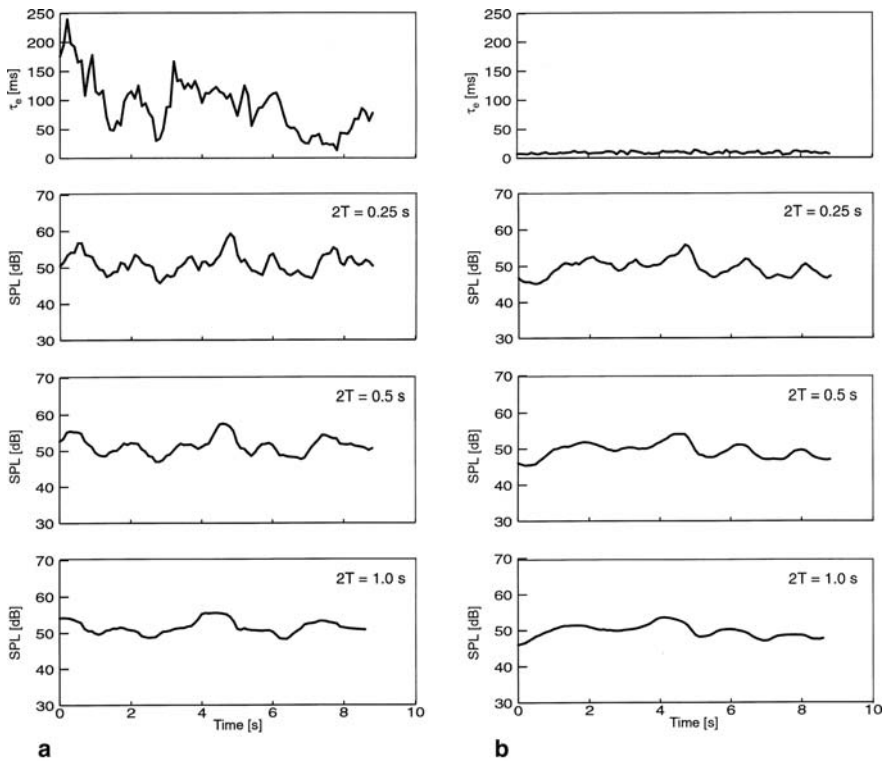
Measurements were taken outdoors near the Osaka International Airport (OIA) in 1999 and near the flight course of the Kansai International Airport (KIA) in 2000. Locations of the measurement are illustrated in Fig. 10.1. At the OIA, two locations were chosen close to the runway to measure the noise from aircraft landing and taking off. The distance between the runway and each measuring point was about 100 m. The ambient noise level in this area was higher because of road traffic ( $60 \pm 2$  dB). It was cloudy and windless at ground level. The temperature was about  $10^\circ\text{C}$  during the measurement. For measurement of noise along the flight course of the KIA, a dummy head was set near the coast. This location is 20 km southwest of the airport, and the flight course for landing is about 1.0 km from the shore. The altitude of the plane used in the measurement was about 1.0 km above sea level, according to flight data reported from the airport. It was cloudy and windless at ground level during the measurement also. The average temperature for the day was  $12^\circ\text{C}$ . The ambient noise level in this area was  $43 \pm 2$  dB. Noise signals were recorded through two half-inch condenser microphones set at both ear positions of



**Fig. 10.1** Locations of two international airports and the measurement points

a sphere representing a human head. This dummy head was made of 20-mm-thick Styrofoam with a diameter of 200 mm of the sphere. Microphones were set at 1.5 m above the ground.

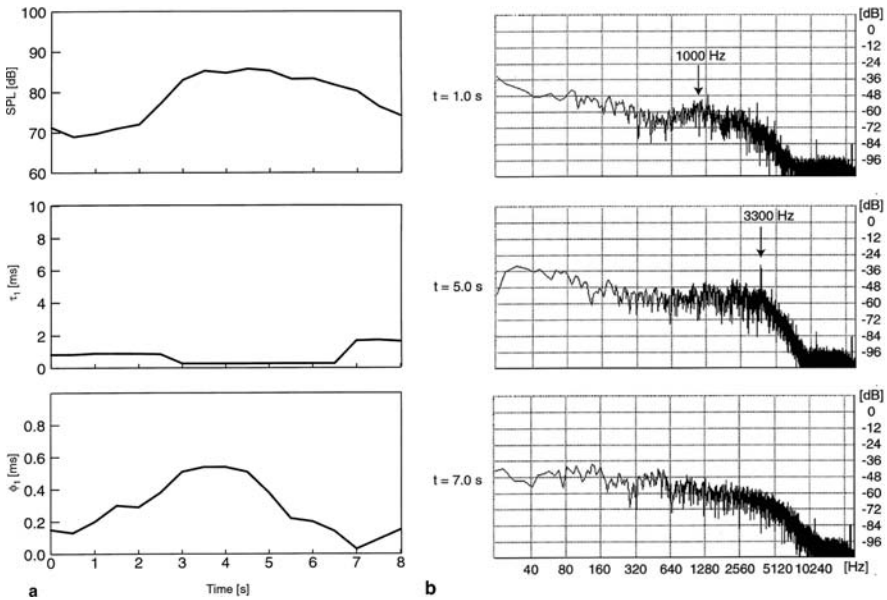
The duration of the noise depended on the location of the aircraft, its speed, and the distance to the receiving position. The values of  $(\tau_e)_{\min}$  extracted from its ACF were between 10 ms and 20 ms. Loudness variation was matched with SPL variations when it was analyzed by  $2T = 0.25$  s or 0.5 s. Thus, it was reconfirmed that the integration interval recommended be  $2T \approx 30 (\tau_e)_{\min}$  as discussed in Section 5.3. In the current study, the integration interval was chosen as 0.25 s for signals with  $(\tau_e)_{\min} \approx 10$  ms and as 0.5 s for signals with  $(\tau_e)_{\min} \approx 20$  ms. As is shown in Fig. 10.2, when  $2T$  was 1 s, it was too long to capture the fluctuation of loudness, but a finer variation of loudness could be observed when  $2T = 0.25$  s.



**Fig. 10.2** Effective duration and temporal integration windows for measurement of aircraft noise. Examples of sound pressure level (SPL) obtained by the geometric mean of  $\Phi_{ll}(0)$  and  $\Phi_{rr}(0)$  with a flat filter at the two ears for two different types of noise signals. (a) Noise with  $(\tau_e)_{\min} = 20$  ms ( $2T = 0.5$  s). (b) Noise with  $(\tau_e)_{\min} = 10$  ms ( $2T = 0.25$  s). The SPL measured by three different integration intervals  $2T = 0.25$ , 0.5, and 1.0 s is shown as a function of time

Typically aircraft produce predominantly high frequency noise while approaching and predominantly low frequency noise after they pass overhead and recede in the distance. Such acoustic characteristics of planes landing are clearly represented

by factors derived from the ACF (Fig. 10.3 (a)). Measured SPL (= LL) obtained by the geometric mean of  $\Phi_{ll}(0)$  and  $\Phi_{rr}(0)$  is shown as a function of time up to 10 s during higher levels than those of other ambient noises. Around when  $t = 5.0$  s, the aircraft had just flown overhead. The delay time and the amplitude of the first peak in ACF,  $\tau_1$  and  $\phi_1$ , represent the perceived pitch and its strength (see Section 6.2). Results indicate that the perceived pitch varied throughout the flight. A strong tonal component was observed, as the aircraft approached with the value of  $\phi_1$  being increased up to about 0.5,  $\tau_1 \approx 1.0$  ms (1000 Hz). The strongest pitch of 3300 Hz was perceived when the aircraft passed just overhead, at which the value of  $\tau_1$  was 0.3 ms. After the aircraft passed over, the  $\tau_1$  value increased and the  $\phi_1$  value decreased simultaneously due to the Doppler effect after  $t > 6.5$  s, indicating that the noise was dominated by the lower-frequency components. The power spectrum was measured after passing through without any filters, and the ACF was measured after passing through the A-weighting network at  $t = 1.0, 5.0,$  and  $7.0$  s, and these are illustrated in Fig. 10.3b and c, respectively. These show that  $\tau_1$  and  $\phi_1$  represent the pitch properties of aircraft noise clearly; at  $t = 1.0$  s, there is a weak peak at  $\tau_1 = 1$  ms representing a weak pitch of around 1000 Hz, but this information may not be found in the spectrum. At  $t = 5.0$  s, there is a high-frequency component at  $\tau_1 = 0.3$  ms or 3300 Hz perceived as a tonal sound; and at  $t = 7.0$  s, such a strong peak disappeared and the lower-frequency components increase below 500 Hz, which is perceived like the white noise in the low-frequency range.



**Fig. 10.3** Measured factors extracted from the ACF of aircraft noise. (a) The SPL, the value of  $\tau_1$ , and the value  $\phi_1$  for landing aircraft as a function of time. (b) The power spectra at  $t = 1.0, 5.0,$  and  $7.0$  s. (c) The normalized autocorrelation function NACF at  $t = 1.0, 5.0,$  and  $7.0$  s

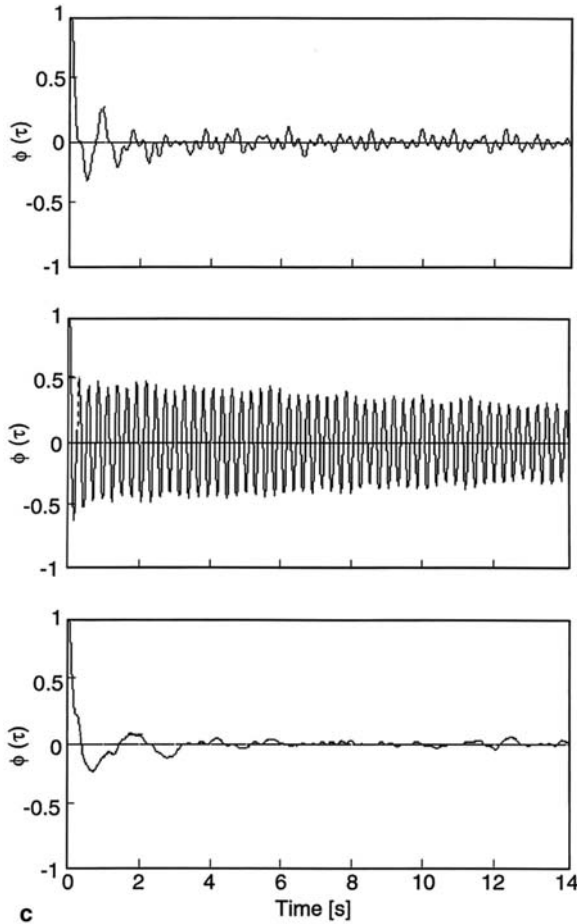
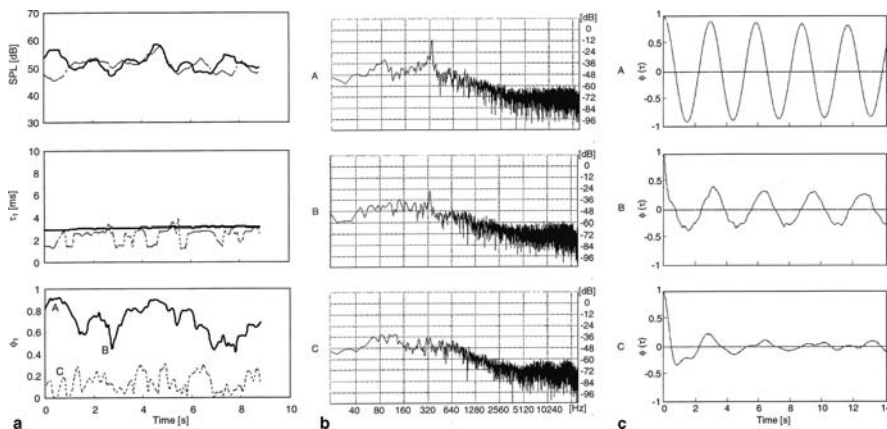


Fig. 10.3 (continued)

Figure 10.4a shows the measured three factors extracted from the ACF for the aircraft during level flying overhead at an altitude of about 1 km. Two typical examples of tonal and nontonal noise analyzed by the power spectra and the normalized ACF, respectively, are shown in Fig. 10.4b and c. The SPLs of the two cases fluctuate in the same manner throughout 10 s, but the values of  $\tau_1$  and  $\phi_1$  were extremely different. The mean values of  $\tau_1$  for the two cases were 3.06 and 2.45 ms, but the  $\phi_1$  values for two cases were quite different. At times with high  $\phi_1$ , a tonal noise at  $\tau_1 = 3.06$  s (about 330 Hz) was heard, and its pitch strength fluctuated due to both  $\phi_1$  and the SPL.

Examples of the measured normalized IACF are shown in Fig. 10.5a for landing, take-off, and two flying conditions. The values of the IACC,  $\tau_{IACC}$ , and  $W_{IACC}$  were extracted from the IACF. For example, the measured IACC of four conditions is shown in Fig. 10.5b as a function of time. For the landing and take-off conditions,



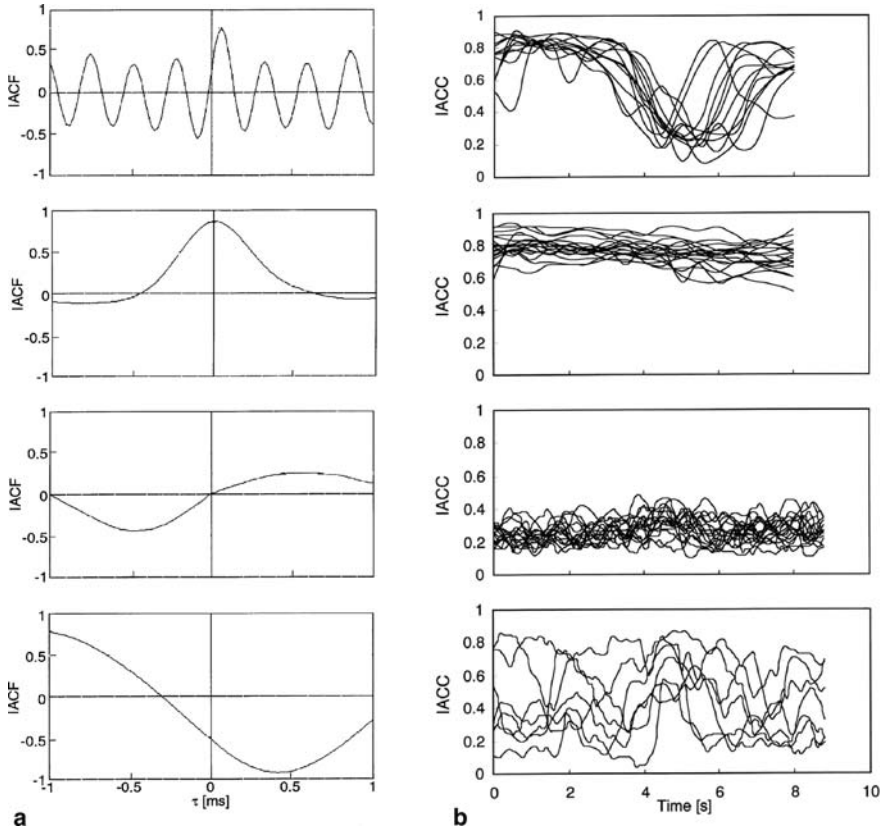
**Fig. 10.4** (a) Examples of measured SPL and values of  $\tau_1$  and  $\phi_1$  for two cases of flying aircraft as a function of time. (—): tonal noise; (----): un-tonal noise. (b) Examples of the measured spectrum at the times when A, B and C indicated in Fig. 10.4(a). (c) Examples of normalized ACF measured at the times A, B, and C indicated in Fig. 10.4a

the IACF had a strong peak at  $\tau_{IACC} \approx 0$ , meaning that the direction of the noise source is perceived clearly just above in the median plane. When  $\tau_{IACC}$  is positive, the localization of the sound source is in the right-hand side, and when it is negative, then the localization is in the left-hand side. The value of the IACC decreased rapidly when the aircraft passed just overhead before landing, because two jet-noise sources with the high-frequency component arrived from different angles from the median plane (Ando and Sakamoto, 1988; Ando, 1998).

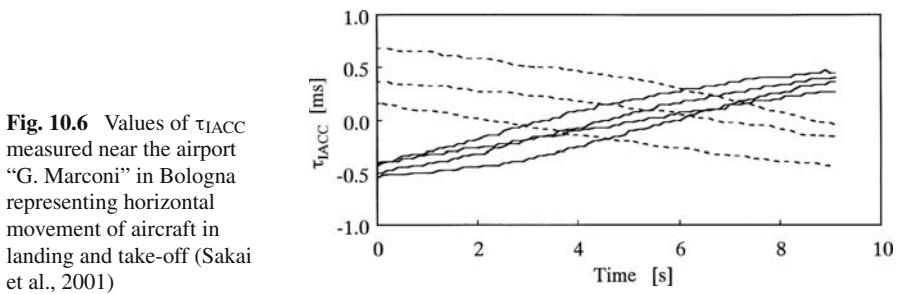
On the other hand, the value of  $W_{IACC}$  is dependent on the dominant frequency of the sound. For the take-off condition,  $W_{IACC}$  was large, because of the low-frequency components. The value of IACC was generally small for the level flying condition, so that subjective diffuseness is high. For such a condition, a flying aircraft at high altitude, the sound signal comes from various directions, because various paths of sound rays may exist due to the ground and the climate conditions of the air. Also, the value of  $W_{IACC}$  was wider for high-level flying caused by the low-frequency components than for low-level landing and take-off conditions due to the higher-frequency components.

As a typical example, Sakai et al. (2001) have measured  $\tau_{IACC}$  of the noise from aircraft in landing and take-off in Bologna. Figure 10.6 shows measured  $\tau_{IACC}$  in the condition of aircraft running from the left to the right (solid curves) and from the right to the left (dotted curves).

Although it has been reported that the dominant frequency component varied throughout the flight for landing aircraft (e.g., Raney and Cawthon, 1979), the current method of the ACF analysis provides much more precise information about noise properties than does the previous method of the sound-level meter based on the envelope information and the spectra measured by the 1/3 or 1/1 octave-band filters.



**Fig. 10.5** Spatial aspects of aircraft noise. (a) Four measured examples of the normalized IACF. (b) Examples of measured IACC for the conditions of landing (*top*), take-off (*second row*), level flying 1 (*third row*), and level flying 2 (*bottom*), respectively



**Fig. 10.6** Values of  $\tau_{IACC}$  measured near the airport “G. Marconi” in Bologna representing horizontal movement of aircraft in landing and take-off (Sakai et al., 2001)

### 10.3 Flushing Toilet Noise

The purpose of this study is to identify factors of the flushing noise of an upstairs toilet, which despite having a low SPL caused annoyance for an apartment resident. We analyzed the temporal and spatial factors of the flushing noise from an upstairs toilet to the head position on a bed, because the resident was very annoyed during sleep. The noise signals were picked up by half-inch condenser microphones placed at two ear entrances on the spherical dummy head, which had been used for the aircraft noise measurements mentioned in the previous section.

The plans of the upstairs and downstairs are shown in Fig. 10.7, wherein recording was performed on two nights (Kitamura et al., 2002). During the measurement, all windows and a bedroom door were closed, and the air conditioner was turned off. The measured temporal factors extracted from the ACF are shown in Fig. 10.8. Solid lines indicate values for a typical example of the flushing toilet noise, and

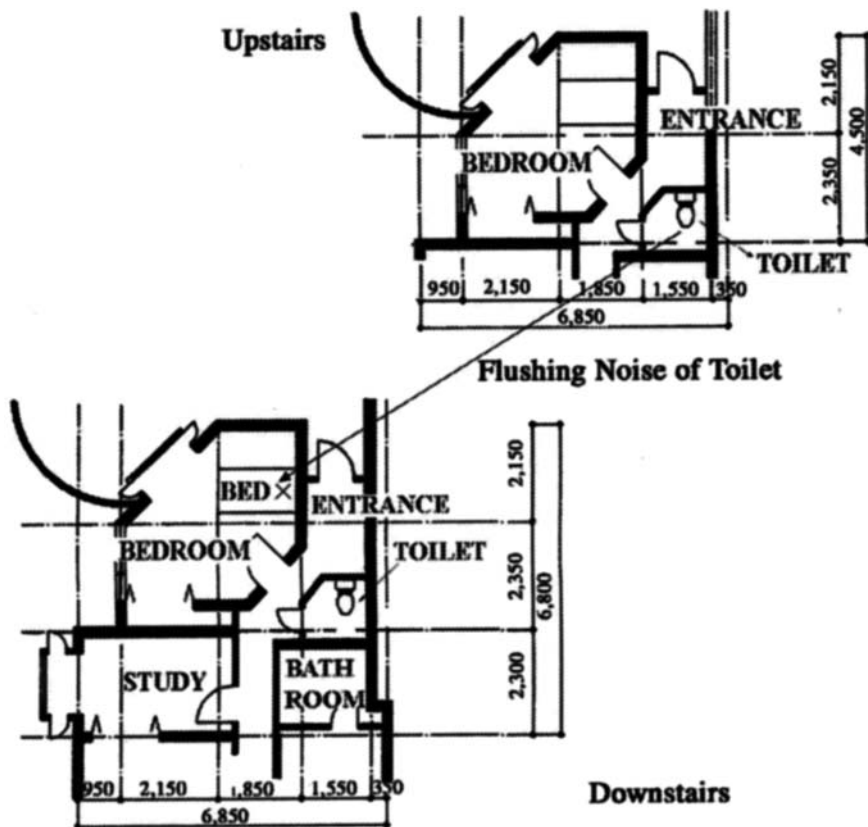
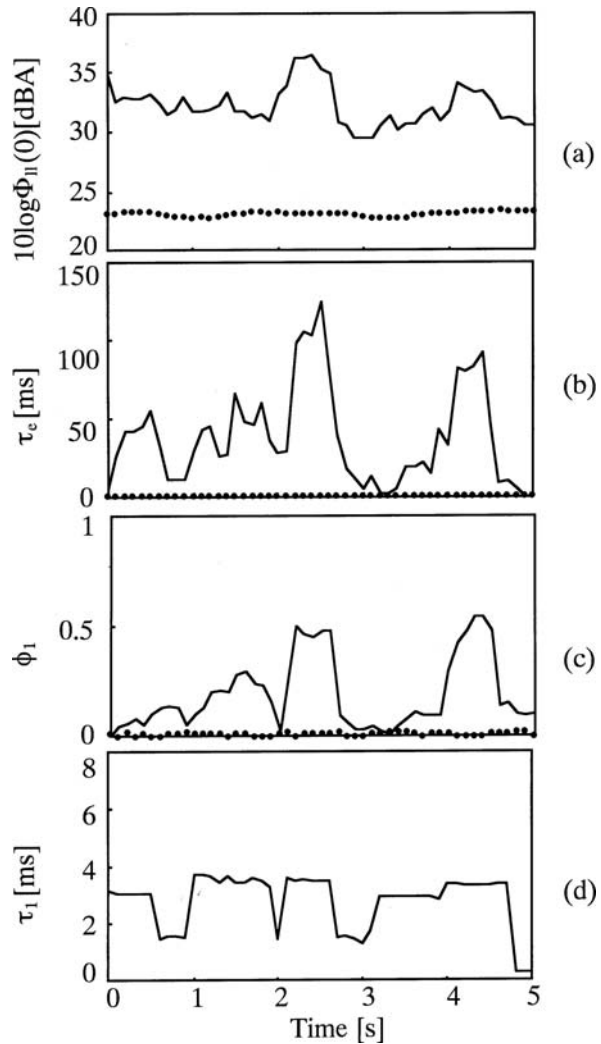


Fig. 10.7 Plans of upstairs and downstairs in an apartment. The flushing noise of an upstairs toilet was recorded on a bed downstairs (X)

dotted lines indicate the background noise. The measurement time was 5 s. The values of all factors were obtained with the integration interval of 0.5 s and the running interval of 100 ms.

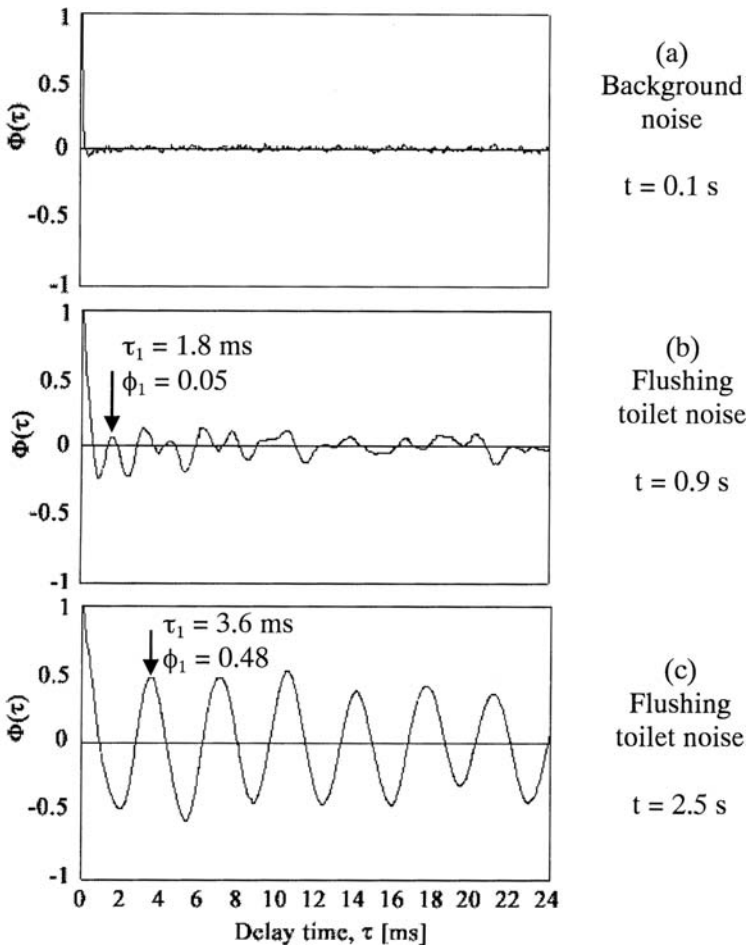
As shown in Fig. 10.8a, the SPL measured by  $\Phi_{11}(0)$  for the flushing toilet noise was between 30 and 35 dBA, but the background noise was about 23 dBA. Thus, the maximum signal-to-noise ratio was 12 dBA. If the difference between the background noise level and the noise signal level is greater than 10 dB, the background noise does not affect significantly the noise signal measurement (Beranek, 1971). As shown in Fig. 10.8b and c, the  $\tau_e$  value for the flushing toilet noise exceeded 100 ms with  $\phi_1 > 0.5$ , and  $\tau_e < 0.1$  ms and  $\phi_1 < 0.0.1$  of the background noise throughout



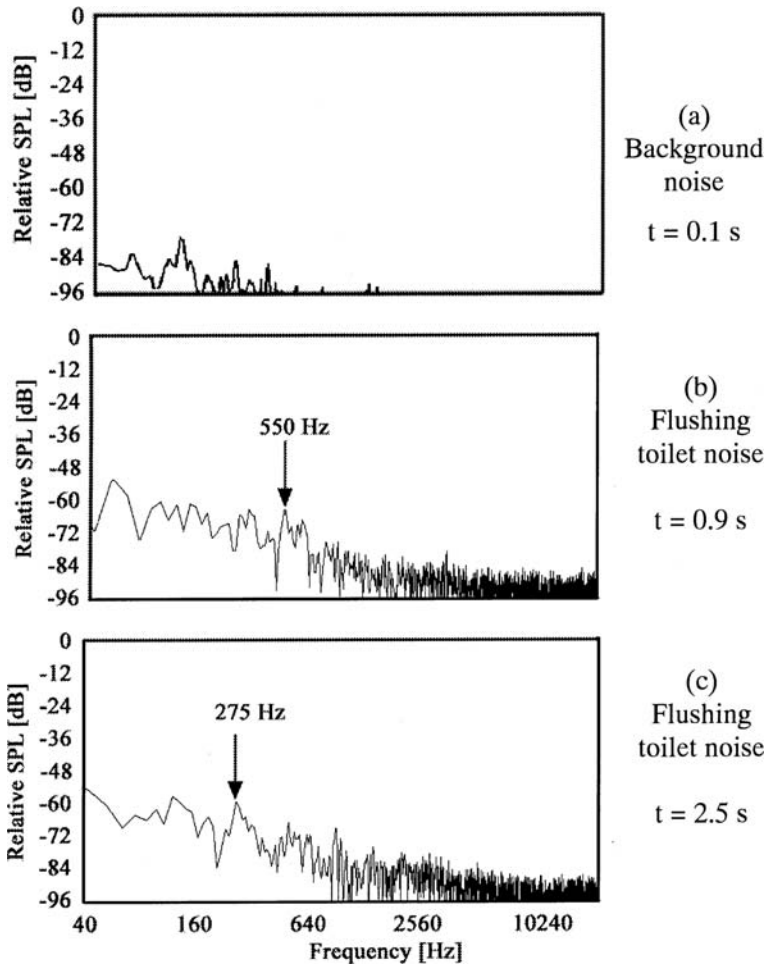
**Fig. 10.8** Measured factors extracted from the running ACF. *Solid line* indicates values for the flushing toilet noise, and *dotted line* indicates the background noise. (a)  $\Phi_{11}(0)$ . (b)  $\tau_e$ . (c)  $\phi_1$ . (d)  $\tau_1$ . The values of all factors were obtained every 100 ms with an integration interval of 0.5 s



the measurement time. Thus, the flushing toilet noise had much more repetitive features than did the background noise. It has been reported that loudness increases in proportion to the value  $\tau_e$  (see Section 6.4; Merthayasa and Ando, 1996). The value of  $\tau_e$  for the flushing toilet noise was the largest near the peak of  $\Phi_{II}(0)$ , as shown in Fig. 10.8a and b. The value of  $\tau_1$  for the flushing toilet noise had a discrete value at 3.6 ms, which means that the perceived pitch was 275 Hz (Fig. 10.6d). The background noise did not have any clear pitch and tonal components, similar to white noise. These are clearly demonstrated in Fig. 10.9, which shows examples of the measured normalized ACF at  $t = 0.1$  s,  $t = 0.9$  s, and  $t = 2.5$  s. The corresponding spectra are shown in Fig. 10.10. In addition, the measured special factors extracted from the IACF are shown in Fig. 10.11a–c. The IACC value for the flushing toilet



**Fig. 10.9** Examples of the NACF analyzed. (a) Background noise measured at  $t = 0.1$  s. (b and c) Flushing toilet noise measured at  $t = 0.9$  s and 2.5 s, respectively

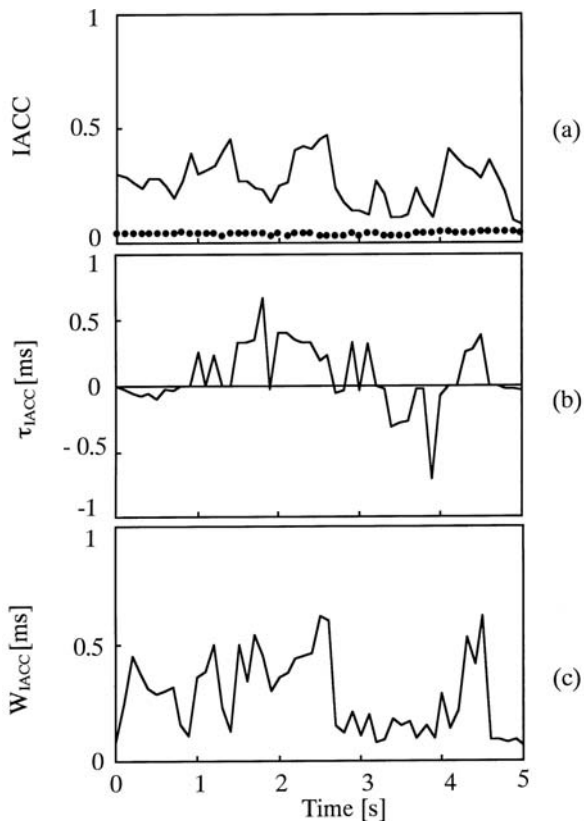


**Fig. 10.10** Power spectra of noise sources. (a) Background noise measured at  $t = 0.1$  s. (b and c) Flushing toilet noise measured at  $t = 0.9$  s and 2.5 s, respectively

noise was much higher than that for the background noise (below 0.05) in the measurement time. Thus, no specific directions may be perceived for the background noise. On the contrary, all spatial factors, the IACC,  $\tau_{IACC}$ , and  $W_{IACC}$  of the flushing noise, changed dramatically as a function of time. These results signify that subjective diffuseness, localization of sound source, and the ASW of the flushing toilet noise changed greatly as a function of time. Judging from these results, a clear pitch and the value of  $\tau_e$  increased the annoyance of the flushing toilet noise as discussed in Section 11.1, although the SPL was low.

Thus far, it was found that the temporal and spatial factors extracted from the ACF and IACF of the flushing toilet noise had specific characteristics. These facts

**Fig. 10.11** Measured factors extracted from the running IACF. *Solid line* indicates values for the flushing toilet noise, and *dotted line* indicates the background noise. (a) IACC. (b)  $\tau_{IACC}$ . (c)  $W_{IACC}$ . The values of all factors were obtained every 100 ms with an integration interval of 0.5 s



imply that both temporal sensations and spatial sensations of the flushing toilet noise changed dramatically. According to our auditory signal processing model, temporal information is mainly processed in the left hemisphere, and spatial information is mainly processed in the right hemisphere (Ando, 1998). Thus, the flushing noise from an upstairs toilet may stimulate both the left and right hemispheres of this resident at the same time. This might partly explain why the resident felt that the flushing noise of an upstairs toilet was very annoying despite its low SPL.

# Chapter 11

## Applications (IV) – Noise Annoyance

Noise is a ubiquitous feature of modern life that is frequently the cause of psychological irritation and stress. We conducted a series of experiments to probe various aspects of noise perception and annoyance. Sections 11.1 and 11.2 discuss experiments in a laboratory that were conducted for annoyance judgments of noise based on both the temporal and spatial factors of environmental noise. In Section 11.3, long-time effects of aircraft noise during pregnancy on postnatal reactions of babies during sleep were investigated. Results show a “prenatal habituation” of babies to aircraft noise, but not music. These differential responses might be explained by the temporal factors extracted from the ACF. In Section 11.4, effects of noise and music on two different mental tasks of children are discussed. Results of two different tasks could be related to the specialization of cerebral hemispheres.

### 11.1 Noise Annoyance in Relation to Temporal Factors

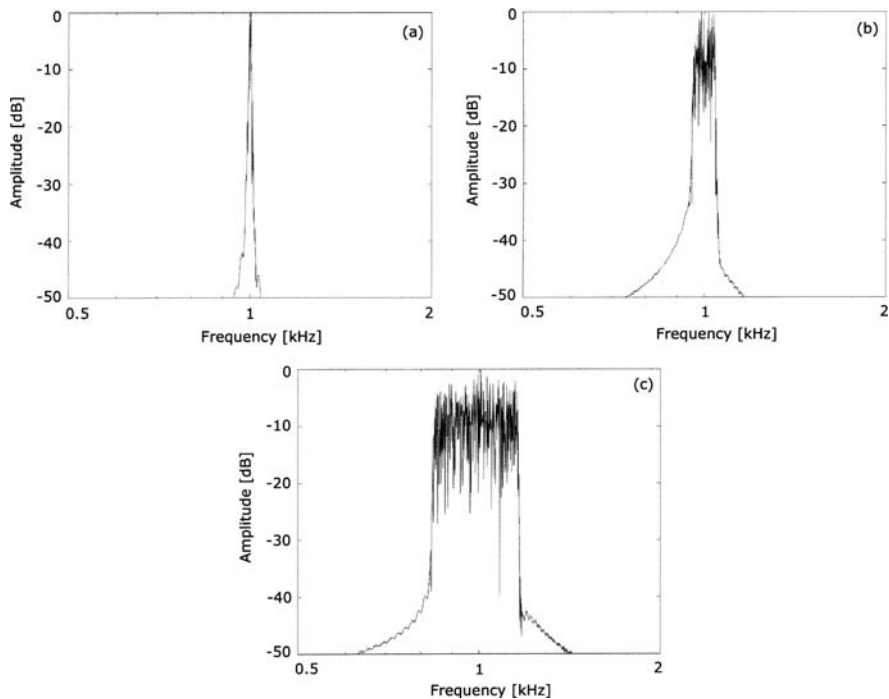
#### 11.1.1 Annoyance of Band-Pass Noise

This study examined annoyance of pure tones and band-pass noises with center frequencies of 1000 and 2000 Hz under equal-SPL conditions. Control of ACF factors of the noise source was realized by use of a 2068 dB/octave sharp filter to create bandwidths of 0, 40, 80, 160, and 320 Hz. The scale value of annoyance was obtained using paired comparisons tests (PCT). Results show that annoyances of pure tones and band-pass noises inside the critical band are not constant, and annoyances of pure tones are greater than those of sharply filtered noises within the critical band. Remarkably, the annoyance of band-pass noise increases with increasing effective duration of the ACF,  $\tau_e$ , which represents the repetitive feature of the source signal within the critical band.

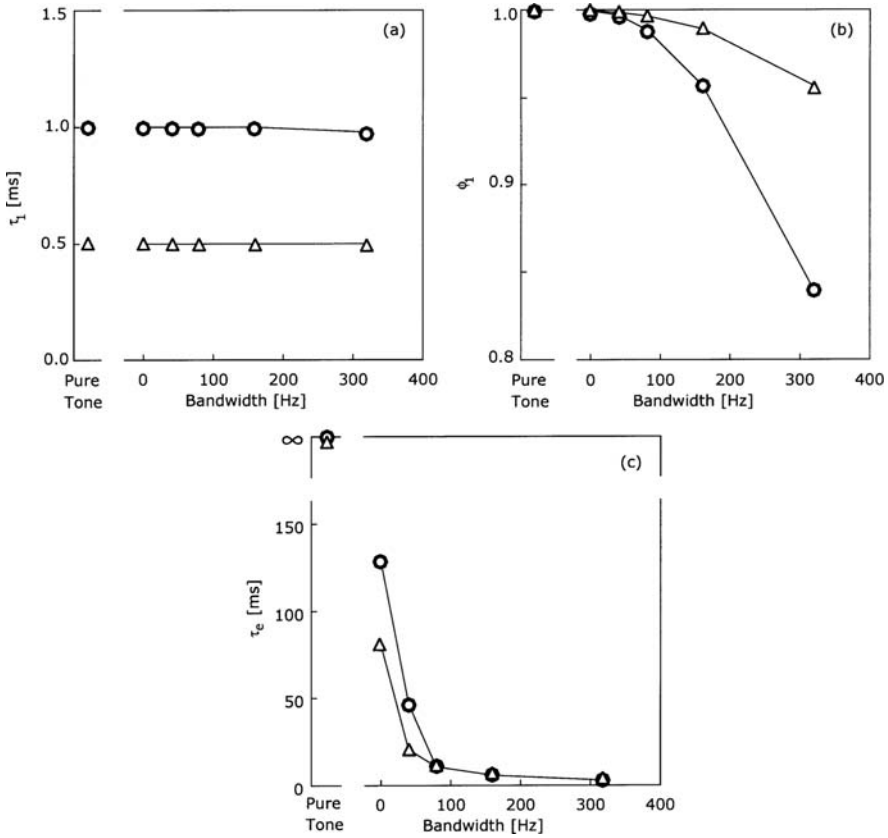
Changes in frequency, duration, and bandwidth all affect perceived loudness of a stimulus even though the level of the stimulus remains fixed (Yost, 2000). Loudness of the noise remains constant as the bandwidth of the noise increases until the bandwidth reaches the critical band. Loudness then increases with increasing bandwidth under the same SPL conditions (Zwicker et al., 1957; Greenwood, 1961a,b; Scharf, 1962; Zwicker and Scharf, 1965). However, it has been found that the loudness of

a sharply filtered noise increases as the effective duration of the ACF,  $\tau_e$ , increases, even when the bandwidth of the signal is within the critical band (see Section 6.4.1). The value of  $\tau_e$  represents repetitive features within the signal itself and increases as the filter bandwidth decreases, and thus loudness increases. As discussed in the following section, ACF factors  $\tau_e$  or  $\phi_1$  had a significant effect on the annoyance of traffic noise.

The current study examines the annoyance of the band-pass noise in terms of factors extracted from the ACF (Soeta et al., 2004). To control the ACF of the band-pass noise, the filter bandwidth was varied at 0, 40, 80, 160, and 320 Hz by using a cutoff slope of 2068 dB/octave, obtained by a combination of two filters (Sato et al., 2002). The filter bandwidth of 0 Hz had only the cutoff slope components in its low- and high-frequency ranges. The critical bandwidths of the center frequencies of 1000 and 2000 Hz are approximately 160 and 300 Hz (Zwicker and Terhardt, 1980). Figure 11.1 shows the power spectra of the noise signal centered on 1000 Hz used in this study. The source signal was characterized by the ACF factors  $\tau_1$ ,  $\phi_1$ , and  $\tau_e$  (Fig. 11.2). The value of  $\tau_1$  corresponds to the center frequency of the band-pass noise;  $\phi_1$  and  $\tau_e$  increase as the filter bandwidth decreases. Note that there is a certain degree of coherence between  $\phi_1$  and  $\tau_e$ . The auditory stimuli were binau-



**Fig. 11.1** Power spectra of band-pass noise centered on 1 kHz after passing through the filter with a cutoff slope of 2068 dB/octave. (a) Bandwidth: 0 Hz. (b) Bandwidth: 40 Hz. (c) Bandwidth: 320 Hz



**Fig. 11.2** Measured ACF factors of the band-pass noise source signal as a function of the bandwidth. (a) Delay time of the first maximum peak of ACF,  $\tau_1$ . (b) Amplitude of the first maximum peak of ACF,  $\phi_1$ . (c) Effective duration of ACF,  $\tau_e$ . Different symbols indicate different frequencies: (○): 1 kHz, (△): 2 kHz

rally presented using headphones (Sennheiser HD-340, Wedemark, Germany). All stimuli were fixed exactly at 74 dBA by measuring  $\Phi(0)$ . The SPL was calibrated by using a dummy head with half-inch condenser-type microphones at both ears. The IACC was kept constant at unity, because the signals fed to both ears were identical.

Eight 21- to 23-year-old subjects with no histories of hearing disorders participated in the experiment. They were seated in a dark, soundproof room with a comfortable thermal environment where they listened to the sound stimuli. Paired comparison tests were performed for all combinations of the pairs of pure tone and band-pass noise, that is, 15 pairs ( $N(N - 1)/2$ ,  $N = 6$ ) of stimuli interchanging the order in each pair per session, with the pairs presented in random order. A total of 10 sessions were conducted for each subject. The duration of the stimuli was 2.0 s, the rise and fall times were 50 ms, the silent interval between the stimuli was 1.0 s, and the interval between pairs was 3.0 s, which was the time allowed for the subject

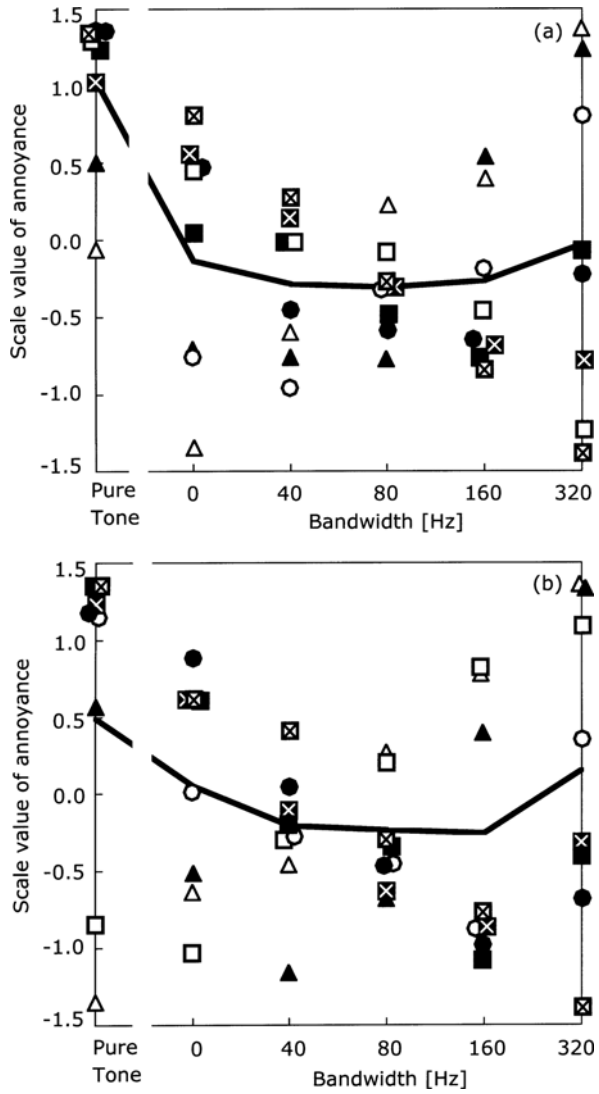
to respond by pushing one of two buttons. They were asked to judge which of two sound signals was more annoying.

The relationship between the scale value of annoyance and the filter bandwidths is shown in Fig. 11.3. For the center frequency of 1000 Hz, the averaged scale value of annoyance for the pure tone was significantly greater than those of sharply filtered noises under the condition of equal SPL ( $p < 0.01$ ). For the center frequency of 2000 Hz, the scale value of annoyance for the pure tone was significantly greater than that of the band-pass noises with 40, 80, and 160 Hz bandwidths ( $p < 0.05$ ). There was agreement among subjects that the most annoying stimulus was the pure tone or band-pass noise with a 320-Hz bandwidth. However, relatively large individual differences were observed of the noise with a 320-Hz bandwidth. The previous research suggests that tonal components increase the perceived annoyance and noisiness of broadband noise (e.g., Kryter and Pearsons, 1965; Hargest and Pinker, 1967; Fujii et al., 2002).

The loudness of the band-pass noise with center frequencies of 1000 and 2000 Hz was also investigated. The same subjects who had participated in the annoyance tests participated in the loudness test. The auditory stimuli were presented in the same way as in the annoyance tests. The subjects were asked to judge which of two sound signals was the louder signal. The relationship between the scale values of loudness and the filter bandwidths is shown in Fig. 11.4. For the band-pass noise centered on 1000 Hz, results of loudness as a function of its bandwidth are similar to those of Section 6.4 with different subjects. It is reconfirmed that the loudness for the pure tone was significantly greater than that of the band-pass noise, and loudness increased with increasing  $\tau_e$ , although one subject in the study had increasing loudness from the pure tone to the widest bandwidth. The averaged scale value of loudness for the band-pass noise with the bandwidth of 320 Hz was significantly greater than those for the other band pass noises for both center frequencies of 1000 and 2000 Hz. This is consistent with previous research by Zwicker et al. (1957), Greenwood (1961a,b), Scharf (1962), and Sato et al. (2002). In this study, three of the eight subjects had identified increasing loudness with increasing bandwidth. The increase of loudness and annoyance by the pure tone was not found in two subjects for the noise centered on 2000 Hz, as indicated with an open triangle and square in Fig. 11.3b. Almost all subjects showed an increase of loudness above the critical band. However, the increase of annoyance above the critical band showed individual variations. The increase in annoyance for the pure tone (Fig. 11.3) was more evident than that of loudness (Fig. 11.4; see also Fig. 6.16). There are many more individual differences in the annoyance judgment than those in the loudness judgment. Loudness and annoyance produced different results. This is consistent with the findings obtained by Berglund et al. (1975), Hellman (1982, 1984), and Kuwano et al. (1988).

The ACF factors  $\tau_e$  or  $\phi_1$ , which show the tonal component included in noise, had a significant effect on annoyance (Fujii et al., 2002). In this study, annoyance increased as the  $\tau_e$  of the sound source increased within the critical band. Therefore, factors extracted from the ACF, such as  $\tau_e$ , may be useful for evaluating noise. It was found that the annoyance of the pure tone and the band-pass noise with equal

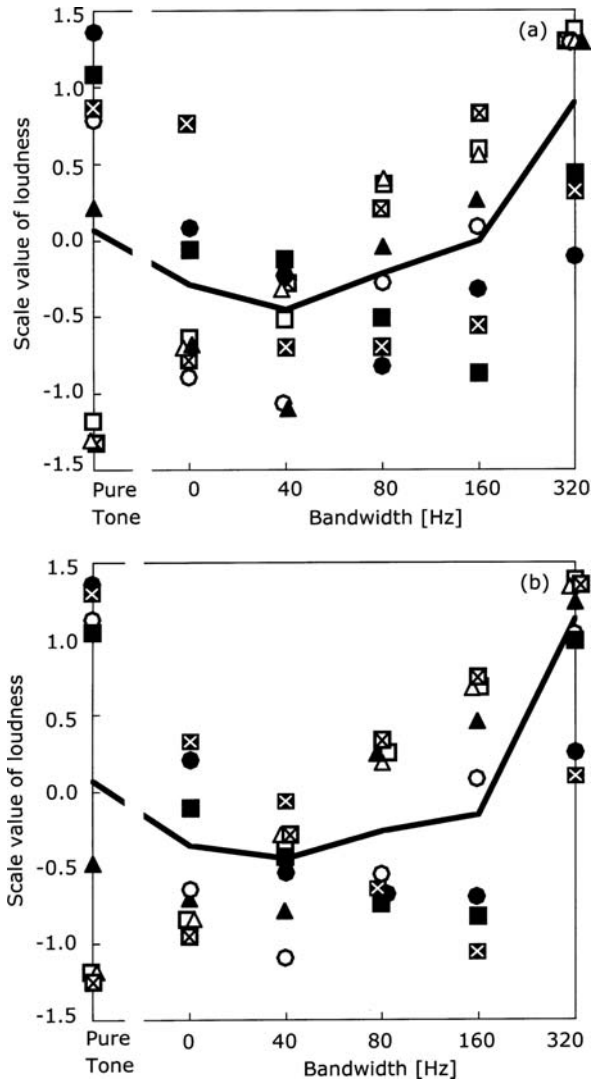
**Fig. 11.3** Scale value of annoyance as a function of the bandwidth of band-pass noise. (a) The center frequency: 1 kHz. (b) The center frequency: 2 kHz. Each symbol represents a single subject. The line represents the mean scale value of eight subjects



SPL is not constant within the critical band. In particular, annoyance of the pure tone is larger than that of the sharply filtered noise within the critical band, thus, annoyance increases with increasing  $\tau_c$ . It is worth noting that annoyance increased with the logarithm of the physical duration ( $\log D$ ) of the white noise in the range of 30 ms to 90 s tested under the conditions of constant SPL (Hiramatsu et al., 1983).



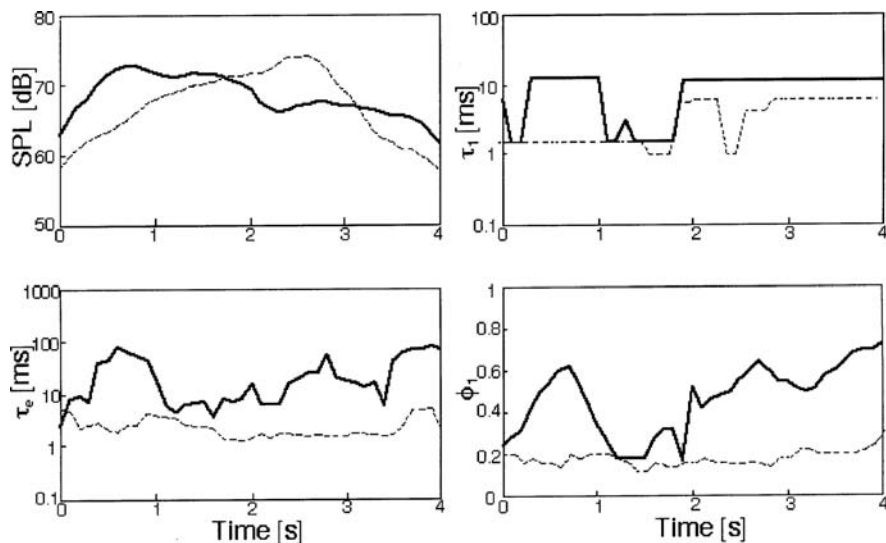
**Fig. 11.4** Scale value of loudness as a function of the bandwidth of band-pass noise. (a) The center frequency: 1 kHz. (b) The center frequency: 2 kHz. Each symbol represents a single subject. The line represents the mean scale value of eight subjects



### 11.1.2 Annoyance of Traffic Noise

The purpose of this section is to discuss the annoyance of traffic noise in relation to 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 Fig. 11.5 as a time function (Fujii et al., 2002). Thick lines and thin lines show two extremes of



**Fig. 11.5** Two extreme examples of measured SPL and three ACF factors as a function of time. *Solid lines* show factors of tonal noise (a motor bike), and *dashed lines* show factors of un-tonal noise (a passenger car)

noise: one has a clear pitch as a tonal and the other has a weak pitch as a nontonal because of a small value  $\phi_1$  near 0.2, as defined in Section 5.2. The value of  $\tau_1$  varied discretely 1 ms and 10 ms, meaning that perceived pitch varied 1000 Hz and 100 Hz for both noises. The strength of perceived pitch increases in proportion to the value of  $\phi_1$  (Section 6.2). For the tonal noise, the  $\phi_1$  value reaches a maximum around 0.7. Thus, a strong tonal noise is heard having a pitch of 100 Hz ( $\tau_1 = 10$  ms). When the  $\tau_1$  value varies with a high  $\phi_1$  value, we perceive a variation of the pitch. If the  $\phi_1$  value for the un-tonal noise remained constant around 0.2, in spite of the variation of  $\tau_1$ , then perceived pitch for the un-tonal noise is weak enough, and it is hard to discriminate pitch fluctuation.

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  $\sigma^2$  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 is indicated in Table 11.1, however, most of these standard measures were highly intercorrelated 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, minimum SPL,  $L_{10}$ , and  $L_{90}$  were selected in the subsequent analysis.

The calculated power spectrum of the unfiltered signal and the autocorrelation function of the A-weighted filtered signal for the tonal and nontonal (atonal, un-tonal) noise are depicted in Fig. 11.6. For the tonal noise, there are several discrete

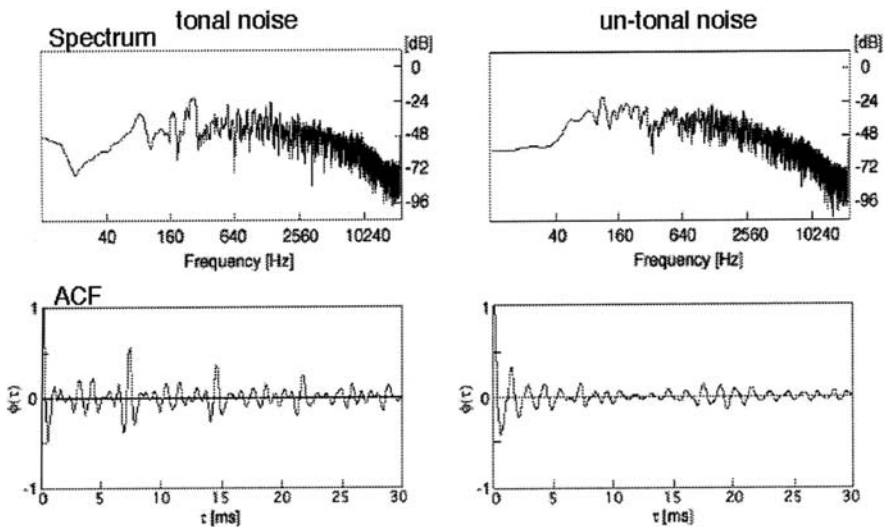
**Table 11.1** Correlations between previously defined eight standard noise measures for nine different traffic noises

Factor	Max	Min	Mean	$\sigma^2$	L <sub>50</sub>	L <sub>eq</sub>	L <sub>90</sub>	L <sub>10</sub>
Max	1							
Min	-0.69*	1						
Mean	-0.24	0.62	1					
$\sigma^2$	0.89**	-0.91**	-0.53	1				
L <sub>50</sub>	0.06	0.24	0.84**	-0.11	1			
L <sub>eq</sub>	0.57	-0.31	0.44	0.48	0.82**	1		
L <sub>90</sub>	-0.85**	0.87**	0.66*	-0.97**	0.27	-0.32	1	
L <sub>10</sub>	0.70*	-0.46	0.28	0.63*	0.70**	0.98**	-0.49	1

\*\*p < 0.01, \*p < 0.05.

peaks in the power spectrum, but it is difficult to identify which peaks correspond to the clear pitch that is heard. 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 (Meddis and Hewitt, 1991a,b; Cariani and Delgutte, 1996a,b). The information can be used to identify the noise source. Also, the envelope of the ACF represented by the value of  $\tau_c$  is a good measure of the repetitive feature of the sound signal.

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 ( $73 \pm 2$  dB). To make the envelope of



**Fig. 11.6** Power spectrum and the ACF measured of tonal and un-tonal noises

noise equal, an 0.5-s rise and fall time was added to all stimuli. To characterize the acoustic properties of a stimulus, we used the median and variance of each factor.

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 judge 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 the scale value (SV) of annoyance is shown in Table 11.2. Contrary to the assumption, perceived annoyance was not correlated with 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 annoyance. In other words,  $\text{Var\_SPL}$ ,  $\tau_e$ , and  $\text{Var\_}\tau_1$  are more effective on the scale value of annoyance than SPL in the range of about 5 dB. The values of  $\tau_e$  and  $\phi_1$  were significantly correlated with 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 nontonal noise. The comments of subjects also indicated that they judged a sound having a clear pitch to be more annoying.

**Table 11.2** Correlations between the ACF factors (median and variance) and annoyance

Factor	SPL	$\tau_1$	$\phi_1$	$\tau_e$	$\text{Var\_SPL}$	$\text{Var\_}\tau_1$	$\text{Var\_}\phi_1$	$\text{Var\_}\tau_e$
SPL	1							
$\tau_1$	-0.66	1						
$\phi_1$	-0.29	0.82**	1					
$\tau_e$	0.34	0.33	0.74**	1				
$\text{Var\_SPL}^1$	-0.11	0.02	0.22	0.03	1			
$\text{Var\_}\tau_1$	-0.57	0.46	0.37	0.35	-0.04	1		
$\text{Var\_}\phi_1$	-0.09	0.50	0.30	0.78**	-0.35	0.12	1	
$\text{Var\_}\tau_e$	-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*

\*\* $p < 0.01$ , \* $p < 0.05$ . Symbol Var represents the variance defined by  $\sigma^2 = 1/n \sum (x_i - m)^2$ , where  $x_i$  is the value of data  $i$ ,  $m$  is the mean value, and  $n$  is the number of the data.

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 its accuracy is not well established (May, 1978). Instead, as

is discussed here by using the values of  $\tau_e$  and  $\phi_1$ , effects of the tonal component on perceived annoyance may be clearly described. As indicated in Table 11.2,

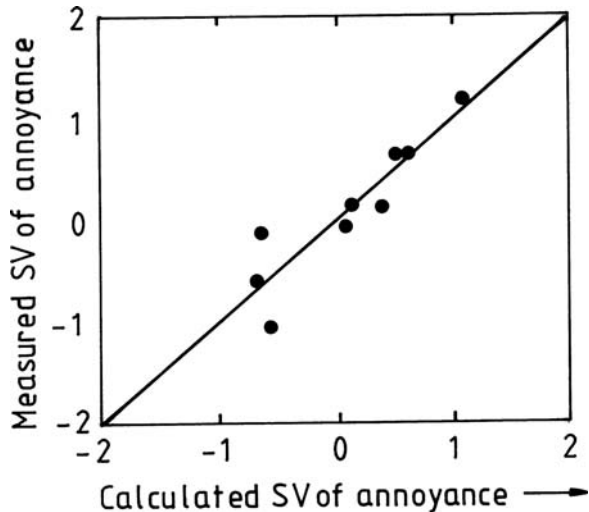
1.  $\text{Var\_SPL}$ , (fluctuations in loudness)
2.  $\tau_e$  as a representative of  $\phi_1$ ,  $\text{Var\_}\phi_1$ , and  $\text{Var\_}\tau_e$ , and (fluctuations in pitch strength and temporal coherence)
3.  $\text{Var\_}\tau_1$  as a representative of  $\text{SPL}$  and  $\tau_1$  (fluctuations in pitch and timbre)

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,

$$S = S_L \approx a\text{Var\_SPL} + b\tau_e + c\text{Var\_}\tau_1 \quad (11.1)$$

where coefficients obtained are  $a \approx 0.64$ ,  $b \approx 0.50$ , and  $c \approx 0.36$ . Using these tentative constants in Equation (11.1), the total correlation coefficient 0.91 was obtained with the significance level  $p < 0.05$ , as shown in Fig. 11.7. Similarly, the noise whose pitch and timbre fluctuates is more annoying than one having a constant quality (Molino, 1979).

**Fig. 11.7** Relationship between measured and annoyance calculated by using linear combination of the ACF factors ( $r = 0.91$ ,  $p < 0.05$ )



This result shows that the combination of the ACF factors was sufficient 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 and applied for any environmental noise evaluations (Sato et al., 2007).

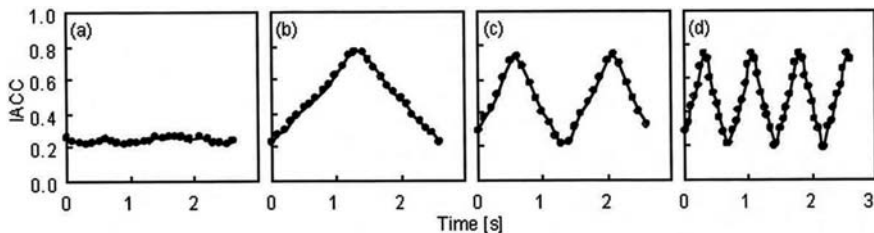
## 11.2 Noise Annoyance in Relation to Spatial Factors

This study examines effects of spatial factors extracted from the IACF on annoyance of noise stimuli in a laboratory. In the first experiment, using paired comparison tests the subjects judged their annoyance while changing fluctuations in the IACC and the SPL. In the second, they judged their annoyance while changing fluctuations in the  $\tau_{\text{IACC}}$  and the SPL. Remarkable findings are that annoyance increased with increasing rates of the IACC and the  $\tau_{\text{IACC}}$  fluctuations, in reference to increasing the SPL.

Environmental noises have been evaluated according to the SPL and their frequency characteristics. The noise criterion (NC) curve, the preferred noise criterion (PNC) curve, and the balanced noise criterion (NCB) curve have been developed for measurements of the SPL and its frequency characteristics (Beranek, 1957, 1971; Beranek et al., 1971; Beranek, 1988). Evaluations of temporal fluctuations involved in both traffic and industrial noise have used the equivalent sound level ( $L_{\text{eq}}$ ). The purpose of the current study, however, is to investigate effects of fluctuation of two spatial factors extracted from the IACF on annoyance of noise stimuli, which has not been known previously (Sato et al., 2004).

### 11.2.1 Experiment 1: Effects of SPL and IACC Fluctuations

Annoyance judgments were performed using paired comparisons while changing fluctuations in the interaural correlation magnitude IACC and sound pressure level SPL. The sound source was white noise. The rate of the IACC fluctuation was simulated by the frontal direct sound ( $L_0$ ) and two symmetric lateral reflections ( $L_1$  and  $L_2$ ) in a soundproof chamber. To produce an incoherent condition, the time delays between the direct sound and the two reflections were fixed at  $\Delta t_1 = 20$  ms and  $\Delta t_2 = 40$  ms. To change the IACC, a logarithmic envelope change was applied to the amplitude of the reflections. The period of the IACC change,  $P_{\text{IACC}}$ , was set at 0.375, 0.75, 1.5 s, and  $\infty$ . The SPLs were set at 65, 70, and 75 dBA, and the resulting measured total SPLs were  $65 \pm 0.7$ ,  $70 \pm 0.8$ , and  $74.9 \pm 0.8$  dBA, respectively. Values of the running IACC measured with  $2T = 0.1$  s are shown in Fig. 11.8.



**Fig. 11.8** Measured IACC for experiment 1. (a)  $P_{\text{IACC}} = \infty$ . (b)  $P_{\text{IACC}} = 1.50$ . (c)  $P_{\text{IACC}} = 0.75$ . (d)  $P_{\text{IACC}} = 0.375$

When the IACC fluctuates, annoyance is assumed to be affected by the maximum value of its time gradient, which is defined by  $Flu\_IACC$ , such that,

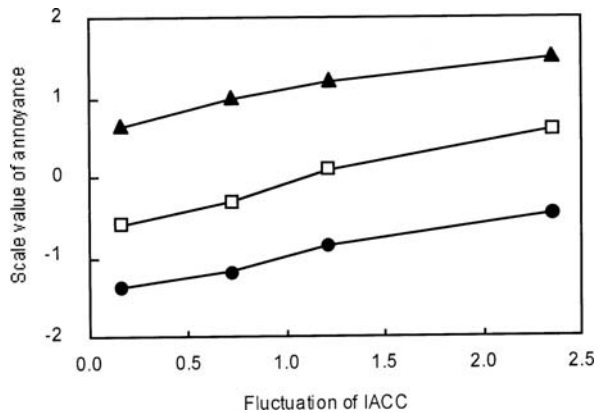
$$\left| \frac{dS_{IACC}(t)}{dt} \right|_{\max} \approx \left| \frac{S_{IACC}(i\Delta) - S_{IACC}(i-1)\Delta}{\Delta} \right|_{\max} \equiv Flu\_IACC \quad (11.2)$$

where  $\Delta$  is a small time interval. We have selected  $\Delta = 0.1$  s obtaining its convergence in this study. Values of  $Flu\_IACC$  that resulted were 2.34, 1.21, 0.72, and 0.17 for values of  $P_{IACC} = 0.375, 0.75, 1.5$  s, and  $\infty$ , respectively. Fluctuations in other spatial factors,  $\tau_{IACC}$  and  $W_{IACC}$ , could be fixed within  $-0.01 \pm 0.01$  ms and  $0.02 \pm 0.01$  ms, respectively, because listeners always localized at the front in the median plane and the identical noise signal.

Paired comparisons tests were conducted for the 12 stimuli (four levels of  $Flu\_IACC \times$  three levels of SPL). Five subjects 22 to 24 years old with normal hearing ability participated in the experiment. A single test session consisted of 66 pairs  $(N(N-1)/2, N = 12)$  of stimuli. Ten sessions were performed for each subject. A single test session was divided into two parts, each of which lasted 5.5 min, to prevent subject fatigue. Fifty responses (5 subjects  $\times$  10 sessions) to each stimulus were obtained. Consistency tests indicated that all subjects had a significant ( $p < 0.05$ ) ability to distinguish various degrees of annoyance.

As shown in Fig. 11.9, the scale value of annoyance increased as the  $Flu\_IACC$  and the SPL increased. A  $Flu\_IACC > 0$  was consistently judged to be more annoying than the condition of the  $Flu\_IACC = 0$  when the SPL was constant. Results of the analysis of variance for scale values of annoyance indicate that the factors  $Flu\_IACC$  and the SPL are significant ( $p < 0.01$ ). Effects of the interaction between  $Flu\_IACC$  and the SPL were not significant, so that the  $Flu\_IACC$  and the SPL contributed to the scale value of annoyance independently, thus,

$$S = S_R \approx f_R(Flu\_IACC) + f_R(SPL) \approx a(Flu\_IACC) + b(SPL) \quad (11.3)$$



**Fig. 11.9** Average scale value of annoyance as a function of  $Flu\_IACC$  and as a parameter of SPL. ●, SPL = 65 dBA; □, SPL = 70 dBA; and ▲, SPL = 75 dBA

where coefficients  $a \approx 0.46$  and  $b \approx 0.20$  were obtained by multiple regressions. It is remarkable that the calculated scale values agree well with the measured ones with a correlation coefficient of 0.99 ( $p < 0.01$ ).

For each individual, weighting coefficients  $a$  and  $b$  in Equation (11.3) were also obtained (Table 11.3). Two of five subjects indicated that the scale value did not satisfy the model’s goodness of fit, because the probability of their judgments in the PCT was beyond the linear range of the scale value ( $0.05 < p < 0.95$ ). Therefore, coefficients in Equation (11.3) for subjects A and D could not be obtained.

**Table 11.3** Contributions and coefficients in Equation (11.3) for each individual in experiment 1

Subject	Contribution (%)		Total	Coefficient	
	<i>Flu_IACC</i>	SPL		<i>a</i>	<i>b</i>
A	11.0	88.2	99.2	–	–
B	18.8	79.6	98.4	0.47	0.19
C	11.2	88.0	99.2	0.36	0.20
D	32.7	65.1	97.8	–	–
E	14.5	84.6	99.1	0.38	0.18
Global	17.1	82.6	99.7	0.46	0.20

### 11.2.2 Experiment 2: Effects of Sound Movement

Annoyance judgments were performed using paired comparison tests while changing fluctuations in the  $\tau_{IACC}$  and the SPL. The source signal was the band-pass filtered noise with a center frequency of 500 Hz (bandwidth = 160 Hz). The horizontally swaying sound images were simulated by the two lateral, symmetric loudspeakers ( $L_1$  and  $L_2$ ) in the soundproof chamber. In these experiments, the horizontal angles of lateral sounds were  $\pm 54^\circ$ .

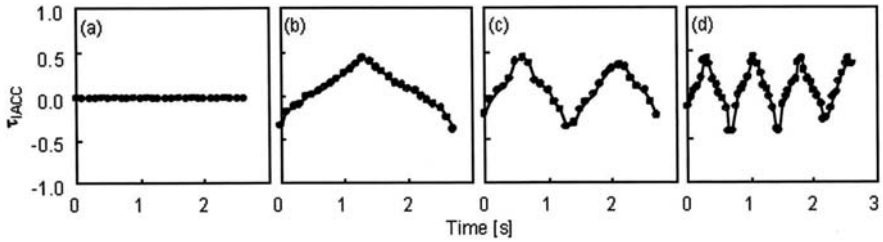
To produce the sound field with the fluctuation of  $\tau_{IACC}$ , the logarithmic envelope was applied to the amplitude of noise. The period  $P_{\tau_{IACC}}$  was set at 0.375, 0.75, 1.5 s, and  $\infty$ . The SPL values were set at 65, 70, and 75 dBA and the ranges of the measured total SPL were  $64.8 \pm 1.5$ ,  $69.6 \pm 1.6$ , and  $74.7 \pm 1.5$  dBA, respectively. The measured values of  $\tau_{IACC}$  are shown in Fig. 11.10. To follow the fluctuations of  $\tau_{IACC}$ , the IACF was calculated for the integration intervals  $2T = 0.3$  s.

When the value of  $\tau_{IACC}$  fluctuates, then annoyance can be affected by the maximum value of its time gradient, which is defined by *Flu\_τ<sub>IACC</sub>*

$$\left| \frac{dS_{\tau_{IACC}}(t)}{dt} \right|_{\max} \approx \left| \frac{S_{\tau_{IACC}}(i\Delta) - S_{\tau_{IACC}}(i-1)\Delta}{\Delta} \right|_{\max} \equiv Flu_{\tau_{IACC}} \quad (11.4)$$

Here,  $\Delta$  is selected 0.3 s small enough for obtaining the convergence. Values of *Flu\_τ<sub>IACC</sub>* obtained were 3.73, 2.07, 1.50, and 0.07 for  $P_{IACC} = 0.375, 0.75, 1.5$  s, and  $\infty$ , respectively. Fluctuations of the measured IACC and  $W_{IACC}$  were almost constant within  $0.91 \pm 0.08$  ms and  $0.27 \pm 0.01$  ms, respectively.



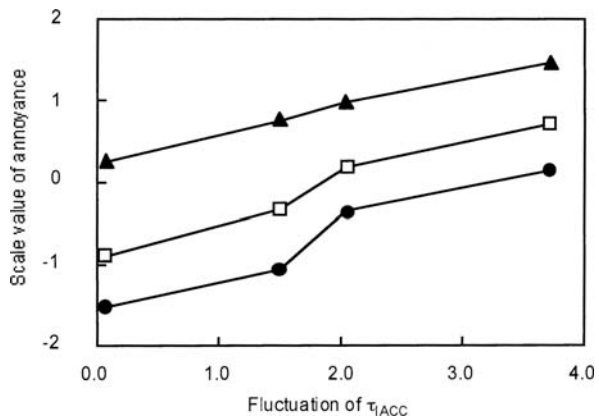


**Fig. 11.10** Measured  $\tau_{IACC}$  for experiment 2. (a)  $P\tau_{IACC} = \infty$ . (b)  $P\tau_{IACC} = 1.50$ . (c)  $P\tau_{IACC} = 0.75$ . (d)  $P\tau_{IACC} = 0.375$

Paired comparisons tests were conducted for 12 stimuli (four levels of  $Flu_{\tau_{IACC}}$   $\times$  three levels of SPL). The same five subjects from experiment 1 participated. They were asked to judge which of two stimuli they perceived to be more annoying. The duration of the stimuli was 3 s, the rise and fall times were 50 ms, and the silent interval between the stimuli was 1 s. Each pair of stimuli was separated by an interval of 3 s and presented randomly. A single test session consisted of 66 pairs  $(N(N - 1)/2, N = 12)$  of stimuli. Ten sessions were performed for each subject. A single test session was divided into two parts, each of which lasted 5.5 min. Fifty responses (5 subjects  $\times$  10 sessions) to each stimulus were obtained.

Figure 11.11 shows that the scale values of annoyance increased as  $Flu_{\tau_{IACC}}$  and the SPL increased. When the SPL was constant, the moving sound images ( $Flu_{\tau_{IACC}} > 0$ ) were always more annoying than the fixed sound image ( $Flu_{\tau_{IACC}} = 0$ ). Results of an analysis of variance for scale value of annoyance indicate that the factors  $Flu_{\tau_{IACC}}$  and SPL are significant ( $p < 0.01$ ). Effects of the interaction between  $Flu_{\tau_{IACC}}$  and SPL are not significant. Thus,  $Flu_{\tau_{IACC}}$  and the SPL contribute to the scale value of annoyance independently, so that:

$$S = S_R \approx f_R(Flu_{\tau_{IACC}}) + f_R(SPL) \approx a(Flu_{\tau_{IACC}}) + b(SPL), \quad (11.5)$$



**Fig. 11.11** Average scale value of annoyance as a function of  $Flu_{\tau_{IACC}}$  and as a parameter of SPL. ●, SPL = 65 dBA; □, SPL = 70 dBA; and ▲, SPL = 75 dBA

where coefficients  $a \approx 0.42$  and  $b \approx 0.16$  were obtained by multiple regression. The calculated scale values agree well with the measured ones with a correlation coefficient of 0.98 ( $p < 0.01$ ).

As the contribution of the factors  $Flu_{\tau_{IACC}}$  and the SPL to the scale value of annoyance for each individual, coefficients  $a$  and  $b$  in Equation (11.5) were also obtained (Table 11.4). One of five subjects indicated that the scale values did not satisfy the model of goodness of fit. Therefore, the coefficients given in Equation (11.5) for subject E were not obtained. A comparison of the results from experiment 1, for all subjects, shows the effects of the SPL are relatively small, and  $Flu_{\tau_{IACC}}$  cannot be ignored for the evaluation of annoyance. As listed in Table 11.4, the contribution of the  $Flu_{\tau_{IACC}}$  to the scale value of annoyance apparently was greater than that of  $Flu_{IACC}$ . It is noteworthy that the variations in the ranges of SPL were the same from 65 to 75 dBA in both experiments. Because  $\tau_{IACC} = 0$  is one of the preferred conditions for listening in the sound field (see Section 3.2.4), annoyance increased with the stimuli of  $Flu_{\tau_{IACC}} > 0$ .

**Table 11.4** Contributions and coefficients in Equation (11.5) for each individual in experiment 2

Subject	Contribution (%)		Total	Coefficient	
	$Flu_{\tau_{IACC}}$	SPL		a	b
A	40.2	59.3	99.5	0.37	0.14
B	25.5	71.2	96.7	0.31	0.17
C	15.6	83.5	99.1	0.24	0.19
D	79.2	19.6	98.8	0.50	0.08
E	46.7	51.6	98.3	–	–
Global	40.8	58.9	99.7	0.42	0.16

The value of the IACC was between 0.2 and 0.8 in experiment 1, however, subjects did not perceive a fluctuation in the IACC when the IACC was such a low value (see, e.g., Section 7.3). Or, the just noticeable difference (JND) of the IACC in the sound field with a lower IACC is larger than that with a higher IACC (Gabriel and Colburn, 1981). In addition, the sound image is blurred when the IACC is low.

On the other hand, in experiment 2, the value of  $\tau_{IACC}$  was changed from  $-0.4$  to  $0.4$  ms, and  $P_{\tau_{IACC}}$  was  $0.38$  s at the minimum. The threshold of the interaural time delay (for the 1000-Hz tone) is short enough,  $10 \mu\text{s}$  (Klumpp and Eady, 1956). The duration of  $0.38$  s is long enough to perceive the movement of the sound source (Grantham, 1986, Chandler and Grantham, 1992). In addition, all stimuli had clear sound image, because the values of the IACC for all stimuli in this experiment were greater than 0.82. Therefore, subjects perceived the fluctuation of localization according to  $\tau_{IACC}$  during the whole period of the stimuli much more vividly than the condition of the IACC fluctuation. Under two experimental conditions, 3.73 of the  $Flu_{IACC}$  is equivalent to an increase of 4.9 dB in the SPL, and 2.96 of the  $Flu_{\tau_{IACC}}$  is equivalent to an increase of 9.7 dB in the SPL.

The results of the two experiments lead to the following conclusions:

1. Moving spatial sound sensations were always more annoying than fixed sound localization under the condition of a constant SPL. The annoyance increased with a greater fluctuation rate of the IACC as well as the  $\tau_{\text{IACC}}$ .
2. Fluctuations of the IACC and the SPL independently contribute to the scale value of annoyance.
3. Fluctuations of the  $\tau_{\text{IACC}}$  and the SPL independently contribute to the scale value of annoyance.
4. The contribution of fluctuations in  $\tau_{\text{IACC}}$  to annoyance was greater than that of the IACC when the range of the SPL was from 65 to 75 dBA.

Therefore, in order to describe subjective evaluations of moving noise sources, we should make binaural measurements to obtain both the spatial factor extracted from the IACF and the temporal factor extracted from the ACF.

### 11.3 Effects of Noise and Music on Children

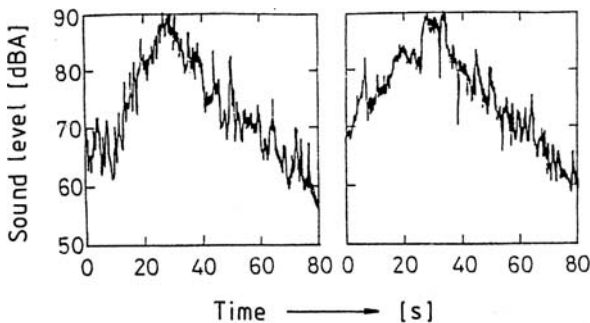
We sought to understand the differential effects of noise and music on the performance of mental tasks by children. Under conditions of quiet (no stimulus), noise, and music, children carried out cognitive tasks that are thought to lateralized in one hemisphere or the other (Table 11.5). The tasks either involved addition (left hemisphere specialized) or pattern search (right hemisphere specialized).

Tests were carried out in classrooms (the reverberation time 0.5–0.9 s in the 500-Hz octave band) of two schools in a quiet living area (Ando et al., 1975; Ando and Kang, 1987; Ando, 1988). The total number of subjects participated in the experiments was 559 (Table 11.5). The no-stimulus, quiet condition was tested in a normal classroom without any reproduced sound. The noise group was tested while being exposed to jet plane noise of  $95 \pm 5$  dBA, peak. The music group was tested while listening to an excerpt of music from the fourth movement of Beethoven's Ninth Symphony ( $85 \pm 5$  dBA, peak). As shown in Fig. 11.12, the time pattern of the

**Table 11.5** Number of subjects monitored while performing two different metal tasks

Task	Age(years)	No-stimulus group	Noise group	Music group	Total
Addition (Left-hemisphere task)	9–10	120	123	36	279
Patterns search (Right-hemisphere task)	7–8	123	119	38	280
Total subjects		243	242	74	559

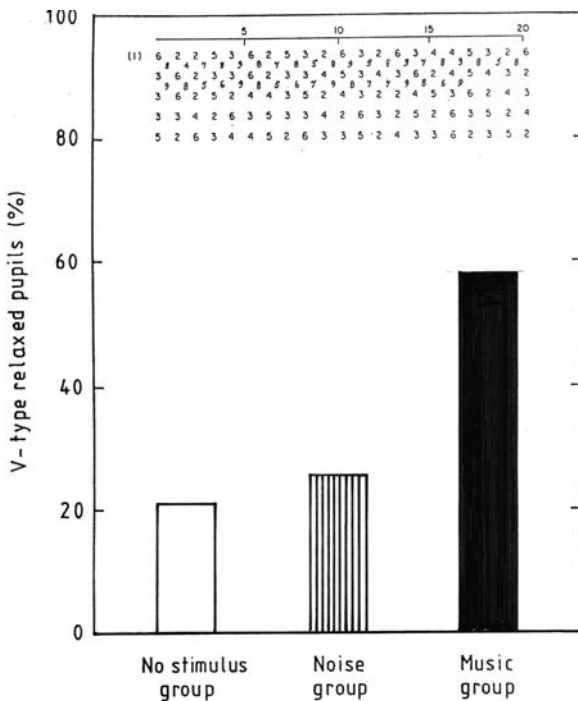
**Fig. 11.12** Sound-pressure levels of stimuli reproduced in classrooms as a function of time. *Left:* Aircraft noise adjusted by a peak of 90 dBA in this figure. *Right:* Music piece before the chorus of Beethoven's Ninth Symphony adjusted by a peak of 90 dBA



music was similar to that of the jet noise. The spectra of the two sound signals were similar also (Ando et al., 1975). Music and aircraft noise were reproduced from two loudspeakers set at the front of the classroom, during every alternative period during the addition and search tasks given by

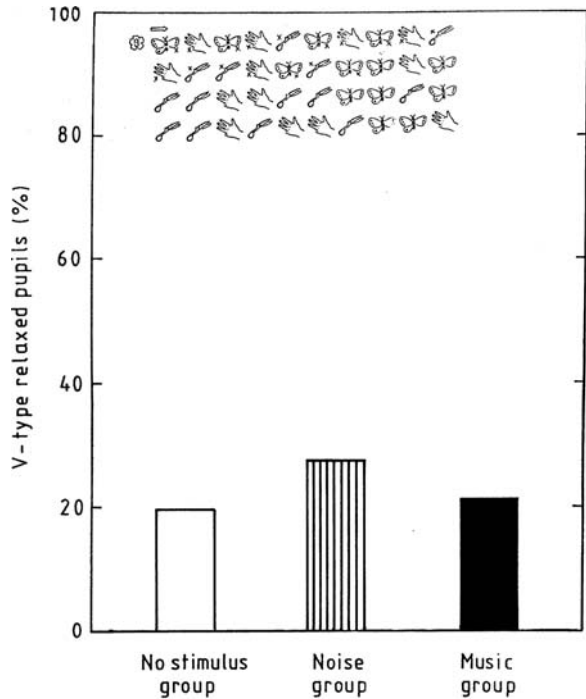
$$i = 2n \tag{11.6}$$

where  $n = 1, 2, \dots, 7$  for the adding task, and  $n = 1, 2, \dots, 5$  for the search task. Examples of one task period are shown in the upper part of Fig. 11.13 (60 s/period)



**Fig. 11.13** Proportion of V-type relaxed children during the adding task (left-hemispheric task) without any stimuli, with aircraft noise stimulus reproduced and music stimulus reproduced. The upper part indicates the task of one period (60 s) in  $N = 15$

**Fig. 11.14** Proportion of V-type relaxed children during the search task (right-hemispheric task) without any stimuli, with aircraft noise stimulus reproduced and music stimulus reproduced. The upper part indicates the task of one period (30 s) in N = 10



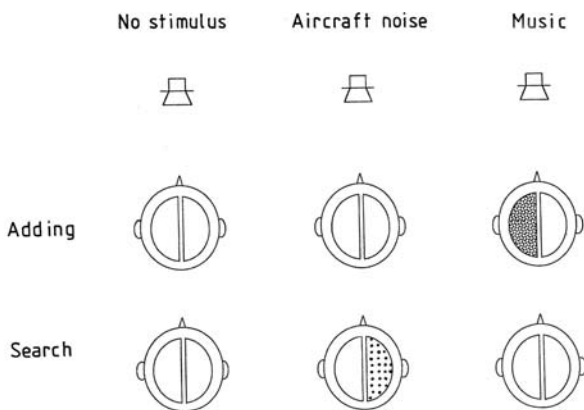
and Fig. 11.14 (30 s/period). The individual work produced in each period, called the “working curve,” was drawn for all test results. The mean work performance is not discussed here, because there were no significant differences between the different conditions. Of particular importance in evaluating the tests results is the “V-type relaxation.” This score is classified into two categories according to the occurrence of a sudden large fall in the working curve during each task. This is assessed by  $M_i < M - (3/2)W$ ,  $i = 1, 2, \dots, N$ , where  $M$  is the mean work performance and  $W$  is the average variation of the curve excluding an initial effect at the first period,  $i = 1$ . Such relaxation is thought to be caused by an abandonment of effort when mental functions are disturbed.

As shown in Fig. 11.13, the percentage of V-type relaxed children given the additional task ( $N = 15$ ) was much greater in the music group than in either the no-stimulus group or the noise group ( $p < 0.01$ ). As shown in Fig. 11.14 for pattern-search task ( $N = 10$ ), the percentage of relaxed children was similar under all test conditions, except for a slight increase in the noise group. The results of the mental tasks were not dependent on the sex, birth order, or birth weight of a child or on whether or not the mother was a working mother (Ando et al., 1975).

Significant differences in the factors  $\tau_1$  and  $\tau_e$  extracted from the running ACF of the noise and the music as a function of time may be found in measured results

(Ando, 2001b). Because of the central auditory signal-processing model (Fig. 5.1), these temporal factors may stimulate the left hemisphere activated by the fluctuation of these temporal factors. It is worth noting that the  $\tau_c$  value is deeply related to the most preferred temporal factors ( $\Delta t_1$  and  $T_{sub}$ ) of the sound field as expressed by Equations (3.3) and (3.4), which are associated with the left hemisphere (Table 5.1).

Effects of temporary music and noise stimuli on mental tasks were closely related to the content of the task being performed or to specialization of cerebral hemispheres. In the case of the addition task, there were no significant differences between the noise group and the no-stimulus group in the percentage of V-type relaxed children. This may support the theory that noise and calculation tasks are separately processed in the right and left hemispheres, respectively (Ando, 1988). Thus as illustrated in Fig. 11.15, no interference effects of the noise were evident in the adding task. However, the percentage of relaxed children in the music stimulus group differs significantly from that in the noise group and the no-stimulus group. This may be explained as an interference effect in the left hemisphere – music perception and calculation being processed sequentially in this hemisphere. On the other hand, music perception as a sequence of time and the spatial pattern task (search task) may be independently processed in the left and right hemispheres, respectively. In the search task, therefore, although no significant differences in the number of V-type relaxed children could be observed under the no-stimulus and music conditions, a difference was observed ( $p < 0.1$ ), so that interference of the noise and the search task in the right hemisphere seems to be discernible (Fig. 11.15).



**Fig. 11.15** Explanations of interference effects between mental tasks and sound stimuli by mean of the specialization of cerebral hemispheres. Aircraft noise (with less fluctuation of  $\tau_c$ ) and music (with a greater fluctuation  $\tau_c$ ), respectively, may be associated mainly with the right hemisphere and the left hemisphere. The adding task and search task, respectively, may be associated mainly with the left hemisphere and the right hemisphere

Differences in interference effects occurring during left and right hemispheric tasks, respectively, may be well described in terms of the temporal factors extracted from ACF and the spatial factors extracted from IACF as listed in Table 11.6. On the other hand, annoyance may be described by all of these factors.

**Table 11.6** Effects of noise on two difference tasks and annoyance, in relation to the temporal and spatial factors extracted from the ACF and the IACF, respectively

Factors		Effects of noise on two different tasks and annoyance		
		Left-hemispheric task	Right-hemispheric task	Annoyance
ACF	$\tau_1$	X		X
	$\phi_1$	X		X
	$\tau_e$	X		X
IACF	LL		X	X
	$\tau_{IACC}$		X	X
	$W_{IACC}$		X	X
	IACC		X	X

X: Factors may influence the corresponding task, and annoyance. LL =  $10 \log [\Phi(0)/\Phi(0)_{ref}]$ , where  $\Phi(0) = [\Phi_{ll}(0) \Phi_{rr}(0)]^{1/2}$ .

## Chapter 12

# Introduction to Visual Sensations

Part I of this book considers the basis for auditory perceptual qualities and preferences, especially as they relate to architectural acoustics. The auditory primary sensations were classified as either temporal or spatial sensations, which are described in terms of a signal-processing model of the central auditory system. The model consists of a monaural autocorrelation mechanism (ACF) and a binaural crosscorrelation mechanism (IACF). The temporal factors extracted from the ACF associated with the left hemisphere and the spatial factors extracted from the IACF associated with the right hemisphere, respectively, effectively describe temporal and spatial sensations. Such specializations of human cerebral hemispheres play an important role in making temporal and spatial factors independent with respect to subjective judgments. The auditory temporal sensations are loudness, pitch (missing fundamental), timbre, and duration sensations. The auditory spatial sensations include sound localization, apparent sound width (ASW), and subjective diffuseness (envelopment).

Part II considers perceptual qualities and preferences in vision. Visual sensations can be considered in terms of temporal and spatial sensations that are based on analogous correlation mechanisms, and similar methods can be used to assess their associated subjective preferences. The visual temporal sensation of flicker rate is analogous to auditory pitch and the missing fundamental phenomenon (Fujii et al., 2000). Considerable physiological and psychological evidence exists for temporal auto- and crosscorrelation representations in vision and in many other sensory modalities (Cariani, 2001). The visual spatial sensations include contrast, regularity, and coarseness of the focal vision (Fujii et al., 2003), as well as localization. Visual sensations in relation to temporal factors may well be described in terms of brain activities (Soeta et al., 2002a, b; Soeta et al., 2002). Table 12.1 summarizes temporal and spatial sensations in vision as well as their subjective preferences. The table provides the features of temporal and spatial autocorrelation representations that are thought to be associated with each visual perceptual attribute, as well as their corresponding neural activity observables and the sections in this volume where they are discussed.



**Table 12.1** Summary of sensations and subjective preference in relation to factors extracted from the ACF and IACF, and neurophysiological locus and observables described in Part II of this volume

Sensations and subjective preference	Acoustic correlate	Representation	Main factor	Locus	Observable(s)	Section(s)
<b>Temporal sensations and preference</b>						
Periodicity pitch	Fundamental frequency	Temporal ACF	$\tau_1$ extracted from the ACF			13.1
Subjective preference of the period of flickering light		Temporal ACF	$\tau_1, \phi_1$ extracted from the ACF	Left hemisphere	EEG, MEG	14.1, 15.1, 15.2
Subjective preference of the period of moving single target		Temporal ACF	$\tau_1$ extracted from the ACF	Left hemisphere	EEG	14.2, 15.3
<b>Spatial sensations and preference</b>						
Contrast		Spatial ACF	$\Phi(0)$ and $\delta_1$ extracted from the ACF	(Right hemisphere) <sup>1</sup>		13.2
Regularity		Spatial ACF	$\delta_1$ and $\phi_1$ extracted from the ACF	(Right hemisphere)		13.2
Coarseness		Spatial ACF	$\Phi(0)$ and $\delta_1$ extracted from the ACF	(Right hemisphere)		13.2
Subjective preference of texture		Spatial ACF	$\phi_1$ extracted from the ACF	(Right hemisphere)		14.3

<sup>1</sup>From previous works, it is considered that the spatial factors are associated with the right hemisphere.

# Chapter 13

## Temporal and Spatial Sensations in Vision

Repetition rates of periodic patterns of flickering light can be distinguished much in the same way that periodic sound patterns evoke distinct pitch percepts. As in audition, one perceives the missing fundamental frequency of a visual flicker stimulus despite the absence of any frequency component at the fundamental. As a typical temporal sensation, we shall show that the missing fundamental phenomenon in vision, like its auditory counterpart, can be predicted from the autocorrelation function ACF of the flickering light signal. In the following sections, three salient spatial sensations of texture: contrast, coarseness, and regularity are described by the spatial factors extracted from the spatial ACF of visual patterns.

### 13.1 Temporal Sensations of Flickering Light

In experiments using flickering lights, we found that a visual “pitch” is evoked at the fundamental frequency of a visual complex flicker pattern (i.e. the repetition rate of the pattern). This “pitch of the missing fundamental” phenomenon in vision is seen even in random-phase conditions. As with auditory pitch, our results also indicate that the perceived flicker rates are not intrinsically detected from the envelopes of the temporal waveforms. One promising operation to account for such a perceived periodicity is to look for features in the temporal autocorrelation function (ACF) of the real temporal waveform.

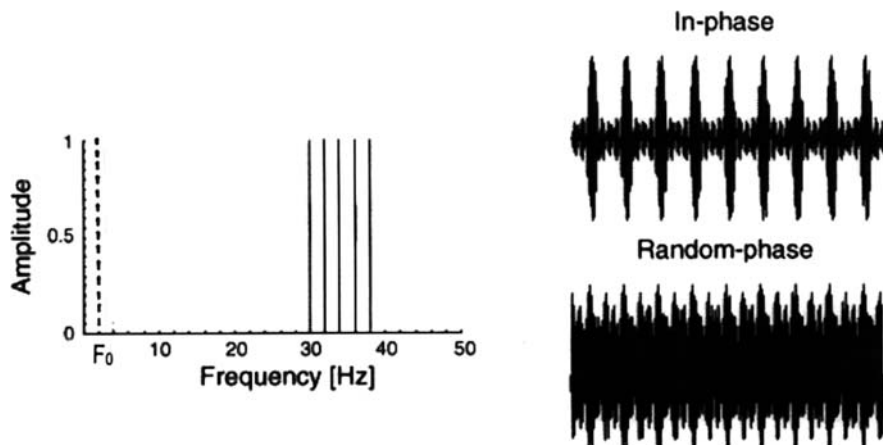
This section describes a phenomenon that is analogous to the auditory temporal sensation (see Section 6.2) called “the pitch of the missing fundamental.” When the signal contains only a number of harmonics without the fundamental frequency, we hear the fundamental frequency as a pitch. Previously, some studies in vision were related to compound waveforms (de Lange, 1952; Bowen et al., 1989; Bowen et al., 1992; Kremers et al., 1993; Eisner, 1995), in which square and sawtooth waveforms were commonly used in comparison with sinusoidal waves. Square and sawtooth waveforms each consist of the fundamental frequency (F0) and a series of sinusoidal components (harmonics). However, no studies in the temporal vision that dealt with a compound waveform without the F0 component had been performed. Missing F0 effect has traditionally only been discussed in the spatial vision literature, (e.g.,

Henning et al., 1975; Nachmias and Rogowitz, 1983). Henning et al. (1975) reported that in their simultaneous visual masking experiment that a masker consisting of only upper harmonics of a “missing” F0 could nevertheless affect detection of a sinusoidal test stimulus with energy only at F0, despite the lack of any spectral overlap between the two signals. They used 1.9 cycles per degree (c/deg) sinusoidal patterns as the test stimulus and an amplitude modulation pattern whose components are 7.6, 9.5, and 11.4 c/deg (i.e., the fourth, fifth, and sixth harmonics of the 1.9 c/deg) as the masking stimulus. That is, the missing fundamental component in the masking stimulus (1.9 c/deg) was perceived and it then disturbed the detection of the test stimulus. Nachmias and Rogowitz (1983) found similar results.

In the following experiment, we measured the subjective flicker rates for compound waveforms consisting of harmonic components without fundamental frequency F0 (Fujii et al., 2000). Because the complex components were combined linearly, there was no Fourier energy at F0. The experiment was conducted under the condition that the complex components were an in-phase and a random-phase waveform. If the perceived rates are based on the actual waveform itself, observers could not detect the rates in the random-phase condition with any clear periodicity. On the other hand, if the F0 component is perceived for both in-phase and random-phase stimuli, there may well exist a correlation mechanism to detect it. Such a correlation mechanism would be similar to the neural mechanism thought to be responsible for periodicity pitch in the auditory system.

Four subjects, ages 23–26 years old, participated in the experiment. All had normal or corrected-to-normal vision. They were well trained before starting the experiment, because they had never participated in such an experiment before. They dark-adapted for about 1 min before all sessions. The light source was a 7-mm-diameter green light-emitting diode (LED), set at a distance of 80 cm from the observer in dark surroundings. The LED stimulus field was spatially uniform, and the size of it corresponded with 0.5 deg. The stimulus waveform was produced with a 16-bit digital-to-analog converter. The mean luminance was set to 20 candela per square meter ( $\text{cd/m}^2$ ) and kept constant during the sessions. To prove the linearity of the apparatus, the luminance waveforms of the stimuli with a luminance meter (TOPCON BM-8, Tokyo, Japan) with a response time of 1 ms were measured. A nonlinear LED output was observed; however, such nonlinear components were sufficiently smaller than signal components ( $-30$  dB) in this experiment.

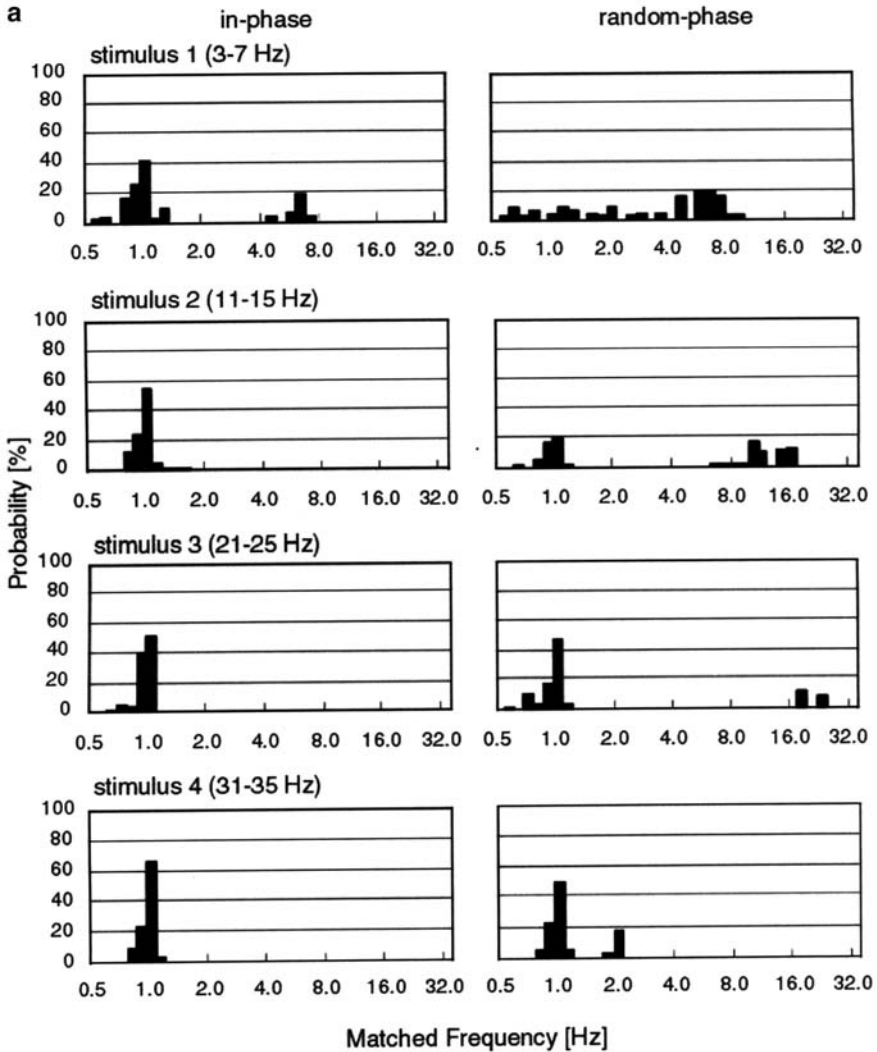
Stimuli in the current study were compound waveforms consisting of five complex components. The frequency of each component corresponded with the  $n$ -th harmonic of the fundamental frequency F0. In series A, we selected four stimuli in terms of the complex frequency range with  $F0 = 1$  Hz. Stimulus 1 consisted of 3, 4, 5, 6, and 7 Hz, and for stimuli 2, 3, and 4, the frequency ranges (11, 12...15), (21, 22...25), and (31, 32...35) Hz, respectively, were selected. In series B, for stimuli 5, 6, 7, and 8, complex components were selected in the frequency range for 30–40 Hz, in which we cannot detect any flickering rate if only a single component is presented. Stimulus 5 with  $F0 = 0.75$  Hz consisted of 30, 30.75, 31.5, 32.25, and 33 Hz. For the stimuli 6, 7, and 8 (with  $F0 = 2, 2.5,$  and  $3$  Hz, respectively), the components were (30, 32...38), (30, 32.5...40), and (27, 30...39) Hz, respectively.



**Fig. 13.1** An example of the spectrum of the complex flicker signal used in the experiment. *Left:* Complex components are 30, 32, 34, 36, and 38 Hz, where the energy of the fundamental frequency ( $F_0 = 2$  Hz) is absent. *Right:* Real waveforms for (*above*) the in-phase condition exhibiting prominent peaks corresponding to the  $F_0$  and (*below*) the random phase condition in which  $F_0$  peaks are not at all obvious

The waveforms of the complex signals used in the experiment are illustrated in Fig. 13.1. The real waveform of the stimuli was affected by the phase of components, so that the in-phase and random-phase stimuli had different waveforms. The in-phase waveform had remarkable peaks corresponding to the  $F_0$ . For the random-phase condition, each component was compounded with different phases so that the waveforms had no significant peaks.

The subjective flicker rate of the stimulus was obtained by means of the method of limits, with a reference stimulus of sinusoidal flicker. These two stimuli were presented in pairs with a blank interval. The task of observers was to judge which of these two stimuli seemed to flicker at the faster rate. As the reference stimulus, we used ascending and descending series. That was, the comparison stimulus was varied in steps, from a low frequency to a high frequency (or vice versa) to measure the value at which the observers' responses reversed. The mean of the two values before and after reversal of the observers' responses was determined as the matched frequency of the test stimulus. When the observers perceived two or more rates for one test stimulus, they were asked to judge with the rate perceived most strongly. This means that the observers matched the sinusoid to the most prominent component of the compound waveforms, and thus, one matched frequency was obtained through one trial. Intervals of the comparison stimulus were 0.1 Hz step for frequencies below 1 Hz, 0.2 Hz step for 1 to 3 Hz, and a 1 Hz step for frequencies above 3 Hz. In the descending series, trials started from a value of a few hertz above the highest frequency of the components in the test stimulus. There were two series of the comparison stimulus (ascending and descending) and two orders of presentation (test-comparison and comparison-test), giving a total of four conditions. For each



**Fig. 13.2** Subjective flicker rate of missing fundamental visual stimuli with different harmonics (a, above) and different F0 values (b, next page) for in-phase and random-phase conditions. Response probability distributions for 4 subjects. **(a)** F0 = 1 Hz **(b)** F0 = 0.75, 2, 2.5, and 3 Hz. Response probability distributions for flicker rate judgments of flicker stimuli with different missing F0s and phase spectra. From top to bottom, F0 = 0.75, 2, 2.5, and 3 Hz

condition, four trials were repeated. Thus, 16 matched frequencies were obtained for each test stimulus.

Results of the probability of responses to each matched frequency are shown in Fig. 13.2 as a histogram. For the in-phase stimuli, observers perceived the rates at F0. This frequency is easily detected, because it is consistent with the time inter-

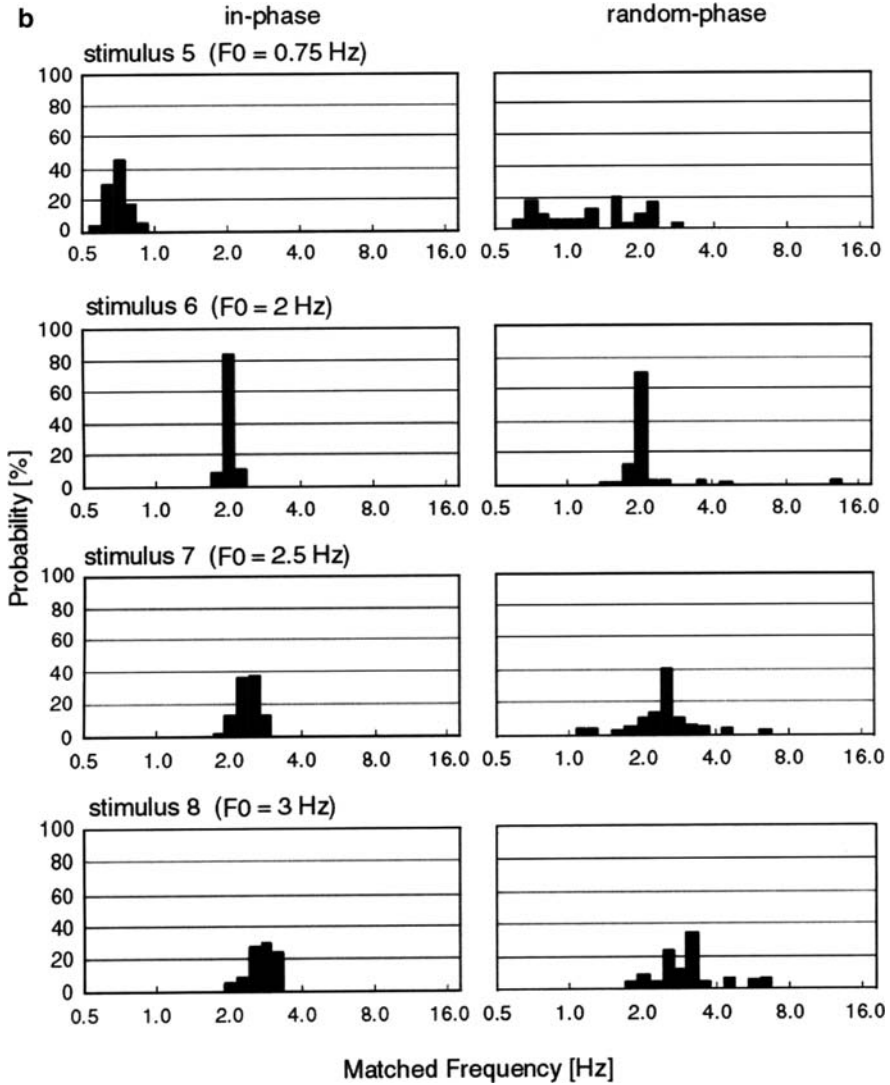


Fig. 13.2 (continued)

val between the periodic peaks appearing in the temporal waveforms as shown in Fig. 13.1. For the random-phase stimuli, matched frequencies were comparable with the several aperiodic peaks, which correspond with the component frequencies. We could detect the flicker rates from local peaks in the waveforms in this low-frequency range only (3–7 Hz). In the high-frequency range, however, the fundamental frequency  $F_0$  was perceived most frequently for both in-phase and random-phase stimuli, which is called the missing fundamental phenomenon, even allowing some exception such as certain multiples of  $F_0$ .

**Fig. 13.3** The probability of matching the flicker rate to a value within 10% of the fundamental frequency  $F_0$  as a function of the fundamental frequency. Filled circles and open circles represent in-phase and random-phase conditions, respectively

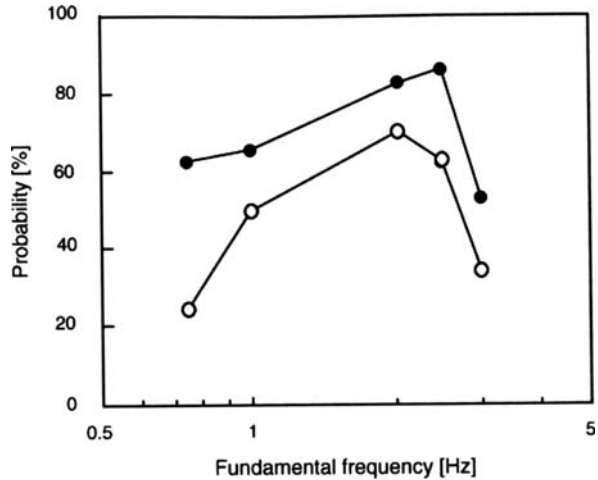
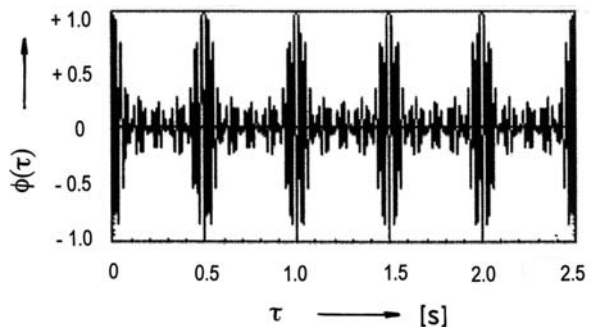


Figure 13.3 shows the observers' responses within  $(1 \pm 0.1) F_0$  as a function of the fundamental frequency  $F_0$ . Both curves have a similar profile, except that the probability was about 10% higher for the in-phase condition. Although probability was affected by the phase, the most frequently perceived rates were about  $F_0$  in all cases. The highest probability is seen at  $F_0 = 2$  and 2.5 Hz for the random-phase and in-phase conditions, respectively. These values correspond with the periods of 500 ms and 400 ms, which are similar to the "sensitive range" reported by Fraisse (1984). He reported that in the sensitive range (500 ms to 700 ms), the sensitivity increased to the periodicity of successive presentation of the stimuli. Our observers might also have responded sensitively to the periodicity of the flickering stimuli in this range. Thus, observers may detect the rates at fundamental frequency, which are not included in the power spectrum of the stimuli. One promising operation that gives the phase-independent prediction for our empirical evidence is the ACF (Fig. 13.4). Actually, the ACF of the real stimulus waveforms had identical profiles for both phase conditions used in the experiment. This result is consistent with the fact that the ACF has particular peaks corresponding to the  $F_0$ . One can postu-



**Fig. 13.4** An example of the temporal ACF of the flicker stimuli for both conditions, in-phase and random phase. The value of  $\tau_1$  corresponds to the "missing" fundamental frequency, which has no energy in the power spectrum representation

late the existence general neural mechanisms that are common to both auditory and visual systems that could analyze temporal patterns of interspike intervals to infer the fundamental period from these autocorrelation-like neural representations.

In the experiment, the observers' responses at  $F_0$  were slightly affected (about 10%) by phase (Fig. 13.3), and some responses were seen at multiples of  $F_0$  with the random-phase stimuli. (i) This difference may be explained simply by the difference of the waveform. There is a periodic structure in the in-phase condition, but such a periodicity disappeared in the condition of the random phase. (ii) A second possibility is that such a phase effect could be explained by the ACF if there is nonlinear distortion of the stimulus waveforms. The ACF would not be identical for in-phase and random-phase stimuli if there is substantial nonlinearity in the visual-brain system.

### **13.1.1 Conclusions**

1. The most frequently perceived flicker rate for complex visual waveforms corresponds to the fundamental frequency, even when there is energy at the fundamental in the power spectrum (the "missing" fundamental is perceived for periodic visual flicker patterns as well as for sounds).
2. Even though the fundamental periodicity is not readily apparent in the waveform plot of the random-phase flicker stimulus, the flicker fundamental is still perceived (again in a manner similar to auditory pitch perception).
3. These phenomena can be explained by assuming a process that detects periodic peaks in the ACF of stimuli for a certain range of fundamental frequencies.

## **13.2 Spatial Sensations**

To understand the mechanism of our visual system, an attempt is made here to characterize the information utilized for texture perception. The spatial ACF analysis of a two-dimensional pattern provides useful measures for representing the three salient spatial sensations of texture: contrast, coarseness, and regularity. The validity of the ACF analysis was examined by use of the calculated factors to describe the subjective scale value of each sensation collected for various kinds of natural textures. When a texture has harmonic structure, for example, the first major peak in the estimated spatial autocorrelation function corresponds to the spatial period of the texture. Results show that these spatial sensations are strongly related to factors extracted from the spatial ACF.

The underlying problem is how does the human visual system characterize salient spatial sensations of texture? Several previous studies have focused on texture properties either through the study of perception or through computational models. However, the relationship between perceptual and computational properties remains unclear. To understand the mechanism of our visual system, it is important to know how the information related to sensations of the textural property may be extracted.



This section contributes to an understanding of the operation of the human visual system, as well as to computational vision.

Historically, the general texture model that was proposed by Julesz (1962) mainly drove approaches. The deficiency of the structural methods is that they are incapable of capturing the randomness-like natural textures. In the decade of the 1970 s, two conflicting prediction models of hierarchical feature detection and multiple spatial frequency channels were discussed (de Valois and de Valois, 1980). Textural features were discussed by first-order and second-order statistics of the image, such as average and variance, or a co-occurrence matrix (Haralick, 1979). Since the 1980 s, the study of texture models has been based on random field statistics (Cross and Jain, 1983; Mao and Jain, 1992) such as the Markov random field and autoregression models. These models can describe texture by using a small number of parameters. These parameters, however, are not based on the mechanism of the human visual system, and therefore they can hardly represent multiple textural property salience to the visual sensations.

More recent texture models are motivated by the psychophysical and physiological findings regarding the early stages of the human visual system. These texture models are based on the psychophysical evidence that two textures are often difficult to discriminate when they produce a similar distribution of responses in a bank of orientation and spatial-frequency selective linear filters (Turner, 1986; Malik and Perona, 1990). In most of the articles (for example, Heeger and Bergen, 1995; Zhu et al., 1998), the texture is parameterized by the first-order statistics, for example, histograms of a set of appropriately chosen multiscale and multiorientation filter outputs. The application of these sets of parameters to the texture synthesis has been successful in generating a new texture that matches the appearance of a given sample. The main disadvantage of this method is the difficulty in synthesizing periodic or structured textures. Another problem that remains to be solved is how to select a set of features to best characterize the images and how to integrate the output of many filters for further processing.

On the other hand, Tamura et al. (1978) measured human discrimination performance with texture features and compared these measures with computational results. From the description in the literature and from observations of the Brodatz (1966) texture album, six features were chosen that correspond with visual perception: contrast, coarseness, regularity, roughness, directionality, and line-likeness. The emphasis was on developing computational analogues of these attributes, but they only modified already developed computational features and combined several features to have a close relationship to a specific property. As a result, this approach failed to clarify how visual properties are related to particular physical properties. Amandasun and King (1989) carried out a similar study. Rao and Lohse (1996) developed a classification method for perceptual texture. Based on psychological similarity judgments, they constructed a three-dimensional space for texture classification. The three orthogonal dimensions were repetitive versus nonrepetitive; high-contrast and nondirectional versus low-contrast and directional; and granular, coarse, and low-complexity versus nongranular, fine, and high-complexity. A similar experiment conducted by Cho et al. (2000) suggested that the dimensionality of

perceptual texture space was at least four. They described four orthogonal attributes, namely, coarseness, regularity, contrast, and lightness.

The studies that were reviewed suggested that the important properties of perceptual texture could be described by using three or four independent factors. Among these orthogonal dimensions, the importance of perceived regularity or periodicity is emphasized in most of the studies. Hence, it is desirable for texture modeling to describe regularity. Actually, there are several studies emphasizing the importance of periodic structure in texture modeling and synthesis. Francos et al. (1993) decomposed the texture spectrum into harmonic (periodic), evanescent (directional), and indeterministic (random) components for texture synthesis. Liu and Picard (1996) extended a similar method for an image retrieval system. Their methods treat periodic and random components separately and succeeded in representing many kinds of natural textures. Portilla and Simoncelli (2000) included local autocorrelation in their synthesis parameter to represent periodic structures in a texture.

The role of periodic structure has also been examined in vision research. Uttal (1975) emphasized the importance of periodic structure in form perception. He conducted an experiment of form detection with dot patterns, measuring the detectability of dot patterns masked by noise patterns. He found that the regularity of dots in a pattern increased the detectability of the pattern. The autocorrelation principle was applied to explain this result. In his autocorrelation model, the periodicity of the signal is emphasized and random noise is minimized. As a result, only the periodic signal is extracted and then perceived as a form. Recently, Ben-Av and Sagi (1995) presented the ACF model for perceptual grouping. They used matrix patterns of discrete elements to quantify the grouping law. The elements were arranged in arrays giving rise to two possible perceptual organizations, either horizontal rows or vertical columns. In the model, the ACF was calculated in vertical and horizontal directions. Perceptual grouping occurred in the direction in which a higher degree of correlation appeared. Their results imply that there is a correlation mechanism involved in the process of perceptual grouping. They proposed a possibility that orientation filters (e.g., Gabor filters) with long-range interactions between them could be used for the estimations of directional autocorrelation.

The underlying problem is how does the human visual system characterize salient spatial sensations of the texture? Several previous studies have focused on the texture properties either through the study of perception or through computational models. However, the relationship between perceptual and computational properties remains unclear. To understand the mechanism of our visual system, it is important to know how the information related to sensations of the textural property may be extracted from the spatial ACF.

The ACF of a two-dimensional texture pattern is defined by

$$\Phi(\Delta x, \Delta y) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} p(x, y) p(x + \Delta x, y + \Delta y) \quad (13.1)$$

where  $p(x, y)$  is the input signal and  $p(x + \Delta x, y + \Delta y)$  is the spatially shifted version of the input. Integers  $M$  and  $N$  refer to the data size in horizontal and vertical

directions. In this study, it was selected as  $M = N = 256$ . For computational efficiency, the ACF is computed as an inverse fast fourier transform (FFT) of the image power spectrum. Usually, the ACF is normalized as

$$\phi(\Delta x, \Delta y) = \Phi(\Delta x, \Delta y) / \left( \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} p(x, y)^2 / MN \right). \quad (13.2)$$

The denominator in Equation (13.2) is the maximum value of the ACF representing the power of the image. Thus, the normalized ACF has the maximum value of 1.0 at the origin,  $\Delta x = \Delta y = 0$ .

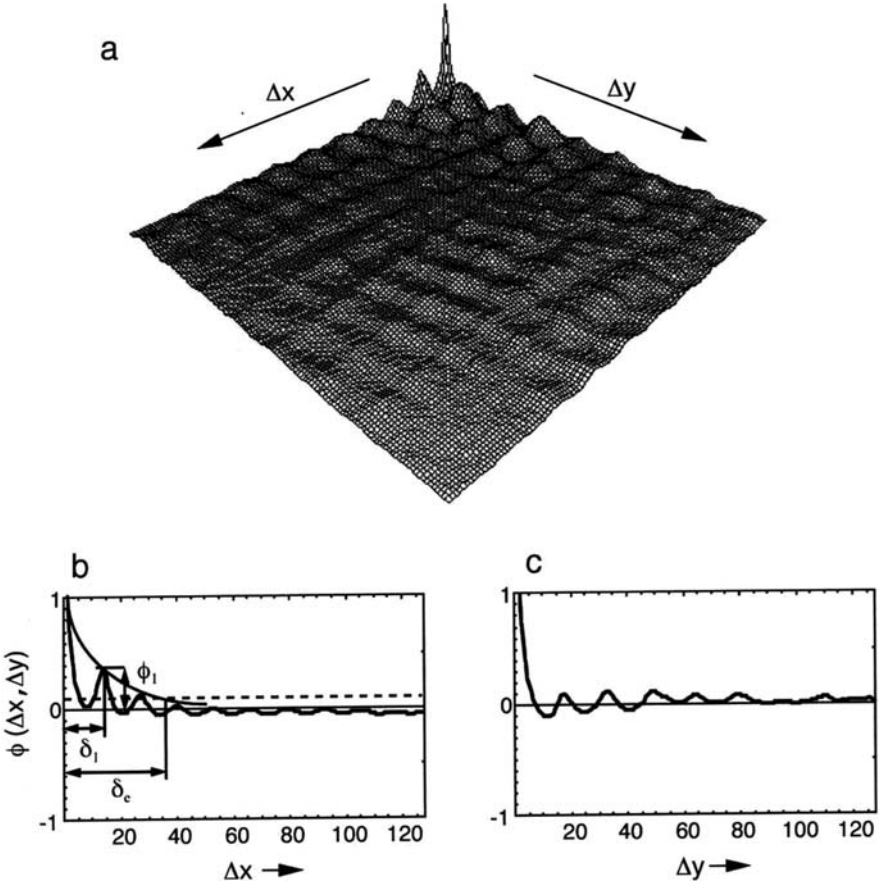
The ACF contains the same information as the power density spectrum. We use the ACF here because it allows us to most simply and directly obtain the major factors involved in spatial texture perception. The spatial ACF reflects the distribution of spatial intervals in the image, that is, the magnitudes for each interval length, as opposed to frequency, in spite of the formal interchangeability of two descriptions. Remarkably, the spatial ACF of the pattern has a dominant periodicity that corresponds to the reciprocal of its fundamental frequency. Even when the fundamental frequency component is subtracted from the stripe pattern, due to the missing fundamental, we can still see the fundamental periodicity (Henning et al., 1975). The frequency domain analysis fails to explain this percept, because the fundamental frequency component is absent. Instead, the ACF of this missing fundamental pattern maintains a dominant periodicity at the fundamental. The perceptual mechanism of the pattern is still under discussion (Badcock and Derrington, 1989; Hammett and Smith, 1994), however, it implies that the ACF is effective for detecting the spatial periodicity of patterns. 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 crosscorrelation that had been proposed 3 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.

From the calculated ACF, four factors may be extracted from the ACF in a similar manner as that described in Chapter 5. As shown in Fig. 13.5, the ACF decays from the origin outwards. To simplify the calculation, we only considered two one-dimensional ACFs along the x and y axes from the origin. This is because the most natural texture is considered to be *isotropic*, in which all orientations occur with the same probability. Consequently, the ACF could be assumed to be circular symmetric. Even for the *anisotropic* texture, at least one of our directional ACFs can hold the periodic structure. The four factors determined are

1.  $\Phi(0,0)$ : the autocorrelation at the origin,  $\Delta x = \Delta y = 0$ .
2.  $\delta_1$ : the displacement of the first maximum peak in the ACF.
3.  $\phi_1$ : the amplitude of the first maximum peak.

4.  $\delta_e$ : the effective range or effective spatial duration of the ACF (defined as the displacement at which the envelope of the normalized ACF decayed at 0.1). The factor  $W_{\phi(0)}$  is not included here because it was not found yet at the time of the study.

Only the factor  $\Phi(0,0)$  was calculated from the two-dimensional ACF, that is, the denominator in Equation (13.2), and it is shown in the dB scale  $[10\log_{10}\Phi(0)]$ . The algorithm calculating the other three factors is as follows. First, the directional



**Fig. 13.5** An example of the spatial ACF analyzed and factors extracted from the ACF. (a) Two-dimensional ACF of texture D3 in Fig. 13.6a. (b) One-dimensional spatial ACF for the x direction. (c) One-dimensional spatial ACF for the y direction. Symbol  $\phi(\Delta x, \Delta y)$  signifies the spatial ACF for the spatial intervals  $\Delta x$  and  $\Delta y$ ,  $\phi_1$  is the amplitude of the shortest major spatial interval,  $\delta_1$  is the length of shortest major spatial interval, and  $\delta_e$  is the effective range of significant intervals. By comparing the amplitudes and shortest spatial intervals in both directions, we extracted three factors from the ACF for the x direction only. Note that the x direction indicates more complicated spatial structures throughout the texture applied here

ACFs along the  $x$  and  $y$  axes are extracted from the two-dimensional ACF. Then, the first maximum peak value in both directions is figured out. From the chosen ACF, the displacement and the amplitude of the maximum peak are picked out as  $\delta_1$  and  $\phi_1$  (see Fig. 13.5b). The value of effective spatial duration  $\delta_e$  is determined by the point at which the normalized spatial ACF decays to 10% of its zero-lag maximal value, as shown in the figure. This is done by exponential fit to normalized ACF peaks above 0.1. When the given ACF has no peaks above 0.1, only the initial decay rate was fitted.

The value of  $\Phi(0)$  corresponds to the energy of the image. The displacement and the amplitude of the maximum peak ( $\delta_1$  and  $\phi_1$ ) are related to the periodicity of the image. The value of  $\delta_1$  is a reciprocal of the fundamental frequency, and  $\phi_1$  is the strength of the harmonic components. The value of  $\delta_1$  represents the spatial pitch, and the value of  $\phi_1$  represents the strength of the spatial pitch. The image width (cm) and the distance (cm) calculate the visual angle in degrees.

Materials applied are shown in Fig. 13.6a (Brodatz, 1966, Tamura et al., 1978, Fujii et al., 2003). Adding 12 botanical textures with a high degree of similarity between them and less harmonic structure (Fig. 13.6b), we discuss the spatial sensations in relation to multiple factors extracted from their spatial ACFs (Fujii and Ando, unpublished data).

According to the multiple regression analysis, the scale value (SV) of each spatial sensation obtained by the PCT was approximately described by the spatial factors extracted from the ACF of the spatial signal of gray level, so that

$$S_{\text{contrast}} \approx c_1 \Phi(0) + c_2 \delta_1 \quad (13.3)$$

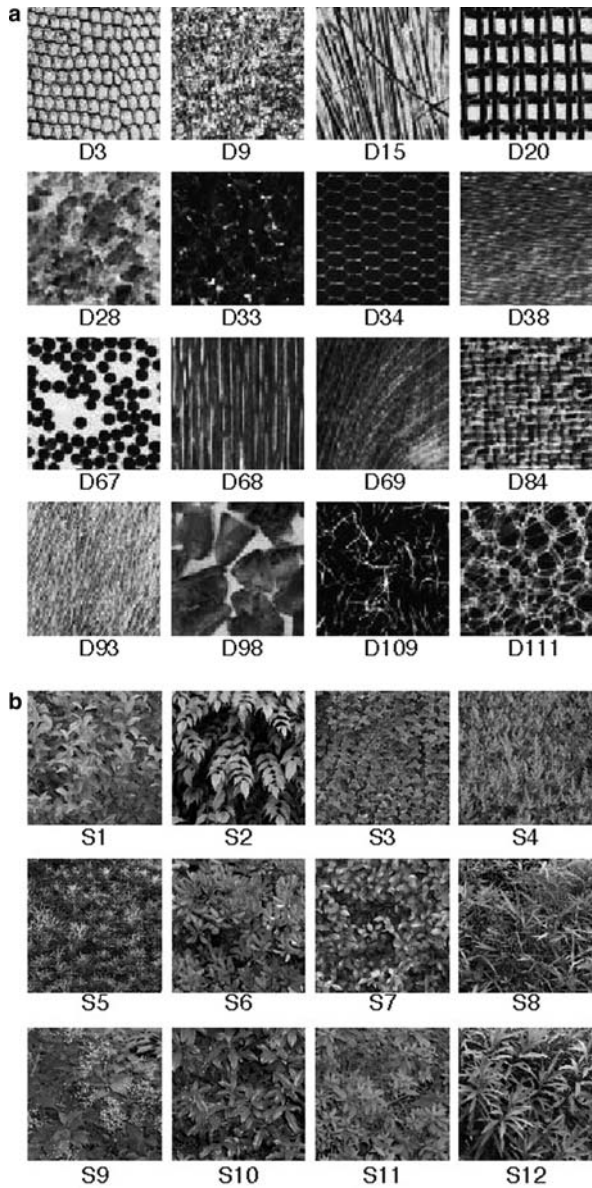
$$S_{\text{coarseness}} \approx c_3 \Phi(0) + c_4 \delta_1 \quad (13.4)$$

$$S_{\text{regularity}} \approx c_5 \phi_1 + c_6 \delta_1 \quad (13.5)$$

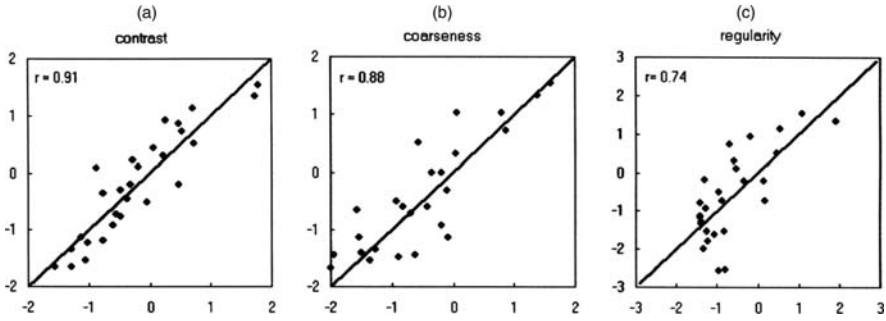
where  $c_1, c_2, \dots, c_6$  are coefficients. As a result of analyses of measured data, coefficients were obtained:  $c_1 = 0.85$ ,  $c_2 = 0.25$ ,  $c_3 = 0.43$ ,  $c_4 = 0.37$ ,  $c_5 = 0.55$ , and  $c_6 = 0.74$ . Figure 13.7 shows relationships between the measured SV (vertical axis) and those (horizontal axis) calculated by Equations (13.3) through (13.5). The correlation coefficients between the measured SV and calculated SV are  $r = 0.91$  (contrast),  $0.88$  (coarseness), and  $0.74$  (regularity).

Thus far, we have clearly shown that the ACF analysis provides useful measures for representing the textural property of contrast, coarseness, and regularity. Multiple regression analysis revealed that perceived contrast is affected by  $\Phi(0)$  and  $\delta_1$ . The simplest explanation of this observation might be the contrast sensitivity function of the human visual system (Campbell and Robson, 1968). Human spatial contrast sensitivity is best at a mid-spatial frequency range around 2–5 c/deg. Perceived regularity was strongly related to the period  $\delta_1$  and height of the peak  $\phi_1$ . As for the random texture, however, the estimated ACF does not have a periodic structure. In this case, the initial decay rate of the ACF was important to represent the texture coarseness as well as regularity. It seems that there are two distinct mechanisms in

**Fig. 13.6** (a) Texture set used by Tamura et al. (1978). (b) Botanical texture set applied for the additional experiment (Fujii et al., 2003)



the human visual system to characterize coarseness and regularity. A recent study also reported that the human visual system might use different cues in the discrimination of texture coarseness depending on the image structure. Sakai and Finkel (1995) presented evidence that the visual system appears to track either mean frequency or peak frequency, depending on the strength of harmonic components in



**Fig. 13.7** Relationships between estimated and measured scale values for visual contrast, coarseness, and regularity. Scale values were estimated using Equations (13.3)–(13.5). Scale values and correlations between estimated and measured values for (a) contrast ( $r = 0.91$ ), (b) coarseness ( $r = 0.74$ ), and (c) regularity ( $r = 0.88$ )

the image. That is, the peak frequency is detected in the presence of strong peak, and the mean frequency is detected in their absence. In the context of our analysis, the peak frequency and mean frequency in the spectrum could be comparable with the maximum peak and the initial decay rate of the ACF. For the similar purpose as our study, Liu (1997) measured the strength of perceived regularity by use of spectrum decomposition. The image was decomposed into the deterministic (periodic) and nondeterministic (random) components in the frequency domain.

What we have done here is just to show how the ACF factors can represent texture sensations and to a certain degree how human subjects may be sensitive to the same parameters captured by the ACF. There is convincing evidence that such a

**Table 13.1** Autocorrelation representations in temporal and spatial vision. Temporal and spatial sensations in relation to factors extracted from the temporal ACF and the spatial ACF

ACF factors	Temporal sensation	Spatial sensations		
	Missing fundamental	Contrast Equation (13.3)	Coarseness Equation (13.4)	Regularity Equation (13.5)
Temporal ACF				
$\tau_1$	$X^1$ ( $F = 1/\tau_1$ )			
$\phi_1$	$X$			
$\tau_c$	$X^2$			
Spatial ACF				
$\Phi(0)$		$c_1 \Phi(0)$	$c_3 \Phi(0)$	
$\delta_1$		$c_2 \delta_1$	$c_4 \delta_1$	$c_6 \delta_1$
$\phi_1$				$c_5 \phi_1$

<sup>1</sup>X: Factors to describe each sensation.

<sup>2</sup>X: Attention should be paid due to the fact that factors of  $\phi_1$  and  $\tau_c$  are mutually related.

mechanism also exists for visual processing (Uttal, 1975; Ben-Av and Sagi, 1995). Whatever the mechanism, it is possible that humans do have access to the correlation structure as extracted by the ACF. It is remarkable that the correlation model in vision can be applied both to explanation of temporal and spatial sensations.

Thus, the so-called seamless correlation model may exist in the brain for auditory and visual sensations. Table 13.1 summarize results of the temporal and spatial sensations of the visual field, which are described in relation to the temporal and spatial factors extracted from the ACF analyzed.



# Chapter 14

## Subjective Preferences in Vision

In order to rationally design visual objects and spaces with both temporal and spatial sensations in mind (“temporal design”), we want to understand subjective preferences for visual stimuli. A series of experiments was carried out to probe the subjective preferences for temporal and spatial factors in vision. This included preferences for rates of flickering lights, for frequencies of oscillating vertical and horizontal movements, and for texture-related spatial regularities. The temporal and spatial factors are extracted respectively from temporal and spatial autocorrelation functions (ACFs).

### 14.1 Subjective Preferences for Flickering Lights

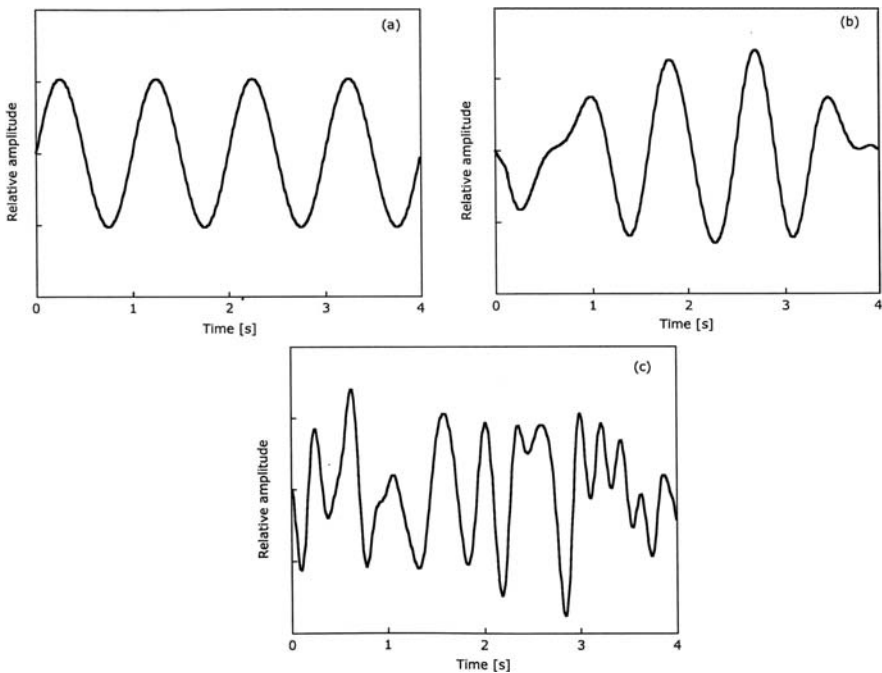
Subjective preference for a flickering light obtained using paired comparison tests (PCT) was investigated in terms of the ACF factors of the signal in the time domain. It has been found that the preferred period of the flickering light with a sinusoidal signal is approximately 1 Hz (Soeta et al., 2002a). The purpose of this study was to find more preferred conditions introducing a fluctuation to the flickering light. For this purpose, the bandwidth of the noise centered on 1 Hz was varied at five levels (1, 2, 4, 8, and 16 Hz) by use of the second-order Chebyshev filter. A remarkable finding is that the most preferred value of  $[\phi_1]_p$  is about 0.46, where  $\phi_1$  can be extracted from the temporal ACF of the stimulus signal. The scale value of preference is formulated approximately in terms of the  $3/2$  power of the normalized  $\phi_1$  of a flickering light by the most preferred value,  $[\phi_1]_p$ . This result may suggest a reason, for example, why one likes the twinkling star. Further, we may produce a visual light and even a music signal based on this temporal factor, and also blending visual light and music.

In the natural environment, for example, we have many visual aspects in the temporal fluctuation, such as leaves in the wind and clouds in the sky, twinkling stars due to air currents, flames, and flows of water in a river. Flames in a bonfire and a glitter of sunlight reflected by the water surface provide us with a lively and splendid environment.

Although a number of studies have dealt with sensitivity to flickering stimuli (e.g., de Lange, 1952; Kelly, 1961; Mandler and Makous, 1984; Kremers et al.,

1993; Wu et al., 1996), this section is concerned with subjective preference in time-varying light. Subjective preference was initially chosen as a primitive and overall response relating to aesthetics. It would lead the individual away from inappropriate environments and toward desirable ones (Schroeder et al., 1974; Kaplan, 1987).

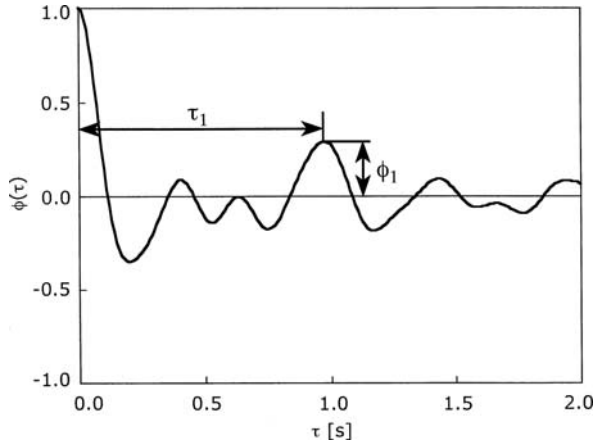
It has been found that the preferred period of a sinusoidally-flickering light is approximately 1 Hz. To obtain more preferred conditions, we introduced a fluctuation to the sinusoidal flickering light centered on 1 Hz (Soeta et al., 2005). The amplitude of the first maximum peak of the ACF,  $\phi_1$ , was controlled by changing the bandwidth of the noise (1, 2, 4, 8, and 16 Hz). This temporal factor that reflects flicker regularity is the visual analogue of auditory pitch strength. An 8-mm-diameter green LED, set at a distance of 1.0 m from the subject in dark surroundings, produced the light source. The stimulus field from the LED was spatially uniform, and its size corresponded with  $0.46^\circ$  of the visual angle. The mean luminance was set to  $30 \text{ cd/m}^2$  and kept constant during the sessions. Examples of the stimulus signal are shown in Fig. 14.1.



**Fig. 14.1** Examples of stimuli used. (a) Sinusoidal wave. (b)  $\Delta f = 1 \text{ Hz}$ . (c)  $\Delta f = 4 \text{ Hz}$

The source signal was characterized by the ACF factors,  $\Phi(0)$ ,  $\phi_1$ ,  $\tau_1$ , and  $\tau_c$  (Fig. 14.2,  $2T = 4.0 \text{ s}$ ). The value of  $\Phi(0)$ , representing average power, was set constant. The  $\tau_1$  corresponding to the center frequency of the band-pass noise, was fixed at 1.0 s (Fig. 14.3a). Note that the factor  $W_{\phi(0)}$  is constant also. The values of  $\phi_1$  and  $\tau_c$  increase as the bandwidth of noise,  $\Delta f$ , decreases (Fig. 14.3b and c). Because

**Fig. 14.2** An example of the normalized autocorrelation function NACF of a signal showing definitions of dominant periodicity  $\tau_1$  and its relative magnitude  $\phi_1$ . For acoustic signals these features correspond respectively to pitch and pitch strength or salience



there is a certain degree of coherence between  $\phi_1$  and  $\tau_c$ , subjective preference for a flickering light is discussed based only on  $\phi_1$  in this study.

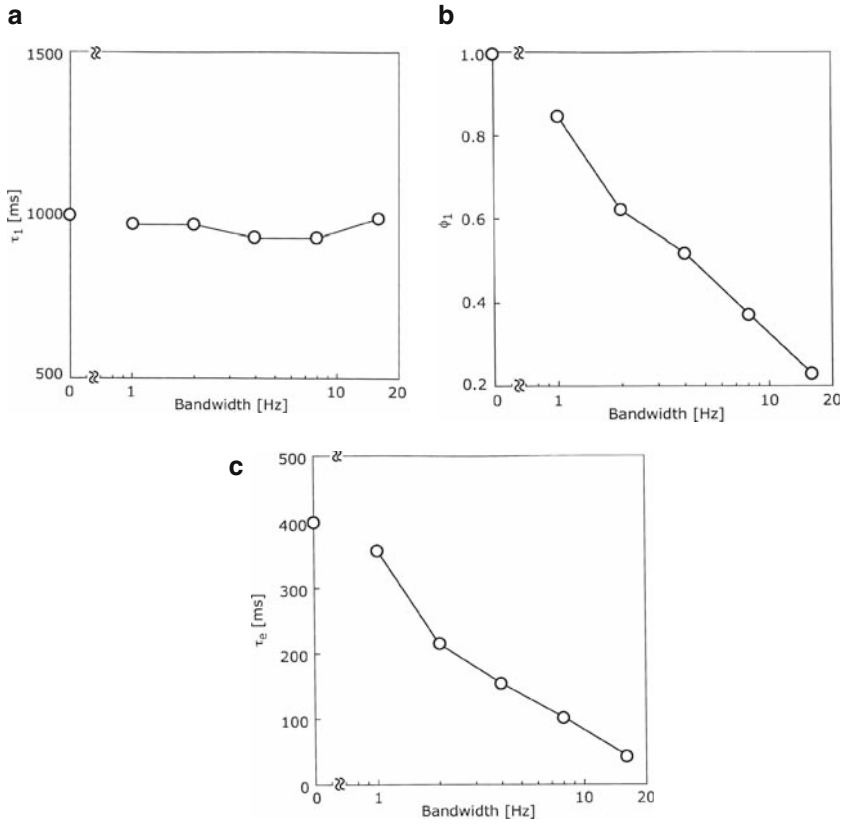
Ten 20- to 23-year-old subjects participated in the experiment. All had normal or corrected-to-normal vision. They adapted to the dark and watched the LED stimulus. Then, the PCT was performed for all combinations of the pairs (i.e., 15 pairs of stimuli interchanging the order in each pair per session), and the pairs were presented in random order. A total of 10 sessions were conducted for each individual subject. The subjects were asked which stimulus they preferred to watch.

The most preferred,  $[\phi_1]_p$ , for each subject was obtained by fitting a suitable polynomial curve to a graph on which scale values were plotted. Figure 14.4 shows an example of the method used for estimating  $[\phi_1]_p$ . The peak of this curve denotes the subject's most preferred value. Table 14.1 shows the results of  $[\phi_1]_p$  for the 10 subjects. The averaged value of  $[\phi_1]_p$  was 0.46.

We also attempted to determine the characteristics of the preference evaluation curve in more detail, in a similar manner to those described in Sections 3.2 and 7.3. The preference evaluation curve can be expressed widely as

$$S = S_L \approx -\alpha|x|^\beta \tag{14.1}$$

where  $\alpha$  and  $\beta$  are the weighting coefficients, and  $x = \log\phi_1 - \log[\phi_1]_p$ . After obtaining the most preferred value  $[\phi_1]_p$  for each subject, values of  $\alpha$  and  $\beta$  were obtained. To simplify Equation (14.1), the coefficient  $\beta$  may be fixed at a certain value so that the individual preference curve can be expressed by the sole coefficient  $\alpha$ . The average value of  $\beta$ , estimated by a quasi-Newton numerical method, was approximately 1.47, as shown in Table 14.2. As discussed later, the scale value of preference for the period of a flickering light and the circular stimulus moving in vertical and horizontal directions is also formulated approximately in terms of the 3/2 power of the normalized period. Thus, the value of  $\beta$  was fixed at 3/2 here again, so that the coefficient  $\alpha$  can represent individual difference in subjective preference. The



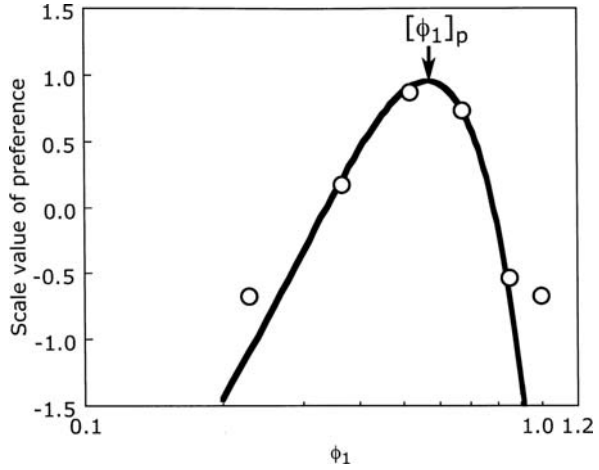
**Fig. 14.3** Measured factors extracted from the temporal ACF of the visual source signal as a function of the bandwidth  $\Delta f$ : **(a)** Delay time of the first maximum peak of ACF ( $\tau_1$ ). **(b)** Amplitude of the first maximum peak of ACF ( $\phi_1$ ). **(c)** Effective duration of ACF ( $\tau_e$ ). The  $\tau_e$  of the pure tone is  $\infty$

values obtained by a quasi-Newton numerical method are listed in Table 14.3. The weighting coefficient  $\alpha$  describes the sharpness of the preference curve with respect to  $\phi_1$ . The large  $\alpha$  value signifies that the subject clearly differentiates the level of preference.

Figure 14.5 shows scale values for all of the subjects and the preference evaluation curve calculated by Equation (14.1). The correlation coefficient between scale values of preference and calculated values by Equation (14.1) is 0.92 ( $p < 0.01$ ). Remarkably, the resulting preferred fluctuation expressed by  $\phi_1$  is an intermediate of the values between those of sine and a perfectly random wave.

The scale value of preference is formulated commonly in terms of the  $3/2$  power of  $x$  of a flickering light. This behavior is consistent with preference judgment for a flickering light without fluctuation (Soeta et al., 2001); a circular target

**Fig. 14.4** An example of obtaining the most preferred flicker regularity value,  $[\phi_1]_p$  ( $\approx 0.58$ ), for a single subject. The scale value at  $\phi_1 = 1$  is not taken into consideration in the curve fitted because the decline of preference saturated already at  $\phi_1 = 0.85$  in this case



**Table 14.1** The most preferred regularity value for flickering light  $[\phi_1]_p$  for each observer and the averaged value

Observer	$[\phi_1]_p$
A	0.51
B	0.50
C	0.47
D	0.58
E	0.45
F	0.90
G	0.27
H	0.33
I	0.33
J	0.31
Averaged	0.46

**Table 14.2** Values of  $\alpha$  and  $\beta$  for each observer obtained for Equation (14.1). It is worth noting that the average value of  $\beta = 1.47 \approx 3/2$ . When  $\beta$  is fixed at  $3/2$ , the individual differences may be expressed by the constant  $\alpha$  (the averaged value of  $\alpha$  is 10.98)

Observer	$\alpha$	$\beta$
A	9.19	1.17
B	4.98	0.72
C	7.98	1.28
D	11.98	1.39
E	11.42	1.36
F	5.53	1.36
G	6.94	1.46
H	24.72	2.37
I	14.93	2.41
J	6.57	1.23
Averaged		1.47

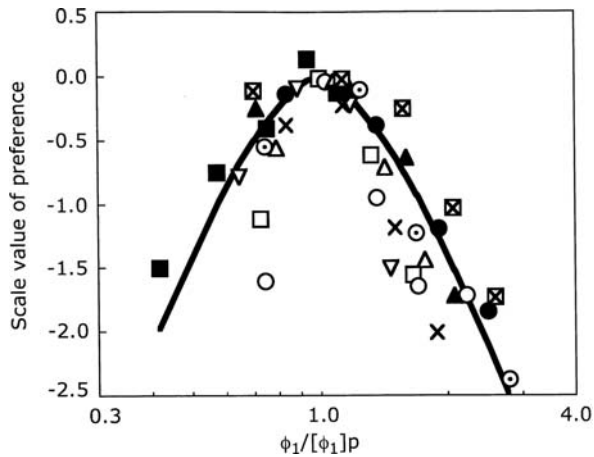
**Table 14.3** The value of  $\alpha$  obtained for each observer representing the individual difference and the averaged value. When the value of  $\beta$  is fixed at  $3/2$  in Equation (14.1), then individual differences may be expressed by the constant  $\alpha$

Observer	$\alpha$
A	15.99
B	18.00
C	11.02
D	14.45
E	13.82
F	6.49
G	7.27
H	8.32
I	5.88
J	8.53
Averaged	10.98

moving in the vertical and horizontal directions (see Section 14.2) and matching the tonal tempo of windblown camphor leaves (see Section 16.1.1). Thus, Equation (14.1) represents the preference evaluation curve, as similar to that of the sound field.

It has been reported that the power spectrum of natural images tends to behave regularly, with values roughly corresponding to  $1/f_s^2$ , where  $f_s$  is the spatial frequency (Field, 1987; Runderman and Bialek, 1994; Schaaf and Hateren, 1996), and the temporal power spectrum of natural time-varying images is given by  $1/f_t^2$  on the temporal frequency  $f_t$  (Dong and Atick, 1995). Previous studies showed that the luminous patterns of fireflies and candlelight have a  $1/f^n$  fluctuation mode (Doi et al., 1997; Inagaki et al., 2001). It has also been reported that the spectral density of fluctuations in the audio power of many musical selections and of English speech varies approximately as  $1/f$ , where  $f$  is the frequency (Voss and Clarke, 1978a,b). Though spectral pattern may be one factor influencing subjective preference, the stimulus used in this study does not have any  $1/f^n$  fluctuation mode. Many repre-

**Fig. 14.5** The normalized scale value of preference for all subjects. The solid curve is the value calculated by Equation (14.1) with constants  $\alpha = 10.98$  and  $\beta = 3/2$  (Table 14.2). Different symbols indicate scale values obtained with different subjects. The abscissa is normalized by  $[\phi_1]_p$ . The scale value at  $[\phi_1]_p$  is adjusted to zero, without loss of any generality



sentations have been traditionally based on recognizing spectral patterns; however, pitch of complex tones and timbre, for example, can be well described by temporal factors extracted from the ACF. It is worth noting that sensory information from other modalities besides audition and vision can be handled using similar autocorrelation and crosscorrelation neurocomputational frameworks (Cariani, 2001).

## 14.2 Subjective Preferences for Oscillatory Movements

Preference judgments using the PCT for sinusoidal movements of a single circular target without any fluctuation on a monitor screen were performed. The period of stimulus movements was varied separately in the vertical or horizontal direction. Results show that the most preferred periods ( $[T]_p$ ) for all subjects are about 1 s in the vertical direction and about 1.3 s in the horizontal direction. The curve of the scale values of preference may be commonly expressed by Equation (14.1) with  $x = \log_{10}T - \log_{10}[T]_p$  and  $\beta = 3/2$ . All observers participating in the vertical direction series showed that the curves for the scale value of preference are significantly steeper in the fast-moving range in reference to  $[T]_p$  than those in the slow-moving range.

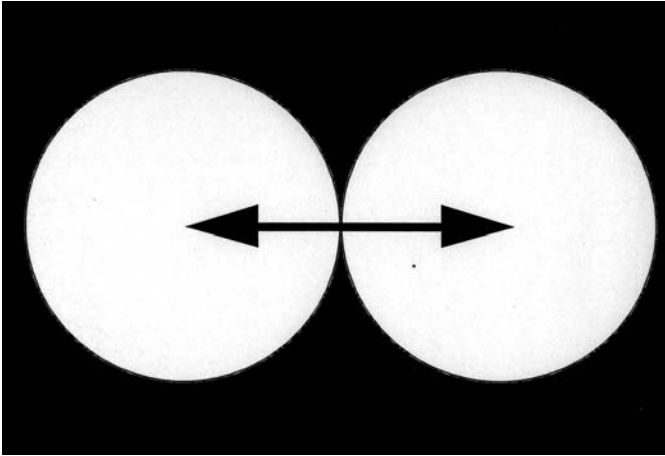
It has been shown as an auditory sensation that the preferred repetition period for noise bursts was approximately 0.55 s. In vision, subjective preference for a flickering light showed that the preferred period was approximately double this (Soeta et al., 2002a; Soeta et al., 2002c). As is described in Section 3.3, results of the scale value of subjective preference from the different test series, using different music programs, yield the common formula of Equation (3.9) as well as Equation (14.1).

Ten subjects (21–26 years of age) participated. All subjects had normal or correct-to-normal binocular vision. The stimuli were displayed on a CRT monitor presenting 30 frames per second. Figure 14.6 shows the stimulus, a single, white, circular target moving sinusoidal, used in the experiment. The diameter of the target was subtended  $1^\circ$  of the visual angle (1.22 cm). The movement of the stimulus is expressed as

$$h(t) = A \cos(2\pi t/T) \quad (14.2)$$

where  $A$  is the amplitude and  $T$  is the period of the stimulus. In all experiments, the amplitude  $A$  was fixed at 0.61 cm on the monitor screen, corresponding to  $0.5^\circ$  of visual angle. The white target and black background corresponded with gray levels 40 and  $0.5 \text{ cd/m}^2$ , respectively. The monitor presenting the stimuli was placed in a dark room 0.7 m away from the subject's eye position to maintain natural binocular vision.

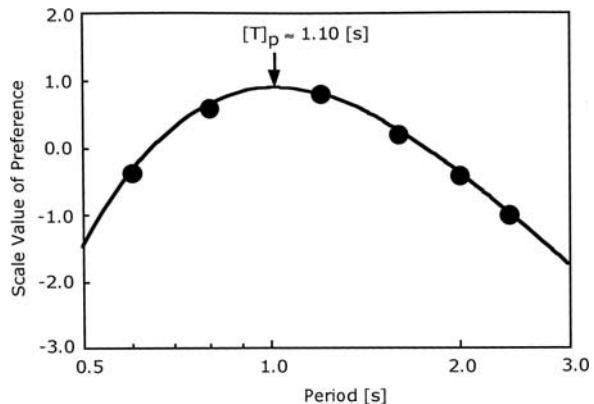
Subjective preference for the period of movements in the horizontal and vertical directions was examined separately. The period of stimulus movement  $T$  in Equation (14.2) was varied at six levels:  $T = 0.6, 0.8, 1.2, 1.6, 2.0,$  and  $2.4$  s. Thirty



**Fig. 14.6** Stimulus target used in the experiment showing an example of oscillatory horizontal movement

pairs combining six different periods constituted each series, and 10 series were conducted for all 10 subjects in the experiments by the PCT.

The most preferred period  $[T]_p$  for each subject was estimated by fitting a suitable polynomial curve to a graph on which scale values were plotted. Figure 14.7 shows an example of the method used for estimating  $[T]_p$ . The peak of this curve denotes the subject's most preferred value. Table 14.4 shows results of the most preferred periods for each subject for both vertical and horizontal directional stimuli. The global value of the most preferred period was about 0.97 s for vertical movement and about 1.26 s for horizontal movement. Results from all subjects indicated that preferred periods in the vertical direction were shorter than that of those in the horizontal direction ( $p < 0.01$ ).



**Fig. 14.7** Viewer preferences for rate of visual oscillatory motion. The arrow indicates the most preferred period  $[T]_p$  ( $\approx 1.10$  [s]) in the vertical direction (subject J)



**Table 14.4** The most preferred periods  $[T]_p$  of vertical and horizontal movements of the target for each subject and the averaged values

Subject	Vertical[s]	Horizontal[s]
A	1.15	1.28
B	1.05	1.82
C	0.78	1.31
D	1.16	1.79
E	0.85	0.91
F	0.83	1.05
G	1.08	1.31
H	0.81	1.04
I	0.93	0.98
J	1.10	1.13
Averaged	0.97	1.26

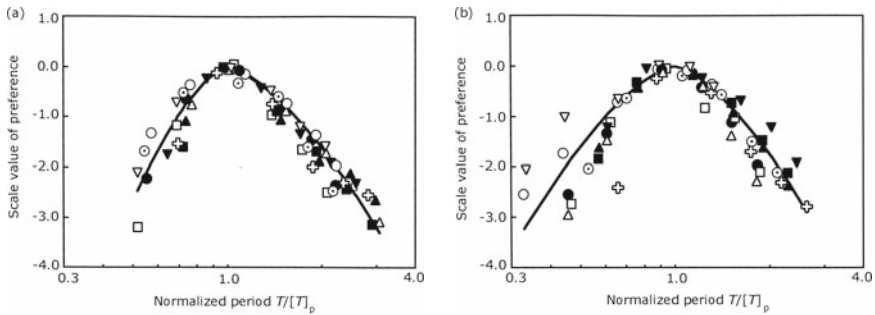
We also attempted to determine the characteristics of the preference evaluation curve in more detail. As shown in Fig. 14.7, the preference evaluation curve here can also be expressed in the form of Equation (14.1), where  $x$  is replaced by  $\log_{10}T - \log_{10}[T]_p$ . After obtaining the most preferred period for each subject, we identified values of  $\alpha$  and  $\beta$  for the period in the fast-moving range in reference to  $[T]_p$ , and also in the slow-moving range (Table 14.5). The values of  $\alpha$  and  $\beta$  in the fast-moving range in reference to  $[T]_p$  for the vertical direction could not be calculated, because there are only two available scale values. The average value of  $\alpha$ , estimated by a quasi-Newton numerical method, was approximately 1.43. Thus, the value of  $\alpha$  can be fixed at  $3/2$  similar to that above. The weighting coefficient  $\alpha$  describes the sharpness of the preference curve with respect to the normalized period. Values of  $\alpha$  of each individual for the period of  $T < [T]_p$  and  $T > [T]_p$  may be obtained as indicated in Table 14.6.

**Table 14.5** The values of  $\alpha$  and  $\beta$  for each subject as calculated by Equation (14.1) and averaged values of  $\beta$ . Because averaged values of  $\beta$  are close to  $3/2$ , it may be fixed at  $3/2$  obtaining a single constant  $\alpha$  representing the individual difference as listed in Table 14.6

Subject	Vertical direction				Horizontal direction			
	$T \leq [T]_p$		$T \geq [T]_p$		$T \leq [T]_p$		$T \geq [T]_p$	
	$\alpha$	$\beta$	$\alpha$	$\beta$	$\alpha$	$\beta$	$\alpha$	$\beta$
A	–	–	11.12	1.33	21.86	1.89	9.08	1.19
B	–	–	8.84	1.49	7.00	1.39	–	–
C	–	–	8.26	1.41	16.03	1.57	17.62	1.54
D	–	–	8.17	1.41	6.20	1.56	–	–
E	–	–	5.53	0.93	–	–	7.68	1.16
F	–	–	8.33	1.28	–	–	10.74	1.60
G	–	–	12.75	1.62	13.03	1.52	14.81	1.53
H	–	–	4.91	0.86	–	–	11.24	1.56
I	–	–	6.57	1.15	–	–	9.41	1.72
J	–	–	15.55	1.72	–	–	11.84	1.54
Averaged	–	–	–	1.32	–	1.59	–	1.48

**Table 14.6** The values of  $\alpha$  obtained in the ranges of fast and slow periods with respect to the most preferred period  $[T]_p$ :  $\alpha_f(\leq [T]_p)$  and  $\alpha_s(\geq [T]_p)$  for each subject. When the value of  $\beta$  is fixed at  $3/2$  in Equation (14.1), then the individual differences may be represented by the constant  $\alpha$

Subject	Vertical direction		Horizontal direction	
	$\alpha_f(\leq [T]_p)$	$\alpha_s(\geq [T]_p)$	$\alpha_f(\leq [T]_p)$	$\alpha_s(\geq [T]_p)$
A	20.73	13.74	13.55	13.81
B	10.45	8.93	7.68	7.71
C	19.44	8.92	14.69	16.58
D	15.05	9.23	5.89	7.88
E	25.19	9.51	29.80	10.87
F	29.21	10.20	14.30	9.60
G	16.46	11.04	12.74	14.10
H	22.04	8.80	13.07	10.49
I	20.02	9.42	11.37	7.44
J	11.84	11.95	13.30	11.09
Averaged	19.04	10.17	13.64	10.96



**Fig. 14.8** (a) The normalized scale values of preference for individual subjects for movement oscillation period in the vertical direction. (b) Those in the horizontal direction. Different symbols indicate scale values obtained with different subjects

Figure 14.8 shows scale values for all subjects and the preference evaluation curve calculated by Equation (14.1). The results indicate that a sinusoidal period of about 1.26 s is preferred for horizontally moving stimuli. This period is approximately twice the period of the most preferred tempo for noise bursts and the same as the most preferred period of a flickering light. For vertically moving stimuli, the most preferred period was about 0.97 s, which is a significantly faster period than that of the horizontally moving stimuli. Moreover, the values of  $\alpha$  in the range of  $T < [T]_p$  were significantly larger than those of  $T > [T]_p$  in regard to vertical movement ( $p < 0.01$ ).

Several investigators found that motion sensitivity to vertical and horizontal movement is isotropic (Ball and Sekuler, 1979; Levinson and Sekuler, 1980; van de Grind et al., 1993; Raymond, 1994; Gros et al., 1998). Kinchla and Allan (1970)

indicated that there is no difference in sensitivity to vertical and horizontal movement, but that judgmental-standards asymmetries exist. Soeta et al. (2003) showed that subjective preference for vertical movement was more sensitive than that for horizontal movement; that is, asymmetry in the sensitivity. Thus, more attention to sensitivity is needed for the vertical direction because of the sharp decline of subjective preference in the fast-moving range.

## 14.3 Subjective Preferences for Texture

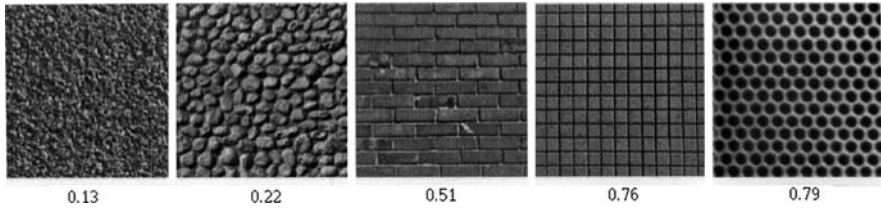
Evaluation of subjective preference of texture was conducted by the PCT. By averaging the scale values of all subjects, it was found that the most preferred value of  $\phi_1$  for texture regularity extracted from the spatial ACF is approximately given by  $[\phi_1]_p \approx 0.41$ . As an application, a suggestion is offered here that artistic expressions of color modulation and sequential form be realized in a drawing.

### 14.3.1 Preferred Regularity of Texture

In a manner similar to that described in Section 14.2, the four factors were extracted from the two-dimensional ACF (Fujii and Ando, unpublished data). To compare the degree of periodicity, the amplitude of the first peak in the ACF,  $\phi_1$ , was considered under the condition of a roughly constant  $\delta_1$ . As a first approximation in the previous section, rating of the  $\phi_1$  value could represent perceived regularity of texture (Equation (13.5)). If the size and shape of the materials and the spacing between the objects in the pattern are completely equal, the calculated ACF does not decay. This means that such a texture is, theoretically, perfectly regular. However, if the materials contain a kind of fluctuation of the object size and spacing and nonuniformity of the light reflection, then the ACF gradually decays, and the value of  $\phi_1$ , which is a measure of perceived regularity, is considered as the measure for the degree of fluctuation in texture.

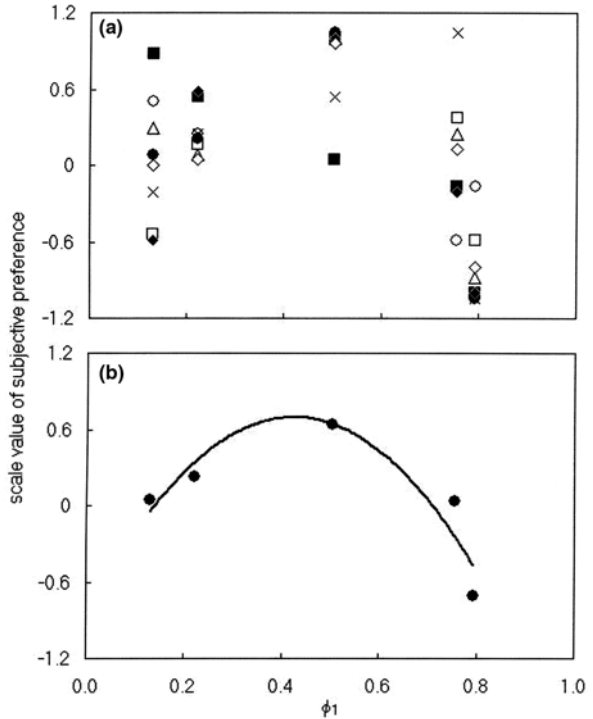
Ten 22- to 24-year-old subjects participated in the experiment. All had normal or corrected-to-normal visual acuity. Stimuli were presented on a display under a dark surrounding. The display was set at a distance of 1.5 m from the subjects. Subjects were presented pairs of two stimuli and asked to judge which they preferred (PCT). All possible pairs from the five selected stimuli as shown in Fig. 14.9 were presented in a random order in one session. All subjects conducted 10 series of sessions, giving a total of 100 judgments.

Results for all subjects are shown in Fig. 14.10. The scale value of the subjective preference has a single peak value for each subject, even allowing some individual differences. The most preferred range was found in the value of  $\phi_1$  for each subject. Subjects did not prefer textures that had a too high or too low a value of  $\phi_1$ . By averaging the scale values of all subjects, it was found that  $[\phi_1]_p \approx 0.41$  was the most preferred value for texture regularity. The coefficients in Equation (14.1) with the number of subjects was  $\alpha \approx 3.9$  and  $\beta = 3/2$ .



**Fig. 14.9** Two-dimensional spatial textures used with the value of  $\phi_1$  for the subjective preference judgment

**Fig. 14.10** (a) Scale values of preference for individual subjects. (b) Averaged values. For averaged values, a fitting curve with the 3/2 power of  $\phi_1$  in Equation (14.1) is shown



It is worth noting that the most preferred subjective preference of the flickering light with fluctuation as described in Section 14.1 is  $[\phi_1]_p \approx 0.46$  as well. It is considered, therefore, that a certain degree of fluctuation in both temporal and spatial factors is a visual property affecting subjective preference.

**14.3.2 Application: Spatial “Vibrato” in a Drawing**

Artistic expressions may be found in Photo 14.1, which was personally provided by Werner Lauterborn in 2003 and photographed by Gisa Kirschmann Schroeder at

**Photo 14.1** An inspiration from the cave paintings in the cave "Cosquer" near Marseille, France (Lauterborn, 2003)



the Third Physics Institute, University of Goettingen. This painting was composed and interpreted by Lauterborn based on an inspiration from cave paintings dated approximately 18,500 years ago that were discovered in the cave "Cosquer" near Marseille, France, which is accessible only via an underwater entrance by diving (Clottes and Courtin, 1995).

It is interesting to point out the fact that color modulation, expressed by Lauterborn as a form of "vibrato," is present in the background being a kind of cave wall. Also, the horsehair sequence of a "vibrato" in the spatial frequency might be analyzed by the ACF of images with the gray level. There is a certain degree of similarity as mentioned above, which is discussed both in the time domain analysis of the flickering light and in the spatial domain relating to subjective preference, respectively. The horsehair sequence is periodic, not perfect or just random, and there is a certain degree of fluctuation like vibrato. In fact, the factor  $\phi_1$  extracted from the ACF is around 0.3, which is a range of the most preferred condition depending on the individual (Fig. 14.10). Regarding such color modulation including the effect of gray level, further attempts could be made to investigate the physical factors with multiple-dimensional correlation analysis. Lauterborn said that Photo 14.1 was painted in the years between 2001 and 2003. As a matter of fact, ideas of such a color modulation and a vibrato in the periodic structure came out without his conscious process. He could not identify how he painted it, because such kinds of ideas of beauty come from "out of time" or "beyond any specified time." Therefore, this process may relate to the "third stage of time" (Ando, 2004).

# Chapter 15

## EEG and MEG Correlates of Visual Subjective Preferences

### 15.1 EEG Correlates of Preferences for Flickering Lights

We have sought to find visual analogies of our comprehensive auditory signal processing model. The last chapter dealt with temporal and spatial sensations in vision that may be mediated by temporal and spatial autocorrelation representations. In the chapter we take up the neural response correlates of visual subjective preferences. For this purpose we analyzed EEG responses to visual stimuli using techniques and analyses similar to those used for auditory stimuli (Chapter 4). Because subjective preference is perhaps the most primitive response of an organism, as in the auditory case, we expected to find response correlates for preferred conditions in the persistence, temporal coherence, and extent of alpha rhythms in EEG and MEG signals. Table 15.1 summarizes our experiments and neural response correlates.

**Table 15.1** Summary of overall argument in this chapter

Acoustic factor	Subjective response	Locus	Neuronal correlate
Period of flickering light, T	Subjective preference	Left hemisphere	Alpha wave in EEG
Period of flickering light, T	Subjective preference	Left hemisphere	Alpha wave in MEG
Period of moving target, T	Subjective preference	Left hemisphere	Alpha wave in EEG

#### 15.1.1 Persistence of Alpha Rhythms

Human cortical responses corresponding to subjective preferences for flicker lights were investigated. We studied the effects of fluctuations in the period and mean luminance of flickering light sources. Paired comparison tests were used to measure subjective preferences. Then, in order to identify neural response correlates of visual preferences, electroencephalographic (EEG) recordings were taken from the same subjects during presentations of more and less favored flicker conditions. Analogous to our findings with auditory subjective preferences, we found that the

effective durations of alpha rhythms, measured at occipital electrodes O1 and O2, were longer for the more preferred visual stimuli. In the preferred conditions, alpha rhythms persist longer, with higher temporal coherence, and extend over wider spatial regions of the cerebral cortex.

As reported by Lindsay (1952), who considered the relation of brain activity and behavioral states, the presence of alpha rhythms in the EEG of a human subject corresponds well to mental states associated with relaxation and free creative thought. The term *alpha* refers to the frequency band between about 8 and 13 Hz. The differentiation of basic emotions (i.e., intention, anxiety, aggression, sadness, and joy) by means of EEG-coherences has been discussed extensively (Hinrichs and Machleidt, 1992). Intention, aggression, and joy are mainly characterized by an increase in alpha-coherence, whereas a decrease is seen for anxiety and sorrow. In Chapter 4, we have discussed the method for using the ACF to analyze brain waves to examine the relationship between brain activities and the scale value of subjective preference as an overall impression of the sound field. We analyzed the effective duration of the normalized ACF envelope ( $\tau_e$ ) of the alpha waves when temporal factors such as the initial time delay gap between the direct sound and the first reflection ( $\Delta t_1$ ) and the subsequent reverberation time ( $T_{\text{sub}}$ ) were varied. Results showed that the  $\tau_e$  of the alpha waves is longer only in the left cerebral hemisphere for the preferred conditions of these temporal factors. The relationship between subjective preference and the ACF of the alpha waves in response to the tempo of a noise burst, for example, has been investigated (Chen et al., 1997). Results showed that the  $\tau_e$  of the alpha waves is longer only in the left cerebral hemisphere for the preferred tempo of a noise burst. Petsche (1996) analyzed EEG changes caused by mental processes of a higher order by using coherence analysis. Acts of creative thinking, whether verbal, visual, or musical, were characterized by a more increased coherence between occipital and frontopolar electrode sites than were other mental tasks. Results were interpreted as showing a stronger involvement of the long cortico-cortical fiber systems in creative tasks.

In this section we examine whether the scale value of subjective preference of visual stimuli reflects the temporal information in EEG in the left or right cerebral hemisphere (Soeta et al., 2002a). First, the PCT was performed for flickering light sources of varied period and mean luminance. From results of scaling the value of subjective preference, the most preferred and relatively less preferred light sources were selected as paired stimuli for brain wave recordings. Then, relationships between the scale value of subjective preference and the factors extracted from the ACF in the alpha waves were examined.

The light source was a 7-mm-diameter green light-emitting diode (LED), and was viewed by the subject at the distance of 0.6m from it in dark surroundings. The LED stimulus field was spatially uniform, and its size corresponded with  $0.67^\circ$  of the visual angle. The luminance of the stimulus is given by

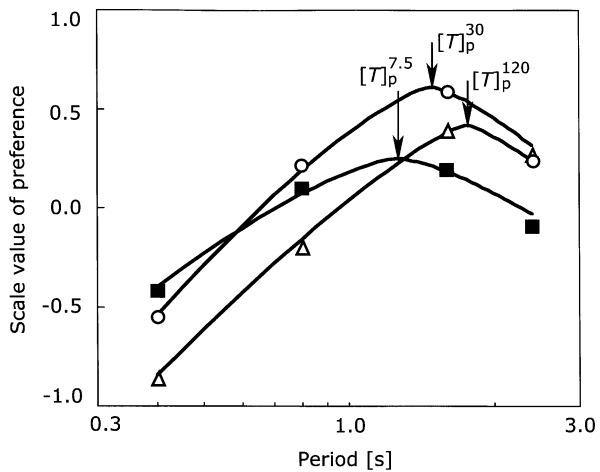
$$I(t) = L_0[1 + m\cos(2\pi ft)] \quad (15.1)$$

where  $L_0$  is the mean luminance,  $m$  is modulation (relative) amplitude fixed at 1.0, and  $f$  is the temporal frequency of the stimulus. The period  $T = 1/f$  was set at 0.4, 0.8,

1.6, or 2.4 s, and mean luminance was set at 7.5, 30, and 120 cd/m<sup>2</sup>. The duration of the stimuli was fixed at 5 s. Ten 23- to 25-year-old subjects participated in the experiment. All had normal or corrected-to-normal vision. They adapted to the dark and looked at the LED stimulus seated in a dark room with a comfortable thermal environment. The PCT was conducted for each subject by having each subject compare 66 pairs per session and having 10 sessions. The subject was asked which stimulus they preferred to watch. The scale value of subjective preference of each subject, which is regarded as the linear psychological distance between light sources, was obtained (Ando and Singh, 1996; Ando, 1998).

The average scale values of preference obtained from the 10 subjects are shown in Fig. 15.1. The most preferred period, [T]<sub>p</sub>, for all subjects were estimated by fitting a suitable polynomial curve to a graph on which the scale values were plotted. The value of [T]<sub>p</sub> for all of the subjects was 1.27 s at a mean luminance of 7.5 cd/m<sup>2</sup>, 1.49 s at a mean luminance of 30 cd/m<sup>2</sup>, and 1.76 s at a mean luminance of 120 cd/m<sup>2</sup>. The most preferred period, i.e. peak of the scale value, shifted gradually toward longer periods as the mean luminance was increased (Fig. 15.2). This may imply that the most preferred condition corresponds to a constant total amount of excitation from the physical environment. Too much or too little excitation may be less preferred.

**Fig. 15.1** Averaged scale values of subjective preference as a function of the flicker period. Different symbols indicate different mean luminance. ■, 7.5 cd/m<sup>2</sup>; ○, 30 cd/m<sup>2</sup>; △, 120 cd/m<sup>2</sup>. Solid curve is expressed by Equation (15.2)



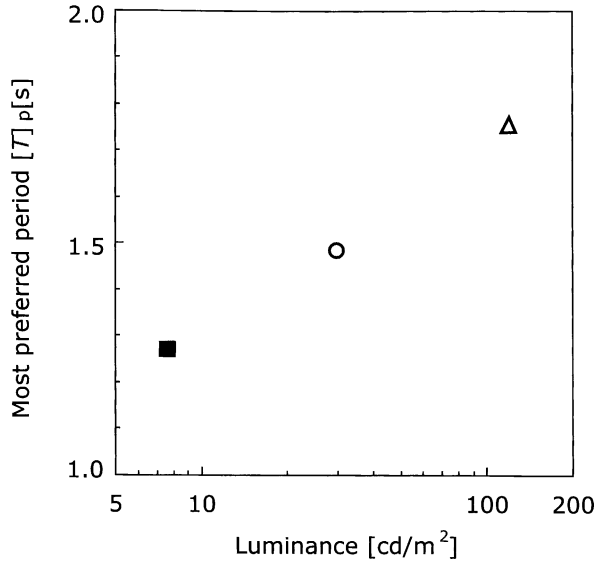
Similar to the above, the preference evaluation curve may commonly be expressed by

$$S = S_L \approx -\alpha |x|^\beta \tag{15.2}$$

where  $\alpha$  and  $\beta$  are the weighting coefficient and a constant, respectively, and  $x = \log_{10}T - \log_{10}[T]_p$ . Values of  $\alpha$  and  $\beta$  were obtained by using the quasi-Newton numerical method, respectively, and were approximately 4.90 and 1.56, respectively. It is interesting that the value of  $\beta \approx 3/2$  was consistent with other preference judgments performed including those for the sound and visual fields. When the hori-

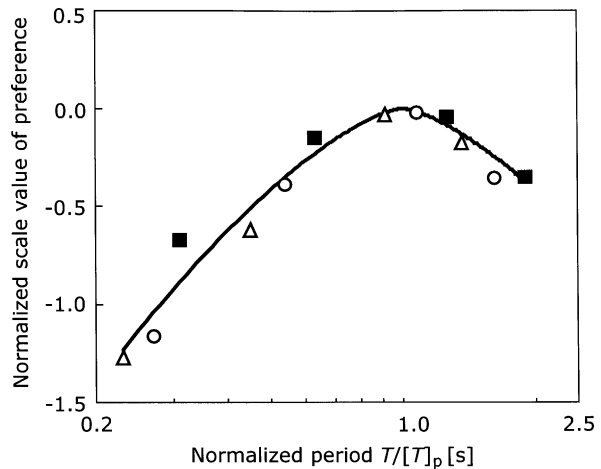


**Fig. 15.2** The most preferred flicker periods  $[T]_p$  obtained by 10 subjects as a function of the mean luminance



zontal axis of Fig. 15.1 is normalized by the most preferred period  $[T]_p$ , then all results may be reduced by a single curve as shown in Fig. 15.2. Without loss of any generality, the scale value can be adjusted to zero at the preferred condition, so that scale values for different  $[T]_p$  values of mean luminance had similar tendencies as shown in Fig. 15.3. Therefore, the preference evaluation curve can be calculated by Equation (15.2) with  $\beta = 3/2$ .

Effects of the period and mean luminance on the scale value of preference were examined for all 10 subjects using the two-way analysis of variance (ANOVA) method. The results clearly indicated that effects of the period were significant



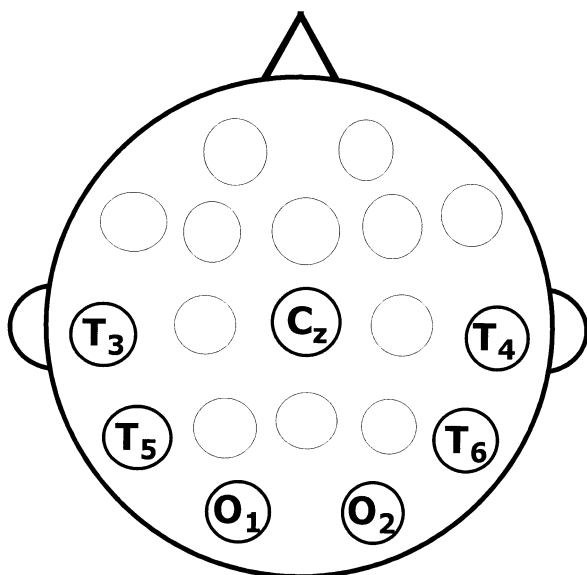
**Fig. 15.3** Scale values of preference as a function of the flicker period normalized by the most preferred periods  $[T]_p$ . Different symbols indicate different mean luminance. ■, 7.5 cd/m<sup>2</sup>; ○, 30 cd/m<sup>2</sup>; △, 120 cd/m<sup>2</sup>. Solid curve is expressed by Equation (15.2)

( $p < 0.01$ ). The period and mean luminance were independent influences on the subjective preference judgment.

Next, the same subjects that were used in the preference tests participated in EEG recordings. The EEG was recorded under three conditions: (1) period varied and mean luminance fixed; (2) period fixed and mean luminance varied; (3) both period and mean luminance varied. To find a significant effect of subjective preference on an EEG, the most preferred flickering light and the relatively less preferred flickering light were selected as paired stimuli. The paired stimuli were set for each subject according to their individual preferences. The subject watched the most and the least preferred flickering lights alternatively. A series of EEG was recorded 3 times for each subject, and each series consisted of 10 stimuli pairs.

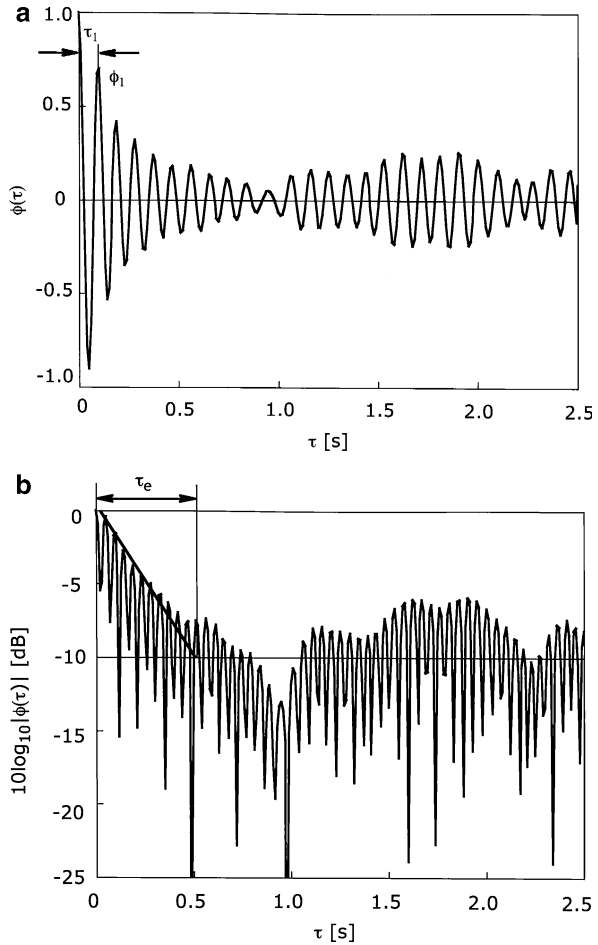
The EEG was recorded from the left and right cerebral hemispheres of subjects using silver electrodes (7 mm diameter) at scalp locations T3, T4, T5, T6, O1, and O2, as shown in Fig. 15.4 (10–20 International Electrode Placement System). A reference electrode was attached to the earlobe of a subject. A ground electrode was placed on the forehead of a subject. The recorded data were filtered with a digital band-pass filter with cutoff frequencies of 8 and 13 Hz (alpha-wave range).

**Fig. 15.4** Top view of a subject's head and electrode positions on the scalp for EEG recordings (10–20 International Electrode Placement System)



An example of a measured ACF is shown in Fig. 15.5a. The ACF may be characterized by four variables (see Sections 2.2 and 5.2). Figure 15.5b shows the absolute value in the logarithmic form as a function of the delay time. To find the degree of ACF envelope decay, the effective duration,  $\tau_e$ , is determined. As shown in Fig. 15.5b, the straight-line regression of the ACF can be drawn by using only the initial declining portion,  $0 \text{ dB} > 10 \log |\phi(\tau)| > -5 \text{ dB}$ . In most cases, the envelope decay of the initial part of the ACF may fit a straight line. The value of  $\tau_1$

**Fig. 15.5** (a) An example of a normalized autocorrelation function ACF of an EEG alpha-band signal (8–13 Hz) showing definitions of the delay time of the first peak  $\tau_1$  and its amplitude  $\phi_1$ . (b) Determination of the effective duration ( $\tau_e$ ) of the alpha rhythm by estimating the slope of the envelope of the autocorrelation function and determining the delay at which it reaches 10% of its maximal, zero-lag value. Effective duration measures duration of temporal coherence, i.e., the duration for which repetitive structure persists in a signal



corresponded mostly with the center frequency in the range 8–13 Hz and thus was not analyzed.

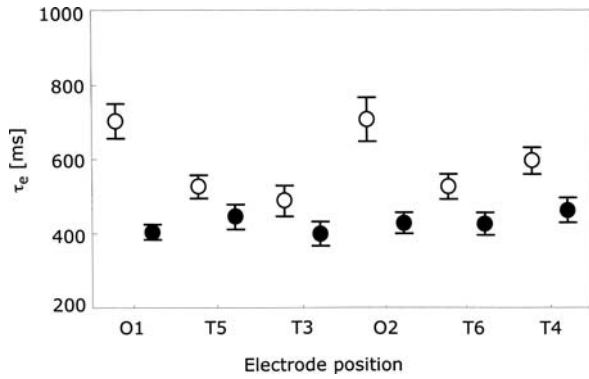
Referring to the results in Section 5.3, the integration interval  $2T$  was selected as 2.5 s in the running ACF analysis to obtain the values of  $\tau_e$ ,  $\Phi(0)$ , and  $\phi_1$ . Table 15.2 shows results of the one-way ANOVA for  $\tau_e$ ,  $\Phi(0)$ , and  $\phi_1$  values of the alpha wave for the 10 subjects. Significant effects were found when the period was varied and the mean luminance was fixed and when both period and luminance were varied. However, significant effects were not found when the period was fixed and the luminance was varied except for  $\Phi(0)$  at O1.

Only when the period was varied were the values of  $\tau_e$ ,  $\Phi(0)$ , and  $\phi_1$  for the most preferred stimuli larger than those for the less preferred stimuli for all subjects, as shown in Figs. 15.6, 15.7, and 15.8. The tendency was especially clear in the

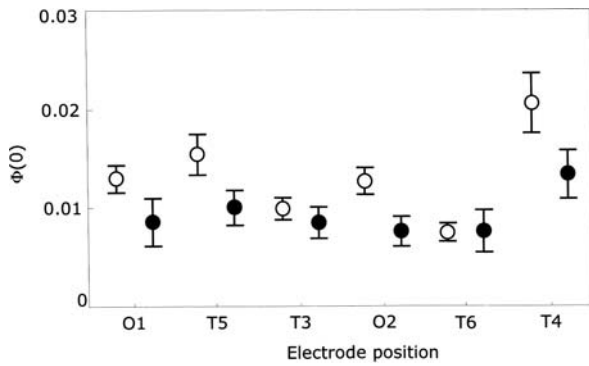
**Table 15.2** Results of one-way ANOVA at each electrode position under three conditions

		Factor					
		$\tau_c$			$\phi_1$		
Condition	Electrode position	F value	Significance level	F value	Significance level	F value	Significance level
(1)	O1	143.2	<0.001	52.4	<0.001	132.8	<0.001
	T5	18.8	<0.001	19.7	<0.001	38.2	<0.001
	T3	15.6	<0.001	12.1	<0.01	25.5	<0.001
	O2	81.2	<0.001	51.3	<0.001	121.6	<0.001
(2)	T6	36.9	<0.001	28.5	<0.001	52.1	<0.001
	T4	26.4	<0.001	6.0	<0.05	32.4	<0.001
	O1	2.0		4.4	<0.05	0.3	
	T5	0.1		2.4		0.1	
(3)	T3	0.1		0.1		0.3	
	O2	0.9		3.6		0.9	
	T6	3.7		1.2		1.5	
	T4	0.1		0.1		0.2	
	O1	143.1	<0.001	39.3	<0.001	132.8	<0.001
	T5	27.6	<0.001	47.9	<0.001	64.0	<0.005
	T3	4.6	<0.05	10.6	<0.005	24.6	<0.005
	O2	25.3	<0.001	10.2	<0.005	34.4	<0.005
T6	10.5	<0.01	27.0	<0.001	32.1	<0.001	
T4	18.4	<0.005	14.2	<0.005	27.4	<0.001	

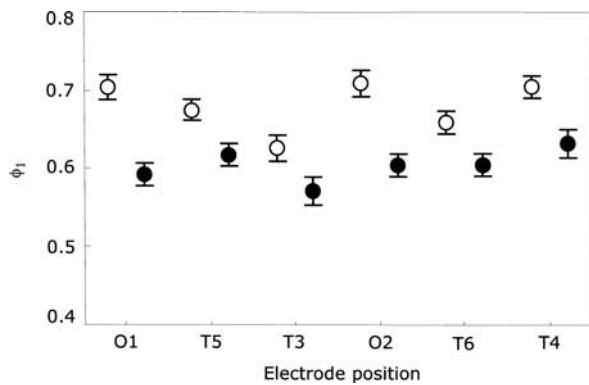
**Fig. 15.6** Effective durations  $\tau_e$  of EEG alpha rhythms at different electrode positions in response to a change in flicker period using preferred and less preferred rates. Error bars represent 95% confidence interval.  $\circ$ , higher preference;  $\bullet$ , lower preference



**Fig. 15.7** Magnitudes  $\Phi(0)$  of EEG alpha rhythms at different electrode positions in response to a change in flicker period using preferred and less preferred rates. Error bars represent 95% confidence interval.  $\circ$ , higher preference;  $\bullet$ , lower preference



**Fig. 15.8** Amplitudes  $\phi_1$  of EEG alpha rhythms at different electrode positions in response to a change in flicker period using preferred and less preferred rates. Error bars represent 95% confidence interval.  $\circ$ , higher preference;  $\bullet$ , lower preference



EEG from occipital sites (O1, O2) ( $p < 0.01$ ). It is significant that the value of  $\tau_e$  correlated with the value of  $\phi_1$  ( $r = 0.75$ ,  $p < 0.01$ ) but not with the value of  $\Phi(0)$  ( $r = 0.19$ ,  $p < 0.01$ ). When both period and mean luminance were varied simultaneously, the values of  $\tau_e$ ,  $\Phi(0)$ , and  $\phi_1$  for the most preferred stimuli were larger than those for the less preferred stimuli. It is remarkable that the ratio of high to low preference in terms of the  $\tau_e$  was larger than that of  $\phi_1$  at the occipital sites (O1 and O2). This indicates that the  $\tau_e$  of alpha wave, i.e. the persistence and temporal coherence of the alpha rhythm, reflects the subjective preference better than does the  $\phi_1$ , i.e. its amplitude. There were no clear differences in the factors extracted from the ACF except when the mean luminance was varied.

As shown in Figs. 15.1 and 15.2, the longer period was preferred when the amplitude sensitivity was higher in the mean-luminance range 7.5–120 cd/m<sup>2</sup>. When the period was varied, the preferred stimulus had a significantly larger  $\tau_e$  than that of the least preferred stimulus, especially at the occipital sites. This kind of tendency for larger  $\tau_e$  under the preferred conditions was commonly discovered in the auditory-brain system, when  $\Delta t_1$ ,  $T_{\text{sub}}$  of a music sound field and the tempo of a noise burst were varied. The  $\tau_e$  signifies the degree of similar repetitive features included in the alpha wave. The fact that alpha wave has a significantly larger  $\tau_e$  indicates that the brain repeats the rhythm for a longer time, on average, under the preferred conditions.

Because O1 (left hemisphere) and O2 (right hemisphere) are closely located, as shown in Fig. 15.4, the difference between their EEG signals was difficult to identify. But, the larger  $\tau_e$  in the left hemisphere may reflect the specialization of the human brain, specifically, the left hemisphere dominance for temporal factors. To further test this hypothesis, that the left hemisphere activity is dominant in the processing of temporal factors in vision as well as in audition, we also measured MEG signals in response to flickering lights.

### 15.1.2 Spatial Extent of Alpha Rhythms

To determine the flow of alpha waves across the cerebral cortex and their possible relation to subjective preference, we compared pairs of EEG signals from several different scalp locations by analyzing their crosscorrelation functions (CCFs). First, PCT was performed to examine subjective preference for the flickering light. The EEG was then recorded from seven electrodes (10–20 International Electrode Placement System) during presentations of the most and least preferred flickering-light conditions. The maximum value of the CCF,  $|\phi(\tau)|_{\text{max}}$ , between the alpha wave measured at different electrodes and its delay time,  $\tau_m$ , was analyzed. Results show that the most preferred flickering light has a significantly larger  $|\phi(\tau)|_{\text{max}}$  than that of the less preferred flickering light and that the value of  $|\phi(\tau)|_{\text{max}}$  decreases with increasing distance between the reference (O1 or O2) and the test electrode. The delay time of the maximum value of the CCF,  $\tau_m$ , increases in a stepwise manner with the distance between reference and test electrodes. This suggests that there are discrete nuclei in the central system.

Analysis techniques based on autocorrelation (ACF) and crosscorrelation (CCF) have been developed to describe the nature of the EEG (Braizer and Casby, 1952; Barlow, 1961; McLachlan and Shaw, 1965; Liske et al., 1967; Hoovey et al., 1972). The ACF is used to determine characteristics over time, that is, the degree of persistence of a signal. Crosscorrelation functions (CCFs) are used to measure mutual relationships between the signals that are detected at two electrode sites, including common frequency components and transmission delays between the two sites.

The alpha rhythm, which has the longest period of any EEG brain rhythm normally seen in the waking state, is thought to be associated with pleasant and comfortable feelings. The relationship between the subjective preference and the alpha wave on the scalp has been studied by using the ACF factor  $\tau_e$  throughout this volume. In these studies, the effective duration of the envelope of the normalized ACF,  $\tau_e$ , of EEG and MEG alpha waves was analyzed. Results showed that the  $\tau_e$  value of the alpha waves is longer when the subject is presented with the preferred condition. The spread of alpha wave over the scalp has also been studied by using the CCF (Inoye et al., 1983; Sato et al., 2003). The propagation of the alpha wave from the right hemisphere to the left that corresponds to the change in the magnitude of IACC has been observed (see Section 4.3.3).

It is assumed that the subjective preference for visual stimuli is reflected in both the intrachannel and interchannel relations between the EEG alpha waves in the time domain. The relationship between the subjective preference and the alpha wave over the scalp was investigated by using CCF analysis. Experiments were conducted under three conditions: (1) variation of period with constant mean luminance fixed; (2) variation of mean luminance with constant period; (3) variation of both period and mean luminance (Soeta et al., 2002b). The paired stimuli were set for each subject according to the scale value of the individual preference as indicated in Table 15.3.

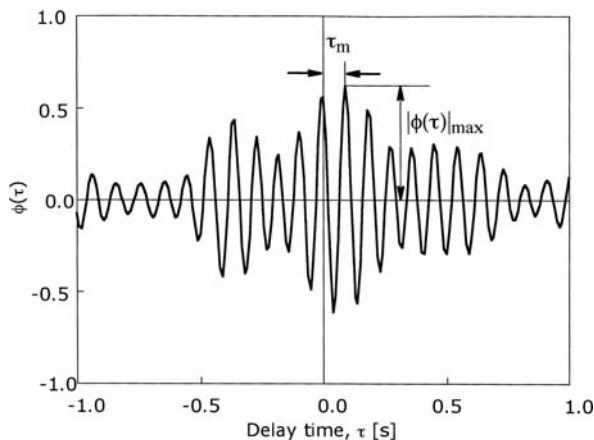
**Table 15.3** Paired stimuli and the differences between scale values of preference for each subject. These are presented under three conditions: (1) variation of period with constant mean luminance (ML); (2) variation of ML with constant period; (3) variation of both period and ML

Subject	Factor varied		
	(1) Period	(2) ML	(3) Period and ML
A	(2.4,7.5)-(0.4,7.5), 1.7	(2.4,7.5)-(2.4,120), 0.9	(2.4,7.5)-(0.4,120), 2.0
B	(0.4,7.5)-(2.4,7.5), 1.6	(0.4,7.5)-(0.4,120), 0.7	(0.4,7.5)-(2.4,120), 1.8
C	(2.4,120)-(0.8,120), 1.7	(2.4,120)-(2.4,7.5), 0.6	(2.4,120)-(0.8,30), 1.4
D	(0.8,30)-(0.4,30), 1.7	(0.4,7.5)-(0.4,120), 0.7	(0.8,7.5)-(0.4,120), 2.1
E	(0.8,30)-(0.4,30), 1.3	(0.8,30)-(0.8,7.5), 0.6	(0.8,30)-(0.4,120), 1.4
F	(2.4,120)-(0.4,120), 1.7	(2.4,120)-(2.4,7.5), 1.1	(2.4,120)-(0.4,7.5), 1.5
G	(1.6,30)-(0.4,30), 1.6	(1.6,30)-(1.6,7.5), 0.4	(1.6,30)-(0.4,120), 1.7
H	(2.4,120)-(0.4,120), 1.4	(2.4,120)-(2.4,7.5), 1.2	(2.4,120)-(0.4,7.5), 1.6
I	(1.6,120)-(0.4,120), 1.3	(1.6,120)-(1.6,7.5), 0.4	(1.6,120)-(0.4,30), 1.4
J	(1.6,120)-(0.4,120), 1.0	(0.4,7.5)-(0.4,120), 1.1	(0.8,7.5)-(0.4,120), 1.3

Values in parentheses indicate (Period, ML), and the single value that follows signifies the difference between scale values of subjective preference of the pair (X)-(Y).

The EEG was recorded from both the left and right cerebral areas of the scalps of subjects by using silver electrodes (7 mm diameter) at points T3, T4, T5, T6, O1, O2, and in addition Cz as shown in Fig. 15.4. The normalized CCF between the alpha waves measured at electrode positions O1 or O2 (reference electrodes) and those at the other electrodes (test electrodes) was analyzed. Subjective preference corresponded well with the effective duration of the ACF,  $\tau_e$ , at both O1 and O2 as discussed above. The integration interval for the CCF was the same (2.5 s) as had been used in the ACF analysis. An example of the normalized CCF is shown in Fig. 15.9. A positive lag ( $\tau > 0$ ) means that the activity at the reference electrode was delayed relative to that at the test electrode.  $|\phi(\tau)|_{\max}$  was defined as the maximum value of the CCF in the range of  $\tau \geq 0$ , and  $\tau_m$  was defined as its delay time.

**Fig. 15.9** An example of the normalized crosscorrelation function CCF between two EEG alpha band signals recorded from different electrodes showing the definitions of the maximum correlation value  $|\phi(\tau)|_{\max}$  and its delay time  $\tau_m$



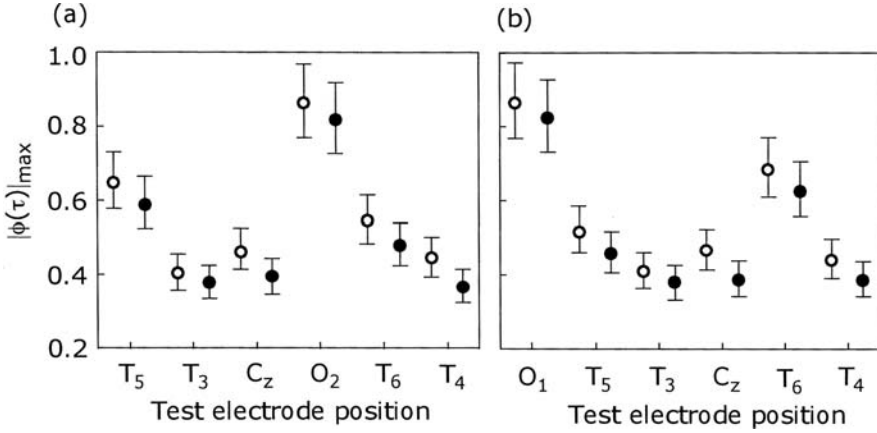
Effects of the subjective preference and the test electrode position on the  $|\phi(\tau)|_{\max}$  values were examined for all 10 subjects by using two-way ANOVA as indicated in Table 15.4. Results clearly indicated that  $|\phi(\tau)|_{\max}$  values were significantly related to the subjective preference when the period alone was varied and when both the period and the mean luminance were varied. When the period was varied, the value of  $|\phi(\tau)|_{\max}$  was significantly greater for the most preferred stimulus than for the less preferred stimulus. This result is true for both reference electrode positions (O1 and O2), as is shown in Fig. 15.10. However, as is shown in Fig. 15.11, there were no clear differences between values of  $|\phi(\tau)|_{\max}$  when the mean luminance was varied. When both period and mean luminance were varied, the value of  $|\phi(\tau)|_{\max}$  was significantly greater for the most preferred stimulus than for the least preferred stimulus. This is shown in Fig. 15.12.

Figure 15.13 shows the cumulative frequency curves of  $\log_{10}\tau_m$  for different test electrodes when the period was varied. The value of  $\tau_m$  increased with the distance between the reference and test electrode under all three conditions regardless of subjective preference. Remarkably, a stepwise phenomenon in the  $\tau_m$  values was

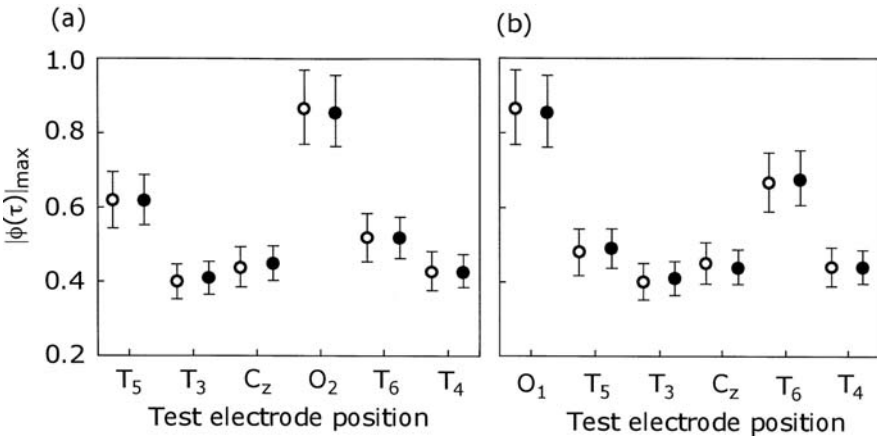


**Table 15.4** Results of two-way ANOVA on  $|\phi(\tau)|_{\max}$  under the three conditions indicated in Table 15.3

Factor varied	Factor	Reference electrode					
		O <sub>1</sub>			O <sub>2</sub>		
		F value	Significance level	F value	Significance level	F value	Significance level
(1) Period	Preference	182.6	<0.001	139.1	<0.001		
	Electrode position	1068.5	<0.001	1005.9	<0.001		
	Preference and electrode position	3.5	<0.005	2.5	<0.05		
(2) ML	Preference	0.3	0.60	0.9	0.33		
	Electrode position	1084.7	<0.001	1186.9	<0.001		
	Preference and electrode position	0.4	0.88	0.4	0.80		
(3) Period and ML	Preference	100.3	<0.001	94.3	<0.001		
	Electrode position	1077.7	<0.001	1032.9	<0.001		
	Preference and electrode position	2.5	<0.05	1.7	0.13		



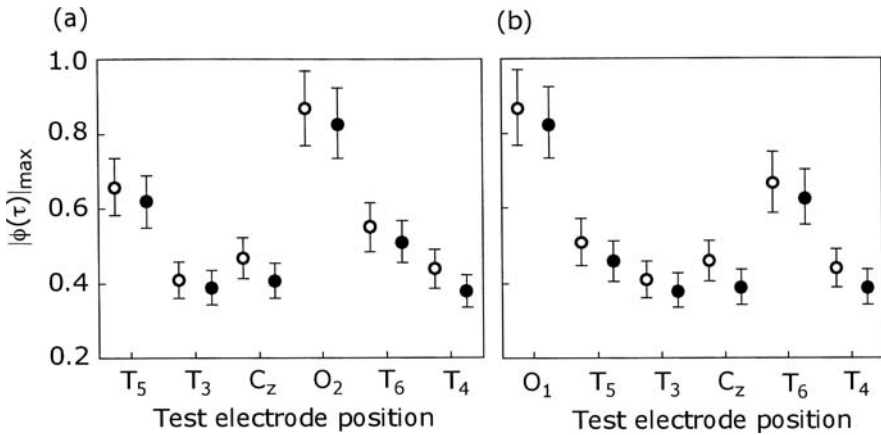
**Fig. 15.10** Maximal crosscorrelation values  $|\phi(\tau)|_{\max}$  for EEG alpha band signals as a function of test electrode position in response to variation of the flicker period between preferred and less preferred conditions. (a) Reference electrode selected: O<sub>1</sub>. (b) Reference electrode selected: O<sub>2</sub>. Error bars represent 95% confidence. ○: Higher preference; ●: lower preference



**Fig. 15.11** Maximal crosscorrelation values  $|\phi(\tau)|_{\max}$  for EEG alpha band signals as a function of test electrode position in response to variation of mean luminance between preferred and less preferred conditions. (a) Reference electrode selected at O<sub>1</sub>. (b) Reference electrode selected at O<sub>2</sub>. Error bars represent 95% confidence. ○: Higher preference; ●: lower preference

found under all three conditions. The values of  $\tau_m$  discovered here were centered on about 10 ms ( $\log \tau_m = -2.0$ ) and 50 ms ( $\log \tau_m = -1.3$ ). This means that the alpha wave propagates stepwise in the delay time.

When the period is varied, the preferred stimulus induces a significantly greater value of the absolute value of normalized ACF  $|\phi(\tau)|_{\max}$  of the alpha waves than that



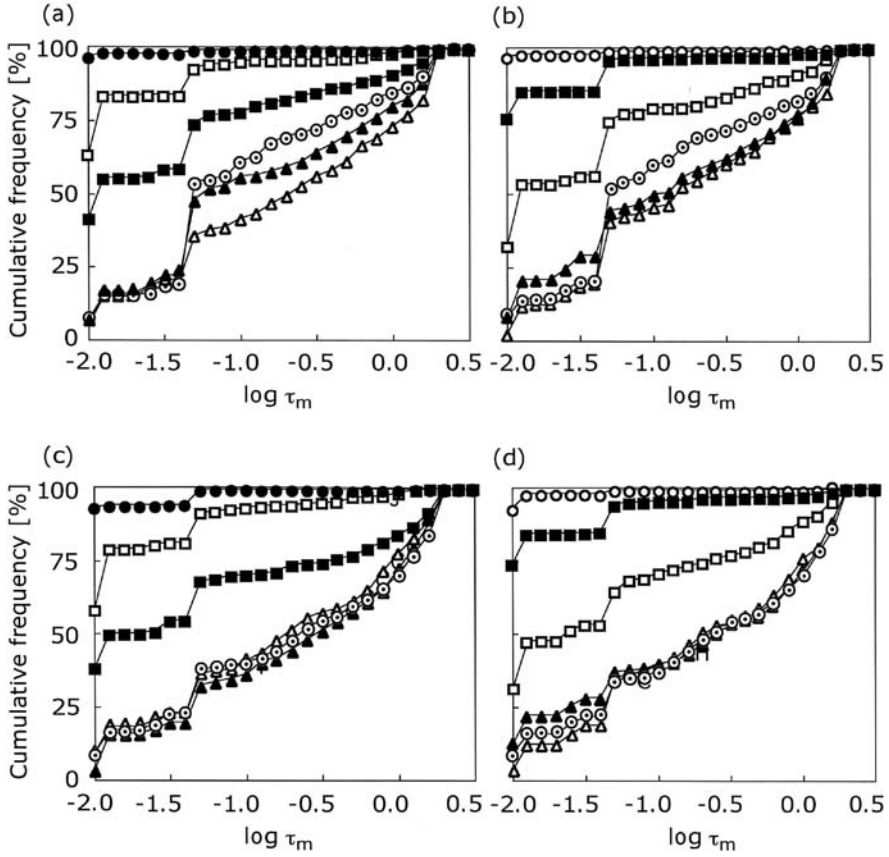
**Fig. 15.12** Maximal crosscorrelation values  $|\phi(\tau)|_{\max}$  for EEG alpha band signals as a function of test electrode position in response to variation of both period and mean luminance between preferred and less preferred conditions. (a) Reference electrode selected: O<sub>1</sub>. (b) Reference electrode selected: O<sub>2</sub>. Error bars represent 95% confidence.  $\circ$ : Higher preference;  $\bullet$ : lower preference

of the less preferred stimulus. The  $|\phi(\tau)|_{\max}$  signifies the degree of similar repetitive features that appear in the alpha waves recorded at two spatially separated electrodes. Significantly greater values of  $|\phi(\tau)|_{\max}$  for the alpha wave signifies that the brain is repeating a similar rhythm over a wider area under a preferred condition.

As discussed in previous sections, a number of studies have found greater  $\tau_e$  values of the ACF of the alpha wave at the preferred stimulus than those at a relatively less preferred one. Significantly larger values of  $\tau_e$  that appear for the alpha wave indicate that the brain is repeating a similar rhythm under these preferred conditions. Thus, the brain repeats a similar rhythm over a wider range in both brain area and time under a preferred condition.

The CCF of the alpha wave clarified the movement of the alpha wave as being over the scalp from the occipital area (O1, O2) to the temporal area (T3, T4) and to the vertex area (Cz). The flow of the alpha waves in relation to  $|\phi(\tau)|_{\max}$  and  $\tau_m$  under a preferred condition is shown in Fig. 15.14. It is clear that the alpha wave propagated from the reference electrode O1 to other regions. A similar tendency was found when the reference electrode was at O2. The values of  $|\phi(\tau)|_{\max}$  and of  $\tau_m$ , therefore, depend on the distance between the reference and test electrodes.

Relationships between EEG-coherence and mental processes have been reported in numerous studies (e.g., Rappelsberger and Petsche, 1988; Hinrichs and Machleidt, 1992; Petsche, 1996). In all of these studies, the focus was on the interchannel relationships between the power spectra in terms of, for example, synchronization of alpha frequency. Here, we have concentrated on the factor in the time domain because some applications of the ACF and CCF have indicated their effectiveness as additional tools in gaining a deeper understanding of EEG dynamics.

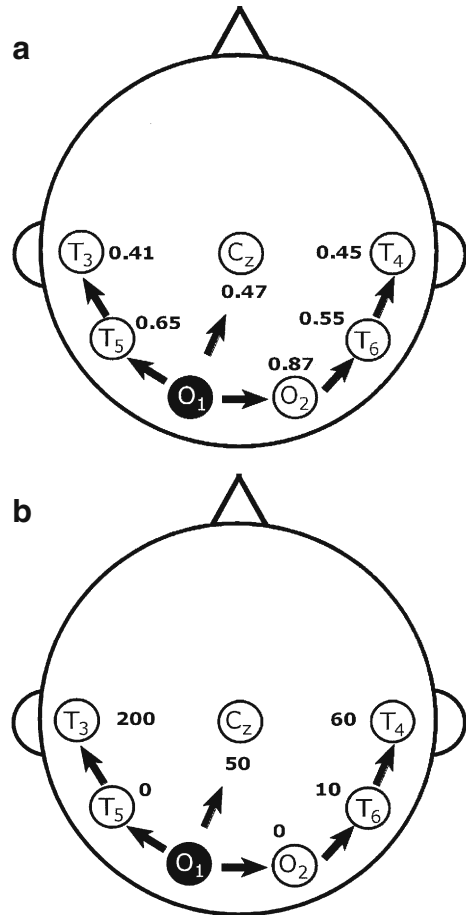


**Fig. 15.13** Cumulative frequency of observed interelectrode signal delays  $\tau_m$  (logarithmic scale) EEG alpha band signals as a function of test electrode position in response to variation of the flicker period under the preferred condition at (a) reference electrode:  $O_1$ ; (b) reference electrode:  $O_2$  and under the non-preferred condition at (c) reference electrode:  $O_1$ ; (d) reference electrode:  $O_2$ . Test electrodes are indicated by  $O_1$  ( $\circ$ );  $T_5$  ( $\square$ );  $T_3$  ( $\square$ );  $O_2$  ( $\bullet$ );  $T_6$  ( $\blacksquare$ );  $T_4$  ( $\blacktriangle$ );  $C_z$  ( $\circ$ )

The results here lead us to the following conclusions:

1. When the period of the flickering light is varied, the preferred stimulus has significantly greater values of  $|\phi(\tau)|_{\max}$  than those of the less preferred stimulus. Together with the result shown above, we conclude that in the preferred condition, the alpha wave repeats over a certain time and this activity spreads over a certain area of the brain.
2. The value of  $|\phi(\tau)|_{\max}$  decreases with increasing distance between the reference and test electrodes.
3. The value of  $\tau_m$  increases in a stepwise fashion with the distance between the reference and test electrodes. This suggests there are discrete nuclei in the central system.

**Fig. 15.14** Flow of EEG alpha waves across both hemispheres as revealed through crosscorrelation comparisons between signals at occipital location  $O_1$  and six other locations. The EEG signals were generated in response to variation of flicker period under a preferred condition. **(a)** The averaged value of maximal crosscorrelation value  $|\phi(\tau)|_{\max}$  for the six pairs of electrode sites, an indication of how similar the alpha band signals are to each other. **(b)** The median (50%) value of interelectrode signal delays  $\tau_m$ (ms) between the six pairs of electrode sites, an indication of the propagation time of the waves



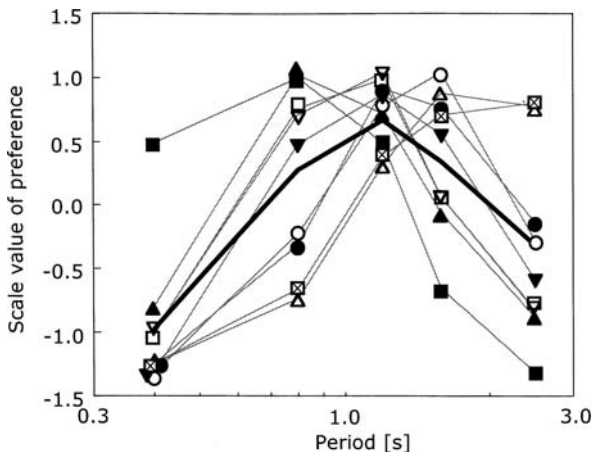
## 15.2 MEG Correlates of Preferences for Flickering Lights

The MEG was recorded during presentations of the most preferred and less preferred flickering lights alternately during change of the temporal factor. Results showed that (1) the effective duration of the ACF,  $\tau_e$ , was longer during the preferred condition than during the less preferred conditions and (2) such results were significant in the left hemisphere when the temporal factor was varied.

### 15.2.1 MEG Correlates of Sinusoidal Flicker

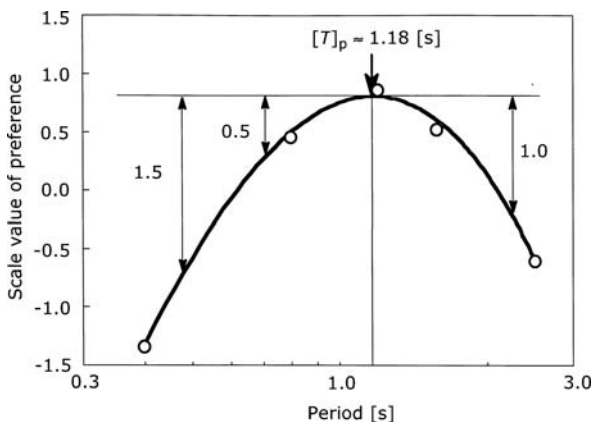
To investigate human cortical responses that correspond with subjective preference and hemispheric specialization for visual stimulus, the ACF of the MEG in relation

**Fig. 15.15** Scale preference values of flicker period  $T$  for different subjects. Different symbols represent preferences of different subjects. The *thick black line* shows averaged scale values of preference for all subjects tested. This reconfirmed the applicability of Equation (15.2)



to the period of a flickering light was analyzed (Soeta et al., 2002). The scale values of individual preference obtained by the PCT together with the averaged scale values of eight subjects are shown in Fig. 15.15. Effects of period on the scale values of preference were examined for all eight subjects by using one-way ANOVA. Results clearly indicated that the effects of period were significant ( $p < 0.01$ ), and the most preferred period,  $[T]_p$ , for each subject was estimated by fitting a suitable polynomial curve to a graph on which the scale values were plotted as shown in Fig. 15.16. The preferred period ranged from 0.6 to 2.0 s, and its averaged value was roughly 1.0 s.

Figure 15.17 shows an example of a recorded MEG alpha wave. We selected 16 channels for the ACF analysis that were located around the occipital area and analyzed each response to the single stimulus in the pair for each subject. The selected



**Fig. 15.16** An example of obtaining the most-preferred period  $[T]_p$  and less preferred periods of the flickering light.  $[T]_{0.5} \approx 0.71$ ,  $[T]_{1.0} \approx 2.22$ , and  $[T]_{1.5} \approx 0.48$  (s). (Suffix number of  $[T]$  denotes the differences between scale values of preference)



**Fig. 15.17** Examples of recorded MEG alpha rhythms. Response durations were 2.5 s

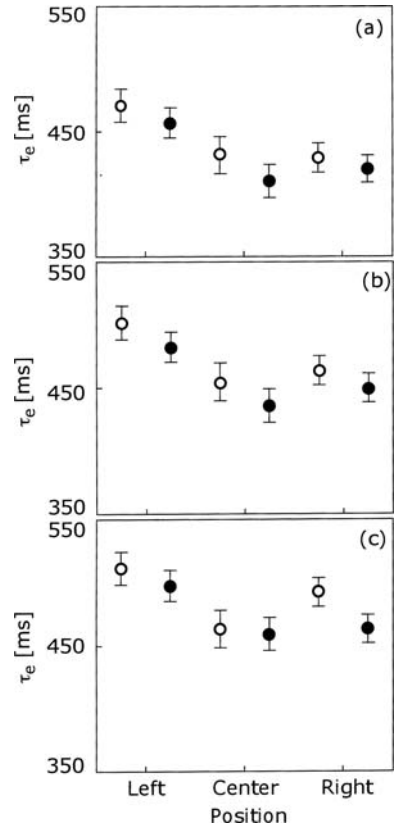
16 channels were divided into three areas as shown in this figure. Table 15.5 shows the results of the two-way ANOVA for  $\tau_e$ ,  $\Phi(0)$ , and  $\phi_1$  values of the alpha wave obtained for the eight subjects. Significant effects of preference were found on  $\tau_e$  and  $\phi_1$  values under all conditions tested. It is clear that the values of  $\tau_e$  correlated with the value of  $\phi_1$  ( $r = 0.75$ ,  $p < 0.01$ ). The values of  $\tau_e$  and  $\phi_1$  for the most preferred stimuli were larger than those for less preferred stimuli for all subjects, as shown in Figs. 15.18 and 15.19, respectively. There were no clear differences, however, in the value of  $\Phi(0)$  as shown in Fig. 15.20. Such a difference of averaged values of  $\tau_e$  and  $\phi_1$  from the left area was significantly larger than those from the central and right areas ( $p < 0.01$ ). The preferred period of the flickering light clearly induces a much longer  $\tau_e$  in the alpha wave than that of the less preferred ones. This

**Table 15.5** Results of two-way ANOVA on the values of  $\tau_e$ ,  $\Phi(0)$ , and  $\phi_1$  under three conditions

Difference of scale value of preference	Factor	$\tau_e$			$\Phi(0)$			$\phi_1$		
		F value	Significance level	F value	Significance level	F value	Significance level	F value	Significance level	
(1) 1.5	Preference	7.9	< 0.01	0.7	0.40	4.1	< 0.05			
	Measurement position	31.8	< 0.001	187.1	< 0.001	48.2	< 0.001			
	Preference and measurement position	0.5	0.61	0.2	0.81	0.2	0.80			
(2) 1.0	Preference	9.6	< 0.005	0.6	0.43	4.9	< 0.05			
	Measurement position	26.6	< 0.001	166.0	< 0.001	26.9	< 0.001			
	Preference and measurement position	0.1	0.88	0.5	0.61	0.5	0.63			
(3) 0.5	Preference	6.9	< 0.01	8.1	< 0.005	15.7	< 0.001			
	Measurement position	19.1	< 0.001	139.7	< 0.001	37.9	< 0.001			
	Preference and measurement position	1.5	0.23	0.0	0.99	1.1	0.31			



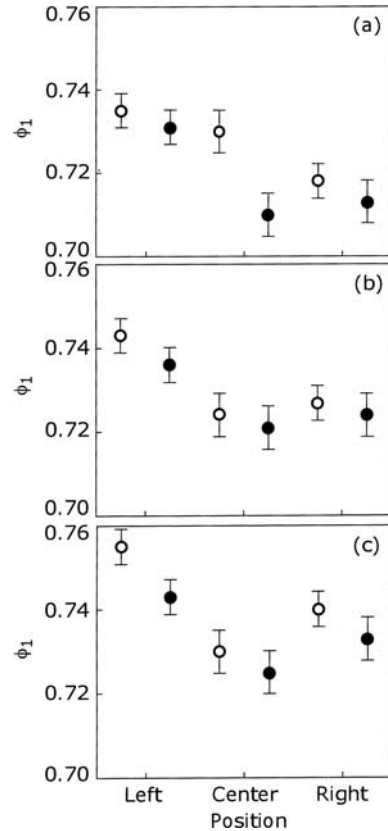
**Fig. 15.18** Effective durations of averaged MEG alpha wave rhythms in response to flickering light for the different occipital sensor regions shown in Fig. 15.17. Error bars represent 95% confidence. The difference of the scale value of preference was (a) 1.5. (b) 1.0. (c) 0.5.  $\circ$ : Higher preference;  $\bullet$ : lower preference



tendency for longer  $\tau_e$  for the preferred period of the flickering light rather than  $\phi_1$  in the alpha wave was also found in a Section 15.1 on EEG. The fact that the brain repeats a similar rhythm under preferred conditions was reconfirmed.

The ratio of high to low preference in terms of the averaged value of  $\tau_e$  obtained here was small in the range 1.01–1.06, but the difference is significant. This is much smaller than that derived in the study on EEG, which was 1.18–1.74. The EEG results from extracellular volume currents triggered mainly by the postsynaptic potential. MEG signals are thought to arise from the intracellular currents that flow from dendritic trees to cell bodies in large numbers (> 50,000) neurons. The MEG and EEG field distributions are mutually orthogonal. Only the current that has a component tangential to the surface of a spherically symmetric conductor produces an exterior magnetic field; the radial source is thus externally silent. Therefore, it one could think that the radial source might more directly reflect the neural activity patterns associated with preferred visual stimuli. Differential results from EEG and MEG for the analysis of human cognition have been discussed previously (Eulitz et al., 1997). Also, numerous studies have reported relationships between EEG and MEG coherence and mental processes (Rappelsberger and Petsche, 1988;

**Fig. 15.19** Values of MEG alpha rhythm regularity  $\phi_1$  in response to flickering light for the three occipital sensor regions shown in Fig. 15.17. Error bars represent 95% confidence. The difference of the scale value of preference was (a) 1.5. (b) 1.0. (c) 0.5.  $\circ$ : Higher preference;  $\bullet$ : lower preference



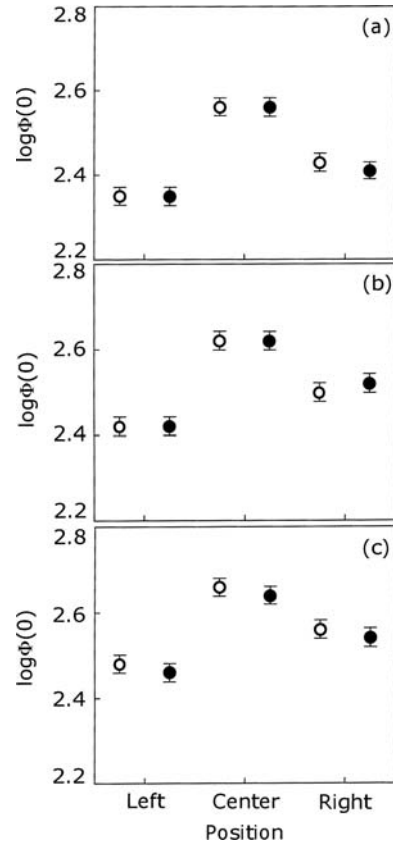
Hinrichs and Machleidt, 1992; Petsche, 1996). Those studies concentrated on the interchannel relations, for example, synchronization of alpha frequency rhythms.

Here, we found that the values of  $\tau_e$  and  $\phi_1$  from the left occipital area were significantly larger than those from the central and right occipital areas. Such a clear tendency was not found in the previous study on EEG mentioned above. The significant values  $\tau_e$  and  $\phi_1$  in the left hemisphere may reflect the specialization of the human brain, reconfirming specifically the left-hemisphere dominance for the temporal factors. Average alpha band signal amplitudes  $\Phi(0)$  from the right occipital area were significantly larger than those from the left occipital area. It is remarkable that the ratio of high to low preference in terms of  $\tau_e$  is greater than that of the value of  $\phi_1$ . This indicates that the effective duration  $\tau_e$  of alpha waves reflects subjective preference better than does alpha rhythm regularity  $\phi_1$ .

These MEG experiments investigated human cortical responses corresponding to subjective preferences for flickering lights. The conclusions that we drew from these experiments are:

1. We reconfirmed that the most preferred flicker periods  $[T]_p$  for individuals were 0.6 – 2 s, with an average value around 1 s.

**Fig. 15.20** Values of MEG alpha band power  $\Phi(0)$  in response to flickering light for the three occipital sensor regions shown in Fig. 15.17. Error bars represent 95% confidence. Differences of scale value of preference were (a) 1.5. (b) 1.0. (c) 0.5.  $\circ$ : Higher preference;  $\bullet$ : lower preference



2. Preferred flicker stimuli evoke significantly longer effective durations  $\tau_e$  of MEG alpha waves than flicker stimuli that are less preferred. There were no clear differences, however, in the average amplitudes of MEG alpha-band signals  $\Phi(0)$ .
3. Averaged values of effective duration  $\tau_e$  and regularity  $\phi_1$  of alpha rhythms from the left area were significantly larger than those from central and right areas.
4. The ratio of high to low preference in terms of effective alpha rhythm duration  $\tau_e$  is greater than that of the value of alpha rhythm regularity  $\phi_1$ . The effective duration of alpha rhythms is a better predictor of preference than the regularity of the rhythms.

### 15.2.2 MEG Correlates of Fluctuating Flicker Rates

As discussed in Section 14.1, the most preferred fluctuation of the flickering light was found in the condition of flicker regularity  $[\phi_1]_p \approx 0.46$ . Okamoto et al. (2007) investigated MEG alpha (8–13 Hz), theta (4–8 Hz) and beta (13–30 Hz) rhythms over the entire head in response to differing degrees of flicker fluctuation.

Results indicated the following:

1. Only in the alpha-wave range of the MEG, was it found that the values of  $\tau_e$  of the MEG signals in the alpha range for the most preferred stimuli are longer when compared with those of the relatively less preferred stimuli.
2. This tendency is most pronounced in MEG signals from left occipital sensors (left hemisphere neural responses) where preferred flicker fluctuations that change the regularity of alpha rhythms also increase their effective durations.

### 15.3 EEG Correlates of Preferences for Oscillatory Movements

This study focused on the EEG correlates of subjective preference for oscillatory horizontal movements of a single target. The EEG was recorded during the presentation of the most or least preferred moving stimuli. The effective duration  $\tau_e$  in the ACF of the alpha wave was analyzed. Results shows that the value of  $\tau_e$  at the most preferred condition was longer than that for stimulus in the less preferred conditions. In addition, the maximum value of the CCF ( $|\phi(\tau)|_{\max}$ ) between EEGs recorded at different electrode sites was analyzed. Results show that the value of  $|\phi(\tau)|_{\max}$  of the alpha wave at the most preferred condition was greater than that of the stimulus in the less preferred conditions. These reconfirm that at the preferred condition, the brain repeats the rhythm in the alpha-wave range in the time domain and that this activity spreads over a wider area of the human cerebral cortex.

The aim of this study was to identify the relationship between human brain response and subjective preference of a single circular target moving sinusoidally in the horizontal direction varying its period (Okamoto et al., 2003). The period is a temporal factor (see Section 14.2). The EEG was recorded during the presentation of stimuli having the most and least preferred periods. Then, the effective durations of EEG alpha rhythms and their correlated behavior at different electrode sites were analyzed to determine the relationship with subjective preference.

Stimuli consisting of white disks with the visual angle of  $1.0^\circ$  in diameter against a black background (Fig. 14.6) were presented on the display placed in front of the subject at the viewing distance of 1.0 m in the dark chamber. The amplitude was fixed at the visual angle of  $0.5^\circ$ . Whereas the most preferred period of the stimuli  $[T]_p$  was found by the PCT at approximately 1.0 s, the subjective preference ratings decreased for shorter and longer periods (Fig. 14.7).

To clarify the effect of subjective preference on EEG, stimuli with the most and least preferred periods were selected as paired stimuli. The thick line shows the preference evaluation curve obtained from the results above. Two pairs of stimuli were presented to determine if either the scale value of subjective preference or the period (velocity) of movement of the stimulus had an influence on the EEG. The pairs were selected as pair 1 (period  $T = 1.0$  and  $0.4$  s) and pair 2 (period  $T = 1.0$  and  $4.0$  s). Eight 22- to 26-year-old subjects participated in this study. All had normal or corrected-to-normal vision.

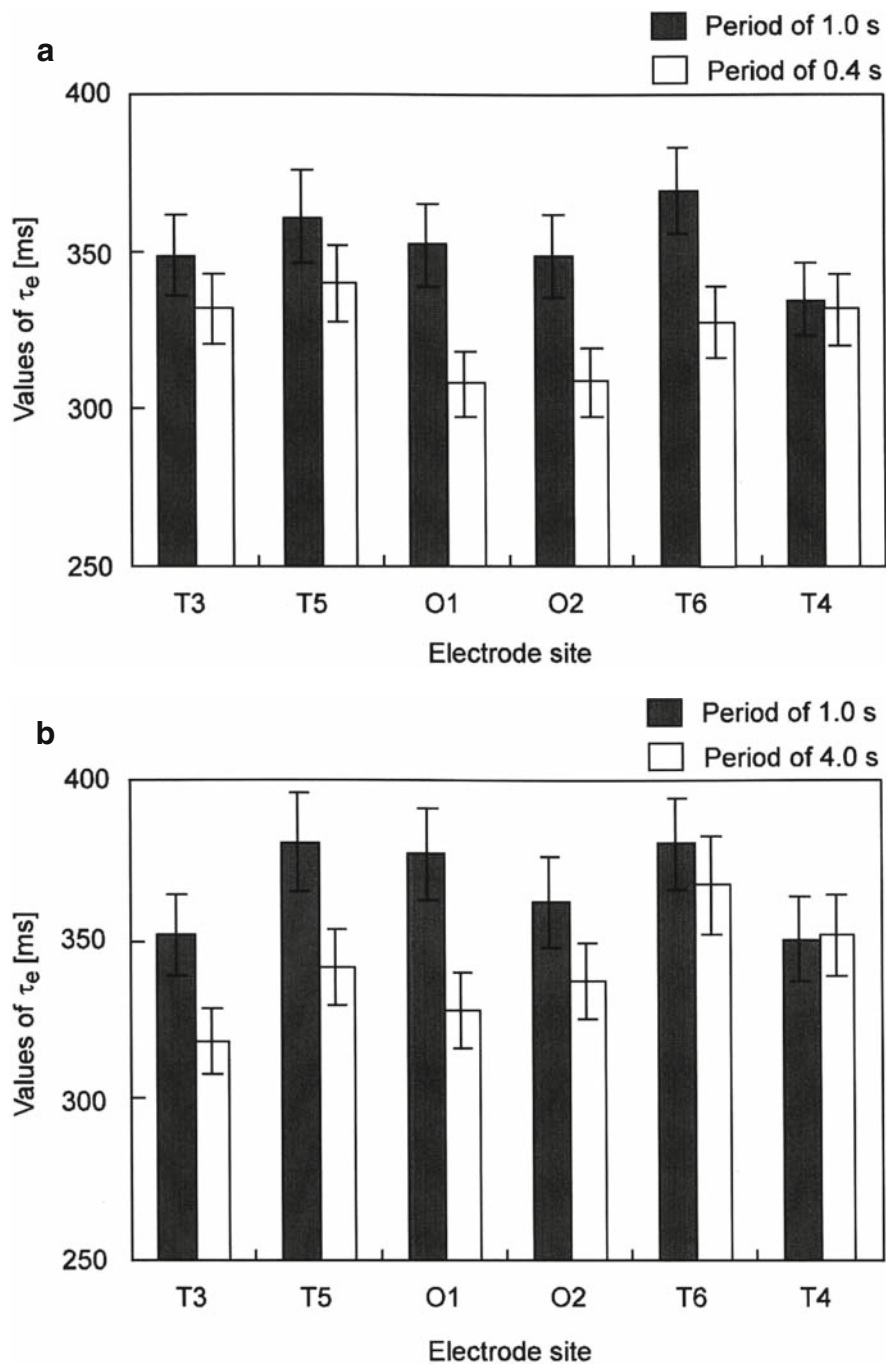
The EEG was recorded using six silver electrodes located at T3, T4, T5, T6, O1, and O2 according to the 10–20 International Electrode Placement System. The

reference electrode was placed at Fz (midline electrodes 30% of distance from nasion toinion). The ground electrode was placed on the forehead. The recorded data were filtered: 4–8 (theta), 8–13 (alpha), and 13–30 (beta). The integration interval was set at 2.5 s in the CCF analysis as well as in the ACF analysis. An example of a normalized CCF is demonstrated in Fig. 15.9. Effects of the stimulus condition and electrode site on the value of  $\tau_e$  were assessed by two-way ANOVA (Table 15.6). A significant effect of the stimulus condition was found in the alpha-wave range when both pair 1 and pair 2 were presented ( $p < 0.01$ ). There were no significant effects, however, in the theta and beta ranges. Figure 15.21 shows the averaged value of  $\tau_e$  in the alpha range for the eight subjects. Results show that the stimulus with the period of 1.0 s induced a longer value of  $\tau_e$  than did those with periods of 0.4 or 4.0 s.

**Table 15.6** Results of two-way ANOVA for the values of  $\tau_e$  of the ACF

Frequency band (Hz)	Factor	F-ratio	P value	F-ratio	P value
		Pair 1 Period of 1.0 and 0.4 s		Pair 2 Period of 1.0 and 4.0 s	
4 to 8	Stimulus condition	0.03	0.87	0.02	0.90
	Electrode site	0.19	0.97	0.96	0.44
	Stimulus condition and electrode site	0.32	0.90	0.24	0.95
8 to 13	Stimulus condition	14.77	< 0.001	11.72	< 0.001
	Electrode site	1.14	0.34	1.85	0.10
	Stimulus condition and electrode site	0.87	0.50	0.96	0.44
13 to 30	Stimulus condition	0.35	0.55	<0.001	0.99
	Electrode site	4.00	0.001	5.30	< 0.001
	Stimulus condition and electrode site	0.90	0.48	0.42	0.84

The values of  $|\phi(\tau)|_{\max}$  extracted from the CCF also were analyzed to estimate the degree of correlation between cortical responses of different positions. The normalized CCF between the brain waves measured at electrode site O1 (reference electrode) and those measured at the other electrode sites (test electrodes) was analyzed because the difference between the values of  $\tau_e$  for the two conditions was “the greatest at O1 (left hemisphere)” in the ACF analysis as shown in Fig. 15.21. The effects of the stimulus condition and electrode site on the  $|\phi(\tau)|_{\max}$  values were assessed by two-way ANOVA in each frequency range (Table 15.7). Significant effects of the stimulus condition and electrode site were reconfirmed in the alpha-wave range only when both pairs were presented ( $p < 0.01$ ); namely, no significant effects from the stimulus conditions in the theta and beta ranges. As shown in Fig. 15.22, results reconfirmed that the stimulus with the period of 1.0 s had a greater value of  $|\phi(\tau)|_{\max}$  in the alpha range than did those with periods of 0.4 or 4.0 s ( $p < 0.01$ ). As shown in the figure, the values of  $|\phi(\tau)|_{\max}$  are greater in the



**Fig. 15.21** Averaged effective durations  $\tau_e$  EEG alpha rhythms range in response to a change of flicker period at different electrode sites. Presentation of (a) pair 1 ( $T = 1.0$  and  $0.4$  s) and (b) pair 2 ( $T = 1.0$  and  $4.0$  s). The preferred flicker period  $T$  is  $1.0$  s. Error bars indicate the standard error of the mean

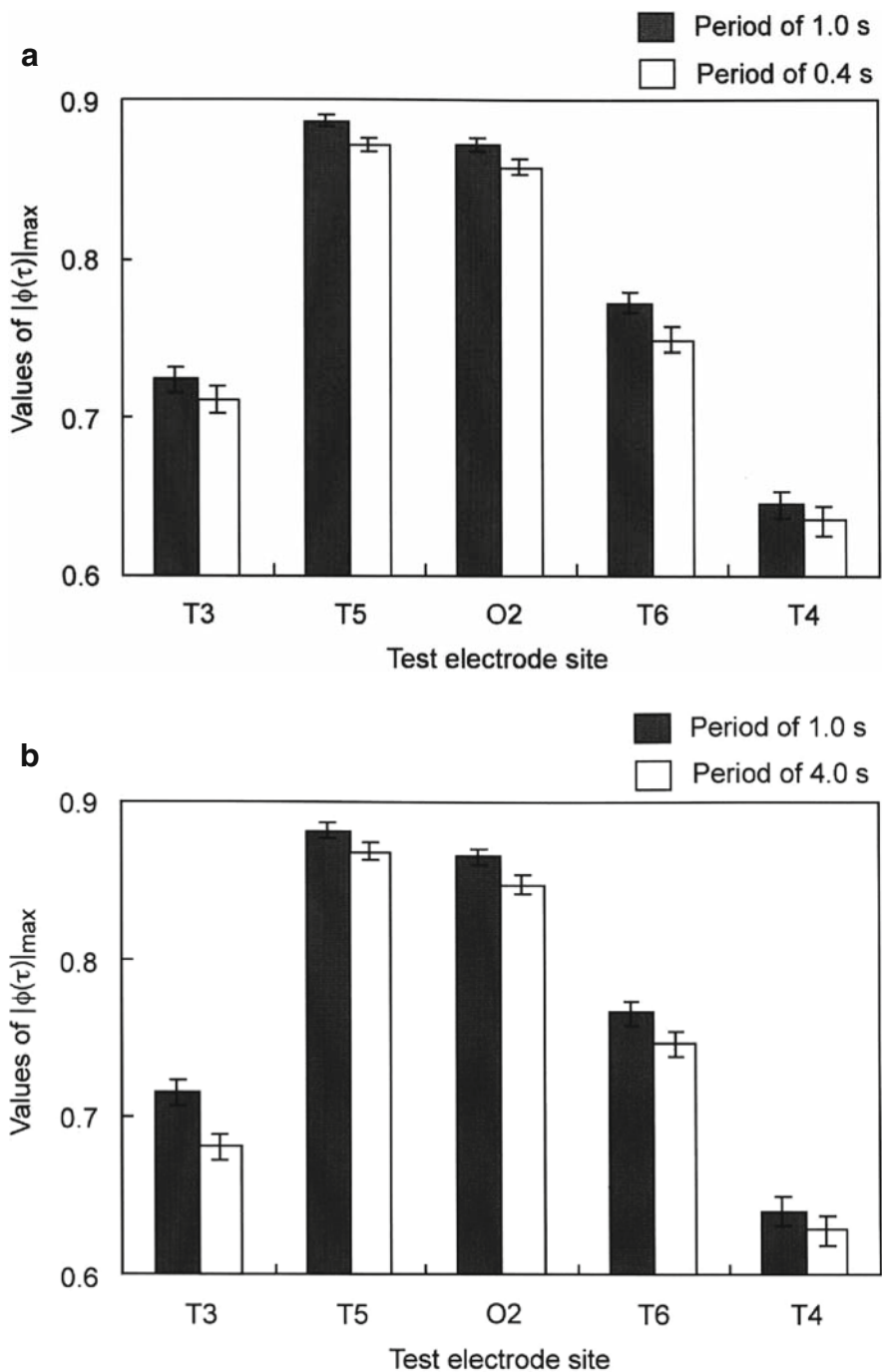
**Table 15.7** Results of two-way ANOVA for the values of  $|\phi(\tau)|_{\max}$  of the CCF

Frequency band (Hz)	Factor	Pair 1		Pair 2	
		F-ratio	P value	F-ratio	P value
		Period of 1.0 and 0.4 s		Period of 1.0 and 4.0 s	
4 to 8	Stimulus condition	0.15	0.70	0.002	0.96
	Electrode site	223.07	<0.001	217.34	<0.001
	Stimulus condition and electrode site	0.05	0.99	0.24	0.92
8 to 13	Stimulus condition	11.06	<0.001	17.73	<0.001
	Electrode site	409.40	<0.001	398.03	<0.001
	Stimulus condition and electrode site	0.23	0.92	0.73	0.57
13 to 30	Stimulus condition	0.09	0.76	0.81	0.37
	Electrode site	402.07	<0.001	428.59	<0.001
	Stimulus condition and electrode site	0.21	0.93	0.08	0.09

posterior temporal and the occipital areas (T5, O2, respectively) than at the other sites for both pairs 1 and 2.

The differences between the scale values of the stimuli with the period of 1.0 s and the stimuli with periods of 0.4 or 4.0 s were not the same for all individual subjects because the stimuli were selected on the averaged scale value of preference obtained from results of previous subjective preference tests. Results indicate that the value of  $\tau_e$  in the alpha range was affected by subjective preference, but not the period of the stimulus itself. The value of  $\tau_e$  of the ACF in the alpha range, which indicates the degree of persistence of the EEG alpha wave, is prolonged at the preferred condition. This may be caused by the brain repeating the rhythm in the alpha range, which reflects temporal behavior in the cortical area.

This tendency has been found also in previous studies as well as in the effects of varying  $\Delta t_1$ ,  $T_{\text{sub}}$  of the sound field (see Section 14.3) and in varying the period of the flickering light (see Section 15.1). In the previous study on the relationship between EEG alpha waves and subjective preference with changes of the period of the flickering light, the value of  $\tau_e$  was longest in the occipital area. In the current study, however, the value of  $\tau_e$  was longest not in the occipital area but in the posterior temporal area. This may be a result of experimental conditions: use of a fixed LED in the previous experiment compared with use of visual motion stimuli in this experiment. A study using human positron emission tomography (PET) found that visual system area V5 is located at the occipital-temporal parietal border (Zeki et al., 1991). The processing of visual motion stimuli in the human visual cortex area V5 has been investigated by using PET, MEG, and EEG. Probst et al. (1993) found that the locations of cortical activation during the presentation of motion stimuli were more lateral than those during the presentation of pattern-reversal (nonmotion) stimuli.



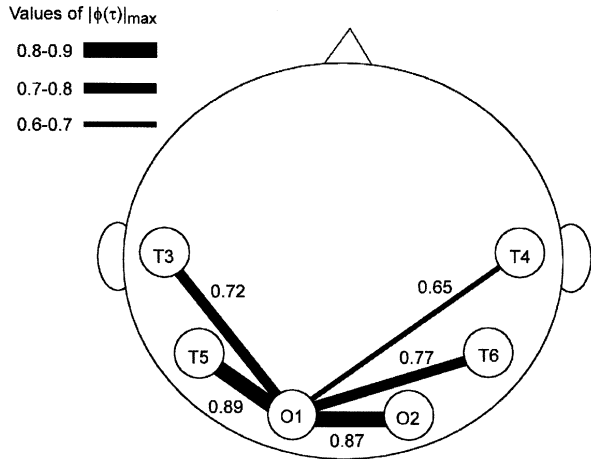
**Fig. 15.22** Crosscorrelation analysis of EEG alpha band signals at electrode occipital site O1 and five other sites. Averaged values of maximal crosscorrelation magnitude  $|\phi(\tau)|_{\max}$  in response to a change of flicker period under presentation of (a) pair 1 ( $T = 1.0$  and  $0.4$  s) and (b) pair 2 ( $T = 1.0$  and  $4.0$  s). The preferred flicker period  $T$  is  $1.0$  s. Error bars indicate the standard error of the mean



A significant effect of the stimulus condition on the values of  $|\phi(\tau)|_{\max}$  was observed in the alpha range. The value of  $|\phi(\tau)|_{\max}$  averaged in the alpha range for the stimuli with the period of 1.0 s was significantly greater than for stimuli with periods of 0.4 or 4.0 s (Fig. 15.22). We reconfirmed that the values of  $|\phi(\tau)|_{\max}$  in the alpha range are related to subjective preference, but not by the period of stimuli itself. The fact that the alpha waves have greater  $|\phi(\tau)|_{\max}$  values shows that the brain repeats its rhythm in the spatial domain over a wider cortical area in the preferred condition than it does in the less preferred condition. This tendency toward greater  $|\phi(\tau)|_{\max}$  values in the alpha range was also found in the effect of varying the period of the flickering light as discussed above.

In the current study, because O1 (left hemisphere) was chosen as the reference electrode, the value of  $|\phi(\tau)|_{\max}$  decreased as the distance between O1 and the other electrodes increased. As shown in Fig. 15.23, values of  $|\phi(\tau)|_{\max}$  were greater in the posterior temporal and occipital areas, T5 and O2, respectively, than in the other areas. These results were observed not only in the alpha but also in the theta and beta ranges.

**Fig. 15.23** Values of maximum alpha rhythm crosscorrelation between EEG recording sites during the presentation of pairs of oscillating moving visual targets. Correlations between signals recorded at different electrode positions and at position O1 are shown for the stimulus pair 1 (oscillatory period  $T = 1.0$  s vs. 0.4 s). The relative thickness of the bars indicates the range of values for the maximum crosscorrelation magnitude  $|\phi(t)|_{\max}$



Thus far, we have analyzed the EEG to determine the relationship between the human brain response and subjective preference for horizontal visual motions of movement over varying periods. Results obtained regarding the temporal and the spatial features of brain rhythms can be summarized as

1. The value of  $\tau_e$  in the alpha-wave range for stimulus with the most preferred period of the movement was significantly longer than that for the stimulus with less preferred periods.
2. This tendency was greatest at electrode site O1 over the left hemisphere.
3. The value of  $|\phi(\tau)|_{\max}$  in the alpha range for stimuli with the most preferred period was significantly greater than that for stimuli with the less preferred period.

These results reconfirm that the brain repeats the rhythm in the alpha range and that this activity spreads wider over the human brain cortex as a result of the presentation of stimuli with preferred motion rather than with the less preferred motion.

## 15.4 Hemispheric Specializations in Vision

We summarize factors extracted from the ACF of the alpha wave corresponding to subjective preference. The specialization of the human cerebral hemispheres for temporal factors of the visual field has not been reported previously, and is presented here for the first time.

We have investigated relationships between alpha rhythms in EEG and MEG recordings and subjective preferences for various aspects of visual stimuli. The effective durations of these alpha rhythms were always longest in all subjects ( $p < 0.01$ ) when the preferred flicker rate was presented. Remarkably, averaged values of  $\tau_e$  and  $\phi_1$  extracted from the ACF of the MEG alpha wave from the left area were significantly larger than those from the central and right areas ( $p < 0.01$ ). Also, effects of the subjective preference and measurement position on values of  $\tau_e$ ,  $\Phi(0)$ , and  $\phi_1$  extracted from the ACF of the MEG alpha wave (eight subjects) were examined, when a single circular target was moved in the horizontal direction. Effective durations were longer and the alpha rhythms more regular, with larger  $\phi_1$ , for the preferred stimuli than those less preferred ( $p < 0.01$ ). It is obvious that the value of  $\tau_e$  correlated relatively highly with the value of  $\phi_1$  ( $r = 0.72$ ,  $p < 0.01$ ) but only very weakly with the value of  $\Phi(0)$  ( $r = 0.23$ ). We found the averaged values of  $\tau_e$  and  $\phi_1$  from the left (O1) area to be significantly longer than those from the central and right areas (O2).

We analyzed the MEG signals from the entire head in response to visual stimuli in which the temporal factor  $\phi_1$ , which reflects the temporal regularity of repeating waveform patterns, was manipulated using different bandwidths of flicker noise centered on 1 Hz (Okamoto et al., 2007). The results showed that the  $\tau_e$  values of the alpha rhythm observed around the left occipital area were significantly larger for the most preferred stimuli than those for the less preferred stimuli. This tendency indicates that the stimuli in the preferred condition, with regard to changes of the temporal factor in the visual signal, increase the stability of the alpha rhythm around the left occipital area.

We also found in the MEG study that the values of  $\tau_e$  and  $\phi_1$  from the left occipital area were significantly larger than those from the central and right occipital areas, but it should be noted that such tendencies were not found in the EEG study. Thus, we again found that the neural correlates associated with temporal factors to be mainly associated with activity in the left hemisphere, analogous to the case for auditory stimuli.

The study on the relationship between the EEG response and the subjective preference by varying of the period of the horizontal movement of the single target showed the following: The value of  $\tau_e$  of the alpha waves for stimulus at the most

**Table 15.8** Cerebral hemisphere specializations related to visual temporal factors changed

Factors changed	EEG, ACF $\tau_e$ value of alpha wave	MEG, ACF $\tau_e$ value of alpha wave
Period of flickering light, T	L > R (sinusoidal wave) <sup>1,2</sup>	L > R (sinusoidal wave, $p < 0.01$ )
Fluctuation of flickering light, $\phi_1$		L > R (sinusoidal wave and band-pass noise, $p < 0.05$ )
Period of moving target, T	L > R (sinusoidal wave, $p < 0.01$ ) <sup>1</sup>	–

<sup>1</sup>A flow of EEG alpha wave from the left hemisphere to the right hemisphere for stimuli of both the flickering light and the moving target in change of the period was observed by the CCF  $|\phi(\tau)|_{\max}$  between alpha waves recorded at different electrodes.

<sup>2</sup>A significant difference was not achieved.

preferred condition was longer than that for stimulus at the less preferred conditions ( $p < 0.01$ ). This tendency was maximum at O1 in the left hemisphere. The value of  $|\phi(\tau)|_{\max}$  of the alpha waves for the most preferred oscillatory movements was greater than that for the stimulus in the less preferred conditions ( $p < 0.01$ ). Considering this fact together with the similar repetitive feature in the alpha-wave increase, the brain repeats the “alpha rhythm” in the time domain, and this activity spreads wider over the area of the brain cortex at the preferred stimulus condition.

Table 15.8 summarizes our experimental EEG and MEG findings related to hemispheric lateralization of neuronal responses with respect to temporal factors of the visual field (rates of flicker and oscillatory movements). It has been reported that the left cerebral hemisphere is greatly concerned with linear, sequential modes of thinking, such as speech and calculation. In contrast, the right hemisphere tends to perceive space in multiple-dimensional and nontemporal terms for the visual field (Sperry, 1974; Davis and Wada, 1974; Galin and Ellis, 1975; Levy and Trevarthen, 1976). Our observations concerning hemispheric lateralizations of responses related to temporal and spatial factors of sounds and sound fields and their subjective preferences (Section 5.6), i.e., processing of temporal factors in the left and spatial factors in the right hemisphere, appear to be broadly consistent with these earlier observations by others.

## Chapter 16

# Summary of Auditory and Visual Sensations

This chapter summarizes both auditory and visual sensations including both temporal and spatial aspects of these percepts in each modality. Each temporal and spatial attributes has a corresponding feature in its respective correlation function.

In the auditory modality, temporal factors are associated with the monaural perceptual qualities of pitch, timbre, and duration and changes in sound quality related to first reflection time, and subsequent reverberations. Auditory temporal factors are those parameters that can be derived from features of autocorrelation functions (ACFs). Auditory spatial factors, on the other hand, are associated with the binaural percepts of sound direction in the horizontal plane, apparent source width and envelopment, as well as loudness. Auditory spatial factors are parameters derived from features of interaural correlation functions (IACFs).

In the visual modality, temporal factors are associated with temporal aspects of visual perception, such as perceived rates of flickering lights and oscillating movements, whereas spatial factors are associated with spatial vision, such as the perception of forms and textures. Temporal visual factors are derived from the features of the temporal autocorrelation functions (ACFs) of changing visual signals, and spatial factors are derived from the spatial autocorrelation functions of static visual patterns.

Thus there may exist deep commonalities between the two modalities, both in the structure of their respective perceptual spaces and in possible similarities in the correlation-based neural information processing mechanisms that subserve their respective percepts. It appears that the behavior of a great many auditory percepts and at least a few visual ones can be explained in terms of features of these various correlation-based representations. In more than a few cases, features of correlation-based representations have more obvious and direct connections to perceptual qualities than do their frequency-domain, spectral counterparts.

Cerebral hemispheric specialization may play an important role in accounting for the independent effects that temporal and spatial factors on their corresponding percepts (temporal and spatial sensations). There may be two distinct modes of representation, temporal and spatial, that are lateralized at the cortical level, such that this lateralization may at least partly explain their high degree of functional independence.

There also appear to be common, possibly universal, neural response correlates of subjective preference that hold for both temporal and spatial factors and their associated perceptual attributes in both the auditory and visual modalities. These are related to the persistence, temporal coherence, and spatial extent of alpha rhythms in auditory and visual cortical populations. The alpha rhythms are signs of subjective preference, but it remains to be explained why some of the factors themselves have preferred values that generate primitive “aesthetic” valences. Some examples of the latter are subjective preferences for vibrato of musical notes, fluctuations in the rates of flickering lights, and less-than-absolute regularities in textures.

## 16.1 Auditory Sensations

### 16.1.1 Auditory Temporal Sensations

Table 16.1 presents an overview of auditory temporal sensations, such as loudness, pitch, timbre and duration, which are described by temporal factors that are features of the autocorrelation functions (ACFs) of sound signals. We want to be as clear as possible about the various dependencies of percepts on single and multiple factors. Loudness, for example, is a function of several factors, most notably sound energy  $\Phi_p(0)$  and dominant periodicity or frequency  $\tau_1$ , but also effective duration  $\tau_e$  (or  $\phi_1$ ) as expressed by Equation (6.8). When the sound signal arrives from only directly in front, timbre is described by the factor of  $W_{\phi(0)}$  extracted from the monaural autocorrelation function ACF. If the sound signal arrives from any direction other than

**Table 16.1** Auditory temporal sensations in relation to the temporal factor extracted from the ACF. The resulting equation, figures or section are indicated at respective sensation

Temporal factors		Temporal sensations			
		Loudness (Equation (6.9))	Pitch (Equation (6.7))	Timbre (Figs. 6.24 and 6.25)	Duration sensation (Section 6.5)
<b>ACF</b>	$\Phi_p(0)$	$\mathbf{X}^{1,2}(\text{SPL})$			
	$\tau_1$	$\mathbf{X}(\tau_1)$	$\mathbf{X}(F = 1/\tau_1)$		$\mathbf{X}(\tau_1)$
	$\phi_1^3$		$\mathbf{X}(\text{Strength of pitch})$		
	$\tau_e$	$\mathbf{X}(\tau_e)$			
	$W_{\phi(0)}$			$\mathbf{X}(W_{\phi(0)})$	
	D				$\mathbf{X}(D)$

<sup>1</sup>**X**: Major factors, which have been discussed in this book and their corresponding temporal sensations.

<sup>2</sup>It is obvious that loudness depends on the sound pressure level (SPL).

<sup>3</sup>**X**: Note that factors of  $\phi_1$  and  $\tau_e$  are mutually related. It is note worthy that temporal sensations are associated with the left hemisphere, so that temporal sensations may be affected by two temporal factors of the sound field:  $\Delta t_1$  and  $T_{\text{sub}}$  in addition. For example, Sayles and Winter (2008) have shown that pitch is influenced by the reverberant energy ratio equivalent to the value A (Equation 3.2), which depend on the distance between the sound source and a receiving position.

this, then we must also take spatial factors extracted from the interaural correlation function IACF into consideration. In fact, timbre of the sound field that we investigated using dissimilarity judgments (Section 6.6) may be described as a function of both temporal and spatial factors. In this table, a capital letter X signifies the most important factor(s) for describing a given sensation. A small x indicates a dependent factor that may warrant special attention. Some factors interact – factors  $\phi_1$  and  $\tau_e$  are to some extent related to each other. The factor D signifies the physical duration of the sound signal. Because of temporal summation, the perception of loudness should be described in relation to sound duration D together with other temporal factors. Pitch has been described in terms of the ACF factor  $\tau_1$  (fundamental frequency  $F0 = 1/\tau_1$ ), and its perceived strength is given by its relative amplitude  $\phi_1$ .

### 16.1.2 Auditory Spatial Sensations

Table 16.2 presents an overview of auditory spatial sensations including the apparent source width ASW, subjective diffuseness and sound source localization. Localization in the horizontal plane, for example, is well described in terms of spatial factors extracted from the interaural correlation function IACF. Localization in the median plane, however, does not necessarily involve the binaural representation (if only because the IACF is little changed by elevation) and can be described in terms of temporal factors extracted from the monaural autocorrelation function ACF alone (Session 7.1.2).

The apparent source width ASW of band-pass filtered noise has been given by Equation (7.4). We have little data on effects of  $\tau_{IACC}$  on this sensation, which might be added to the expression. However, as far as acoustic design of concert halls is concerned, listening level LL is usually governed by the power watt level PWL of the sound source in a room, a parameter that musicians can control. But,

**Table 16.2** Auditory spatial sensations in relation to the factor extracted from the IACF

Spatial factors		Spatial sensations		
		ASW (Section 7.2)	Subjective diffuseness (Section 7.3)	Localization in the horizontal plane (Section 7.1)
<b>IACF</b>	LL <sup>1</sup>	<b>X</b> <sup>2</sup>	<b>X</b>	– <sup>3</sup>
	$\tau_{IACC}$	–	–	<b>X</b>
	$W_{IACC}$	$(W_{IACC})^{1/2}$	–	–
	IACC	$(IACC)^{3/2}$	$(IACC)^{3/2}$	<b>X</b>

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

<sup>2</sup> **X**: Major factors to describe corresponding spatial sensations. (The related section is indicated in the parentheses.)

<sup>3</sup> Factors to be examined whether or not they influence the respective sensation. Note that factors for localization in the median plane may be extracted from the ACF as discussed in Section 7.1.2 (Sato et al., 2001).

at the design stage of the sound field in a concert hall, all we can do is calculate the distribution of relative sound pressure level in an enclosure. The condition of  $\tau_{IACC} \approx 0$  is usually realized in a room by producing the strongest direct sound. The ASW for multiband complex noise is given by Equation (7.5). It is worth noting that  $W_{IACC}$  depends mainly on the frequency composition of a sound signal. If the sound signal contains only low frequencies, then the value of  $W_{IACC}$  is large and the ASW becomes wide.

Finally, subjective diffuseness has been well expressed by Equation (7.6). It is interesting that the power of IACC is commonly  $\beta \approx 3/2$  as expressed by Equations (7.4) through (7.6), and the power of  $W_{IACC}$  in Equations (7.4) and (7.5) is  $1/2$ .

### 16.1.3 Auditory Subjective Preferences

Table 16.3 reviews the most significant factors for describing subjective preference, reverberance and speech intelligibility as an overall subjective response to a sound field. They are related to temporal factors extracted from the ACF of the source signal, spatial factors extracted from the IACF, and additional orthogonal factors of first reflection time  $\Delta t_1$  and subsequent reverberations  $T_{sub}$  of the sound field. Table 16.4 lists cerebral hemisphere dominance in relation to the temporal and spa-

**Table 16.3** Subjective preference of the sound field in relation to the factors extracted from both the ACF and IACF, as well as orthogonal factors of the sound field

Temporal and spatial factors		Subjective responses			
		Subjective preference (Listeners) (Sections 3.1, 3.2, and 3.3)	Subjective preference (Performers) (Section 8.4)	Reverberance (Listeners) (Hase et al., 2000)	Speech intelligibility (Listeners) (Sections 9.1 and 9.2)
<b>ACF</b>					
	$\tau_1$				<b>X</b> <sup>1</sup>
	$\phi_1$				<b>X</b>
	$\tau_e$	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>
<b>IACF</b>	$LL^2$	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>
	$\tau_{IACC}$	( = 0) <sup>3</sup>			
	$W_{IACC}$				
	IACC	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>
<b>Sound field</b>	$\Delta t_1$	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>
	$T_{sub}$	<b>X</b>	- <sup>4</sup>	<b>X</b>	<b>X</b>

<sup>1</sup> **X**: Major factors to describe corresponding subjective attributes. (The related section is indicated in the parentheses.)

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

<sup>3</sup> The preferred condition is obtained under the condition of  $\tau_{IACC} = 0$ .

<sup>4</sup> Factors to be examined whether or not they influence the respective response.

**Table 16.4** Hemispheric specializations of temporal and spatial factors observed by analyses of the AEPs, EEG, and MEG

Factors changed	AEP (SVR) A(P <sub>1</sub> – N <sub>1</sub> )	EEG, ratio of ACF τ <sub>e</sub> values of alpha wave	AEP (MEG) N1m	MEG, ACF τ <sub>e</sub> value of alpha wave
<b>Temporal</b>				
Δt <sub>1</sub>	L > R (speech) <sup>1</sup>	L > R (music)		L > R (speech)
T <sub>sub</sub>	–	L > R (music)	–	
<b>Spatial</b>				
LL	R > L (speech)	–	–	
IACC	R > L (vowel /a/)	R > L (music) <sup>2</sup>	R > L (band noise) <sup>3</sup>	
	R > L (band noise)			
τ <sub>IACC</sub>			R > L (band noise) <sup>3</sup>	
Head-related transfer functions			R > L (vowels) <sup>4</sup>	

<sup>1</sup> Sound source used in experiments is indicated in the parentheses.

<sup>2</sup> The flow of EEG alpha wave from the right hemisphere to the left hemisphere for music stimulus in change of the IACC was determined by the CCF |φ(τ)|<sub>max</sub> between alpha waves recorded at different electrodes.

<sup>3</sup> Soeta and Nakagawa (2006).

<sup>4</sup> Palomaki et al. (2002).

tial factors of the sound field, and Table 16.5 shows hemispheric activities related to subjective preference of the sound field. In the preferred conditions for a wide variety of sensations, alpha rhythms persist more coherently over longer durations (larger values of τ<sub>e</sub>) and over wider areas of the brain.

The most preferred conditions of the four orthogonal factors of the sound field are described in Section 3.2. As is expressed by Equation (3.9), the remarkable finding is that both temporal and spatial factors x<sub>i</sub> (i = 1,2,3) are commonly expressed in terms of the 3/2 power of the factor normalized by its most preferred value. But, as far as the IACC is concerned (i = 4), it is expressed in terms of the 3/2 power of its “real value” indicating that it provides the greatest contribution of the orthogonal factors. Dissimilar signals arriving at two ears, that indicate a small value of the IACC, are almost invariably the preferred condition without exception.

### 16.1.4 Effects of Noise on Tasks and Annoyance

Table 16.6 summarizes the effects of noise on the performance of different mental tasks and the contributions of temporal and spatial factors to annoyance. Effects of noise on left hemisphere intensive tasks such as adding numbers may be described in terms of temporal factors extracted from the ACF of the noise stimulus. This can



**Table 16.5** The  $N_2$ -latency of SVR, the ACF  $\tau_e$  values, and the  $CCF-|\phi(\tau)|_{\max}$  values of both EEG and MEG alpha waves, which correspond with subjective preference in relation to each of four orthogonal factors of the sound field

Factor changed	SVR $N_2$ -latency	EEG, ACF $\tau_e$ value of alpha wave	EEG, CCF $ \phi(\tau) _{\max}$ of alpha wave	MEG, ACF $\tau_e$ value of alpha wave	MEG, CCF $ \phi(\tau) _{\max}$ of alpha wave
<b>Temporal</b>					
$\Delta t_1$	L & R <sup>1</sup> (speech) Figures 4.14 and 4.15	Left H. (music) <sup>2</sup> Figures 4.16 and 4.17	—	Left H. (speech) Figures 4.28 and 4.29	L & R <sup>3</sup> (speech) Section 4.4.2
$T_{\text{sub}}$	—	Left H. (music) Figures 4.18 and 4.21	—	—	—
<b>Spatial</b>					
LL	L & R (speech) Figure 4.14	—	—	—	—
IACC	—	Right H. (music) Figures 4.22–4.23	Right H. ( $\Rightarrow$ Left H.) (music) <sup>4</sup> Figure 4.24	L & R (speech) Soeta et al. (2005)	L & R (speech) Soeta et al. (2005)

H., hemisphere.

<sup>1</sup> Sound source used in experiments is indicated in the parentheses.

<sup>2</sup> For example, the  $\tau_e$  value of alpha wave in EEG increased significantly in the left cerebral hemisphere when the scale value of subjective preference of the music sound field was high.

<sup>3</sup> When the scale value of subjective preference was high, then the  $|\phi(\tau)|_{\max}$  value of alpha wave in MEG increased and thus propagated in wide brain areas on both hemispheres.

<sup>4</sup> Flow of EEG alpha wave from the right hemisphere to the left in change of the IACC with music, which was observed by the CCF  $|\phi(\tau)|_{\max}$  between alpha waves recorded at different electrodes.

**Table 16.6** Effects of noise on mental tasks and annoyance in relation to temporal and spatial factors extracted from the ACF and the IACF, respectively

Temporal and spatial factors	Effects of noise on tasks and annoyance of noise			
	Annoyance of noise in change of spatial factors (Section 11.2)	Annoyance of traffic noise (Section 11.1.2)	Left-hemispheric task (Section 11.3)	Right-hemispheric task (Section 11.3)
<b>ACF</b>				
$\tau_1$		$c(\text{Var}_{\tau_1})^1$	$\mathbf{X}^2$	
$\phi_1$		–	$\mathbf{X}^3$	
$\tau_e$		$b(\tau_e)$	$\mathbf{X}$	
D	– <sup>4</sup>	–		
<b>IACF</b>				
LL <sup>5</sup>	b(SPL)	$a(\text{Var}_{\text{SPL}})^6$		$\mathbf{X}$
$\tau_{\text{IACC}}$	$a(\text{Flu}_{\tau_{\text{IACC}}})$	–		$\mathbf{X}$
$W_{\text{IACC}}$		–		$\mathbf{X}$
IACC	$a(\text{Flu}_{\text{IACC}})$	–		$\mathbf{X}$

<sup>1</sup> Symbol of “Var” represents the variance defined by  $\sigma^2 = 1/n \sum (x_i - m)^2$ ,  $x_i$ : the value of data  $i$ ;  $m$ : the mean value, and  $n$ : the number of the data.

<sup>2</sup>  $\mathbf{X}$ : Major factors, which have well described respective subjective attributes in this book.

<sup>3</sup>  $\mathbf{X}$ : Note that factors of signal regularity or periodicity strength  $\phi_1$  and effective duration  $\tau_e$  are mutually related.

<sup>4</sup> –: Factors to be examined in the future, whether or not they influence on respective subjective attributes.

D: Physical duration of the sound signal.

<sup>5</sup>  $LL = 10 \log [\Phi(0)/\Phi(0)_{\text{reference}}]$ , where  $\Phi(0) = [\Phi_{ll}(0) \Phi_{rr}(0)]^{1/2}$ .

<sup>6</sup> SPL was measured at one of two ear entrances or by a single microphone at the center of the head.

<sup>7</sup>In this context, one might hypothesize that music is associated with the left hemispheric factors more than the right hemispheric factors, and noise is associated with the right hemispheric factors more than left hemispheric factors. However, further investigations on hemispheric specialization for single factors, for example,  $\tau_1$  and  $\phi_1$  are needed.

be considered as an interference effect in the left hemisphere. On the other hand, effects of noise on the right hemisphere intensive tasks such as pattern searches might be described in terms of spatial factors extracted from the IACF of the noise field and analogously as interference effects in the right hemisphere.

When spatial factors (IACC and SPL) of noise fluctuate, then these spatial factors become significant as generators of annoyance (Equation (11.3)). And, when  $\tau_{\text{IACC}}$  and SPL fluctuate, annoyance is expressed by Equation (11.5). In the more general example of evaluating the annoyance of traffic noise, fluctuations of parameters should be investigated in relation to both the temporal and the spatial factors of the noise field. In fact, a linear combination of the three variables ( $\text{Var}_{\text{SPL}}$ ,  $\tau_e$ ,  $\text{Var}_{\tau_1}$ ) as approximately expressed by Equation (11.1) has been used to effectively predict the scale value of annoyance of traffic noise.

## 16.2 Visual Sensations

The temporal and spatial sensations as well as the subjective preference in vision are summarized. These sensations are also described in relation to the temporal and spatial factors of their respective correlation functions.

### 16.2.1 Temporal and Spatial Sensations in Vision

Table 16.7 shows the temporal and spatial sensations in vision, which can be described in terms of temporal factors extracted from the temporal autocorrelation function ACF, in the time domain, and in terms of spatial factors extracted from the spatial ACF. The missing fundamental phenomenon of flickering light is perceived as one of the typical temporal sensations in vision and is predicted by the value of  $\tau_1$ . The pitch frequency  $F$  of the temporally complex pattern of flickering light is given by  $F=1/\tau_1$ , and the strength of this flicker pitch is given by  $\phi_1$  in a manner similar to auditory pitch. As the spatial sensations of two-dimensional space, contrast, regularity and coarseness can similarly be described in terms of spatial autocorrelation features, as expressed by Equations (13.3), (13.4) and (13.5), respectively.

**Table 16.7** Temporal and spatial sensations in vision in relation to factors extracted from the temporal ACF and spatial ACF, respectively

		Temporal sensations	Spatial sensations		
Temporal or spatial ACF	Factors	Missing fundamental of flickering light (Section 13.1)	Contrast (Section 13.2)	Coarseness (Section 13.2)	Regularity (Section 13.2)
Temporal ACF	$\tau_1$	$\mathbf{X}^1$ ( $F = 1/\tau_1$ )			
	$\phi_1$	$\mathbf{X}$ (Strength of the pitch)			
	$\tau_e$	$\mathbf{X}^2$			
Spatial ACF	$\Phi(0)$		$c_1 \Phi(0)$	$c_3 \Phi(0)$	
	$\delta_1$		$c_2 \delta_1$		
	$\phi_1$			$c_4 \delta_1$	$c_6 \delta_1$
	$\delta_e$				$c_5 \phi_1$ —

<sup>1</sup>  $\mathbf{X}$ : Major factors associated with the sensation.

<sup>2</sup>  $\mathbf{X}$ : Note that factors of  $\phi_1$  and  $\tau_e$  are mutually related.

### 16.2.2 Visual Subjective Preferences

Table 16.8 lists the most preferred-temporal conditions for flickering light, oscillating single targets, and fluttering leaves in the wind. The preferred sinusoidal period of a flickering light and oscillating movement of single target both in the vertical and horizontal directions are almost the same, about 1.0 s. A still more preferred condition is obtained by introducing a fluctuation to this 1 Hz periodicity, an example being the twinkling of stars. The most preferred fluctuation level of flickering light is approximately expressed by  $\phi_1 \approx 0.46$ . This signifies that the extreme conditions  $\phi_1 = 0$  (perfectly random) and  $[\phi_1]_p = 1.0$  (perfectly periodic like the sinusoidal) are not preferred, but if a certain degree of fluctuation is introduced, then it becomes more preferred. Concerning the fluttering leaves, the most preferred condition has been apparently observed at  $[\tau_e]_p \approx 0.3-0.4$ , but this bears further investigation.

**Table 16.8** Preferred conditions of visual temporal factors for vision with movement

Temporal	Factors	Flickering light (Section 14.1)	Moving single target (Section 14.2)	Moving leaves (Section 16.1.1)
ACF	$\tau_1$	$[\tau_1]_p \approx 1.0$ s	$[\tau_1]_p \approx 1.0$ s (vertical) $[\tau_1]_p \approx 1.0$ s (horizontal)	–
	$\phi_1$	$[\phi_1]_p \approx 0.46$	–	–
	$\tau_e$	– <sup>1</sup>	–	$[\tau_e]_p \approx 0.3$ to $0.4$ s

<sup>1</sup> Factors to be examined whether or not they influence the respective attribute.

Table 16.9 indicates the most preferred-spatial regularity condition for texture, that is  $[\phi_1]_p \approx 0.41$ . This fluctuation is, in fact, very similar to the optimal condition for temporal fluctuations.

**Table 16.9** Preferred condition of the visual spatial factor extracted from the spatial ACF for texture

Spatial	Factors	Preferred condition (Section 14.3)
ACF	$\delta_1$	– <sup>1</sup>
	$\phi_1$	$[\phi_1]_p \approx 0.41$
	$\delta_e$	–
	$\Phi(0)_S$	–

<sup>1</sup> Factors to be examined whether or not they influence the respective attribute.

Table 16.10 shows brain activities related to subjective preferences that are related to changes of temporal factors for visual fields. In the preferred conditions, as in their auditory counterparts, satisfaction of visual preferences are accompanied by increased persistence and cortical extent of alpha rhythm activity.

**Table 16.10** Left hemisphere specialization observed in EEG and MEG alpha waves in relation to the temporal factors of the visual field

Temporal factor changed	EEG, ACF $\tau_e$ value of alpha wave (Sections 15.1.1 and 15.3)	EEG, CCF $ \phi(\tau) _{\max}$ of alpha wave (Section 15.1.2)	MEG, ACF $\tau_e$ value of alpha wave (Section 15.2)
Period of the flickering light, T	L > R <sup>1</sup> (SW) <sup>2</sup>	L & R <sup>3</sup> (SW)	L > R (SW)
Period of horizontal movement of target, T	L > R (SW)	–	–

<sup>1</sup> The ratio of  $\tau_e$  values of alpha wave in EEG, high to low preference, increased significantly in the left hemisphere.

<sup>2</sup> Sinusoidal wave used to control the period, T.

<sup>3</sup> The  $\phi(\tau)|_{\max}$  value of alpha wave in MEG increased on the wide area of both hemispheres when the scale value of subjective preference was high. The similar repetitive feature in the alpha wave on the wide area of the brain relates to the preferred condition of vision.

Because there are deep commonalities between the neurons and neural systems that subserve all the senses, we believe that this present theory of auditory and visual sensations can be extended and applied to other modalities as well. It could well be the case that in other senses there are analogous temporal and spatial factors that are extracted from their respective neural correlation functions. A possible fruitful direction for further investigations is to study the interactions and/or interferences between different senses, which might also be explained in terms of temporal factors associated with the left hemisphere and/or spatial factors with the right hemisphere.

# References

- Aitkin, L. M., Irvine, D. R. F., and Webster, W. R. (1984) Central neural mechanisms of hearing. In: Brookhart, J.M., Mountcastle, V.B. (eds.), *Handbook of Physiology*, pp 675–737. American Physiological Society, Bethesda, MD.
- Aitkin, L. (1986) *The Auditory Midbrain: Structure and Function in the Central Auditory Pathway*. Humana Press, Clifton, NJ.
- Aitkin, L. (1990) *The Auditory Cortex*. Chapman and Hall, New York.
- Alho, K., Connolly, J. F., Cheour, M., Lehtokoski, A., Huotilainen, M., Virtanen, J., Aulanko, R., and Ilmoniemi, R. J. (1998). Hemispheric lateralization in preattentive processing of speech sounds. *Neuroscience Letters*, **258**, 9–12.
- Alrutz, H. (1981). Ein neuer Algorithms zur Auswertung von Messungen mit Pseudorausschsignalen. *Fortschritte der Akustik, DAGA '81*, Berlin, 525–528.
- Amandasun, M., and King, R. (1989). Textural features corresponding to textural properties. *IEEE Transactions Systems, Man and Cybernetics*, **19**, 1264–1274.
- 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., Nakane, Y., and Egawa, J. (1975). Effects of aircraft noise on the mental work of pupils. *Journal of Sound and Vibration*, **43**, 683–691.
- 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 Imamura, M. (1979). Subjective preference tests for sound fields in concert halls simulated by the aid of a computer. *Journal of Sound and Vibration*, **65**, 229–239.
- Ando, D., and Gottlob, D. (1979). Effects of early multiple reflection on subjective preference judgments on music sound fields. *Journal of the Acoustical Society of America*, **65**, 524–527.
- Ando, Y., and Morioka, K. (1981). Effects of the listening level and the magnitude of the interaural crosscorrelation (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., and Hosaka, I. (1983). Hemispheric difference in evoked potentials to spatial sound field stimuli. *Journal of the Acoustical Society of America*, **74**(S1), S64–S65(A).

- 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., and Kang, S. H. (1987). A study on the differential effects of sound stimuli on performing left- and right-hemispheric task. *Acustica*, **64**, 110–116.
- Ando, Y., Kang, S. H., and Nagamatsu, H. (1987a). 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. (1987b). On the relationship between auditory-evoked potential and subjective preference for sound field. *Journal of the 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. *Journal 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., Nagamatsu, H., and Kang, S. H. (1991). Auditory brainstem response (ABR) in relation to the horizontal angle of sound incidence. *Acoustic Letters*, **15**, 57–64.
- Ando, Y. (1992). Evoked potentials relating to the subjective preference of sound fields. *Acustica*, **76**, 292–296.
- Ando, Y., and Chen, C. (1996). On the analysis of the autocorrelation function of  $\alpha$ -waves on the left and right cerebral hemispheres in relation to the delay time of single sound reflection. *Journal of Architectural Planning and Environmental Engineering, Architectural Institute of Japan (AIJ)*, **488**, 67–73.
- 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., 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 Noson, D. (Eds.). (1997). *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., Sato, S., and Sakai, H. (1999). Fundamental subjective attributes of sound fields based on the model of auditory-brain system. In: Sendra, J. J. (ed.), *Computational Acoustics in Architecture*. WIT Press, Southampton.
- Ando, Y., Sakai, H., and Sato, S. (2000). Formulae describing subjective attributes for sound fields based on a model of the auditory-brain system. *Journal of Sound and Vibration*, **232**, 101–127.
- Ando, Y. (2001a). A theory of primary sensations measuring environmental noise. *Journal of Sound and Vibration*, **241**, 3–18.
- Ando, Y. (2001b). Differential effects of noise and music signals on the behavior of children. *Journal of Sound and Vibration*, **241**, 129–140.
- Ando, Y. (2002). Correlation factors describing primary and spatial sensations of sound fields. *Journal of Sound and Vibration*, **258**, 405–417.

- Ando, Y., and Pompoli, R. (2002). Factors to be measured of environmental noise and its subjective responses based on the model of the auditory-brain system. *Journal of Temporal Design in Architecture and the Environment*, **2**, 2–12. Available at <http://www.jtdweb.org/journal/>.
- 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. Available at <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. Available at <http://www.jtdweb.org/journal/>.
- Ando, Y. (2006). Neural Based Theory of Temporal and Spatial Sensations of Sound. The 9th Western Pacific Acoustic Conference (WESPAC IX 2006), Seoul, 2006.
- Ando, Y. (2007a). Musical performance and the concert hall as a second instrument. *Journal of Temporal Design in Architecture and the Environment*, **7**, 19–32. Available at <http://www.jtdweb.org/journal/>.
- Ando, Y. (2007b). Concert hall acoustics based on subjective preference theory. In: Rossing, T. (ed.), *Springer Handbook of Acoustics*. Springer-Verlag, New York, Chapter 10.
- Ando, Y. (2008). Theory of auditory temporal and spatial primary sensations. *Journal of Temporal Design in Architecture and the Environment*, **8**, 8–26. Available at <http://www.jtdweb.org/journal/>.
- Ando, Y. (2009). Theory of temporal and spatial design of environment. In: Blumel, D. (ed.), *McGraw-Hill 2009 Yearbook of Science and Technology*, pp. 384–389. McGraw-Hill, New York.
- Badcock, D. R., and Derrington, A. M. (1989). Detecting the displacement of spatial beats: No role for distortion products. *Vision Research*, **29**, 731–739.
- Ball, K., and Sekuler, R. (1979). Masking of motion by broad-band and filtered directional noise. *Perception & Psychophysics*, **26**, 206–214.
- Barlow, J. S. (1961). Autocorrelation and crosscorrelation technique in EEG analysis. *Electroencephalography and Clinical Neurophysiology*, **20**, 31–36.
- Békésy, G. (1934). Ueber die Hoersamkeit von Konzert und Rundfunksaelen. *Elektrische Nachrichten Technik*, **11**, 369–375.
- Békésy, G. (1967). *Sensory Inhibition*. Princeton University Press, Princeton, NJ.
- Ben-Av, M. B., and Sagi, D. (1995). Perceptual grouping by similarity and proximity: Experimental results can be predicted by intensity autocorrelations. *Vision Research*, **35**, 853–866.
- Beranek, L. L. (1957). Revised criteria for noise in buildings, *Noise Control*, **3**, 19–27.
- Beranek, L. L. (1971). *Noise and Vibration Control*. McGraw-Hill, New York.
- Beranek, L. L., Blazier, W. E., and Firwer, J. J. (1971). Preferred noise criterion (PNC) curves and their application to rooms, *Journal of the Acoustical Society of America*, **50**, 1223–1228.
- Beranek, L. L. (1988). Balanced noise-criterion (NCB) curves. *Journal of the Acoustical Society of America*, **86**, 650–664.
- Berglund, B., Berglund, U., and Lindvall, T. (1975). Scaling loudness, noisiness, and annoyance of aircraft noise, *Journal of the Acoustical Society of America* **57**, 930–934.
- Boring, E. G. (1942). *Sensation and Perception in the History of Experimental Psychology*. Appleton-Century-Crofts, New York.
- Botte, M. C., Bujas, Z., and Chocholle, R. (1975). Comparison between the growth of averaged electroencephalic response and direct loudness estimations. *Journal of the Acoustical Society of America*, **58**, 208–213.
- Bowen, R. W., Pokorny, J., and Smith, V. C. (1989). Sawtooth contrast sensitivity: Decrements have the edge. *Vision Research*, **29**, 1501–1509.
- Bowen, R. W., Pokorny, J., Smith, V. C., and Fowler, M. A. (1992). Sawtooth contrast sensitivity: Effects of mean illuminance and low temporal frequencies. *Vision Research*, **32**, 1239–1247.
- Braizer, M. A. B., and Casby, J. U. (1952). crosscorrelation and autocorrelation studies of electroencephalographic potentials. *Electroencephalography and Clinical Neurophysiology*, **4**, 201–211.



- Brodatz, P. (1966). *Textures*. Dover, New York.
- 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). Nachhallfreier Musik fuer akustische Modelluntersuchungen. *Rundfunktechn. Mitteilungen*, **13**, 200–201.
- Campbell, F. W., and Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology (London)*, **197**, 551–566.
- Cariani, P. A., and Delgutte, B. (1996a). Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *Journal of Neurophysiology*, **76**, 1698–1716.
- Cariani, P. A., and Delgutte, B. (1996b). Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase-invariance, pitch circularity, and the dominance region for pitch. *Journal of Neurophysiology*, **76**, 1717–1734.
- Cariani P. (1999). Temporal coding of periodicity pitch in the auditory system: An overview. *Neural Plasticity*, **6**, 147–172.
- Cariani, P. (2001). Temporal coding of sensory information in the brain. *Acoustical Science and Technology*, **22**, 77–84.
- Cariani P. (2002). Temporal codes, timing nets, and music perception. *Journal of New Music Research*, **30**, 107–136.
- Cariani P. (2004). A temporal model for pitch multiplicity and tonal consonance. *Proceedings, International Conference on Music Perception & Cognition (ICMPC)*, Evanston, IL.
- Chandler, D. W., and Grantham, D. W. (1992). Minimum audible movement angle in the horizontal plane as a function of stimulus frequency and bandwidth, source azimuth, and velocity. *Journal of the Acoustical Society of America*, **91**, 1624–1636.
- Chen, C., and Ando, Y. (1996). On the relationship between the autocorrelation function of the  $\alpha$ -waves on the left and right cerebral hemispheres and subjective preference for the reverberation time of music sound field. *Journal of the Architectural Planning and Environmental Engineering, Architectural Institute of Japan (AIJ)*, **489**, 73–80.
- Chen, C., Ryugo, H., and Ando, Y. (1997). Relationship between subjective preference and the autocorrelation function of left and right cortical  $\alpha$ -waves responding to the noise-burst tempo. *Journal of the Architectural Planning and Environmental Engineering, Architectural Institute of Japan (AIJ)*, **497**, 67–74.
- 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.
- Cherry C. (1961). Two ears – but one world. In: Rosenblith W. A. (ed.), *Sensory Communication*, pp. 99–117. MIT Press/John Wiley, New York.
- Cheveigne, A. (1998). Cancellation model of pitch perception. *Journal of the Acoustical Society of America*, **103**, 1261–1271.
- Cho, R. Y., Yang, V., and Hallett, P. E. (2000). Reliability and dimensionality of judgments of visually textured materials. *Perception and Psychophysics*, **62**, 735–752.
- Chon, R. (1970). Differential cerebral processing of noise and verbal stimuli. *Science*, **172**, 599–601.
- Clottes, J., and Courtin, J. (1995). *Grotte Cosquer bei Marseille*, Thorbecke Verlag, Speläo.
- Cocchi, A., Farina, A., and Rocco, L. (1990). Reliability of scale-model research: A concert hall case. *Applied Acoustics*, **30**, 1–13.
- Colburn, S. (1996). Computational models of binaural processing. In: Hawkins, H., McMullin, T., Popper, A. N., and Fay, R. R. (eds.), *Auditory Computation*. Springer-Verlag, New York.
- Cross, G. R., and Jain, A. K. (1983). Markov random field texture models. *IEEE Transactions Pattern Analysis and Machine Intelligence*, **5**, 25–39.

- Dallos, P., Popper, A. N., and Fay, R. R. (1996) *The Cochlea*. Springer, New York.
- Damaske, P. (1967/68). Subjektive Untersuchungen von Schallfeldern. *Acustica*, **19**, 199–213.
- Damaske, P., and Ando, Y. (1972). Interaural crosscorrelation for multichannel loudspeaker reproduction. *Acustica*, **27**, 232–238.
- Davis, A. E., and Wada, J. A. (1974). Hemispheric asymmetry: Frequency analysis of visual and auditory evoked responses to non-verbal stimuli. *Electroencephalography and Clinical Neurophysiology*, **37**, 1–9.
- de Boer, E. (1976). On the "residue" and auditory pitch perception. In: Keidel, W. D., and Neff, W. D. (eds.), *Auditory System (Handbook of Sensory Physiology)*, pp. 479–583. Springer-Verlag, Berlin.
- de Cheveigné, A. (2004). Pitch perception models. In: Plack, C. J. and Oxenham, A. J. (eds.), *Pitch*. Springer-Verlag, New York.
- de Lange, H. (1952). Experiments on flicker and some calculations on an electrical analogue of the foveal system. *Physica*, **8**, 935–950.
- de Valois, R. L., and de Valois, K. K. (1980). Spatial vision. *Annual Review of Psychology*, **31**, 309–341.
- Doi, S., Otuka, T., and Takahashi, H. (1997). Experimental investigation on lighting control with 1/f fluctuation, *IEEJ Transactions on Electronics, Information and Systems*, **117C**, 409–415 (in Japanese).
- Dong, D. W., and Atick, J. J. (1995). Statistics of natural time-varying images. *Network*, **5**, 517–548.
- Dubrovskii, N. A., and Chernyak, R. I. (1969). Binaural loudness summation under varying degrees of noise correlation. *Soviet Physics – Acoustics*, **14**, 326–332.
- Edmonds, B. A., and Culling, J. F. (2009). Interaural correlation and the binaural summation of loudness. *Journal of Acoustical Society of America*, **125**, 3865–3870.
- Edward, R. M. (1974). A subjective assessment of concert hall acoustics. *Acustica*, **30**, 183–195.
- Ehret, G., and Romand, R. (1997) *The Central Auditory System*. Oxford University Press, New York.
- Eisner, A. (1995). Suppression of flicker response with increasing test illuminance: roles of temporal waveform, modulation depth, and frequency. *Journal of the Optical Society of America, A*, **12**, 214–224.
- Eulitz, C., Diesch, E., Pantev, C., Hampson, S., and Elbert, T. (1995). Magnetic and electric brain activity evoked by the processing of tone and vowel stimuli. *The Journal of Neuroscience*, **15**, 2748–2755.
- Eulitz, C., Eulitz, H., and Elbert, T. (1997). Differential outcomes from magneto- and electroencephalography for the analysis of human cognition. *Neuroscience Letters*, **227**, 185–188.
- Field, D. I. (1987). Relation between the statistics of natural images and the response properties of cortical cells. *Journal of Optical Society of America*, **4**, 2379–2394.
- Fraisse, P. (1984). Perception and estimation of time. *Annual Review of Psychology*, **35**, 1–36.
- Francois, J. M., Meiri, A. Z., and Porat, B. (1993). Unified texture model based on a 2-D wold-like decomposition. *IEEE Transactions on Signal Processing*, **41**, 2665–2678.
- Fujii, K., Kita, S., Matsushima, T., and Ando, Y. (2000). The missing fundamental phenomenon in temporal vision. *Psychological Research*, **64**, 149–154.
- Fujii, K., Soeta, Y., and Ando, Y. (2001). Acoustical properties of aircraft noise measured by temporal and spatial factors. *Journal of Sound and Vibration*, **241**, 69–78.
- 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. Available at <http://www.jtdweb.org/journal/>.
- Fujii, K., Sugi, S., and Ando, Y. (2003). Textural properties corresponding to visual perception based on the correlation mechanism in the visual system. *Psychological Research*, **67**, 197–208.
- Fujii, K., Sakurai, M., and Ando, Y. (2004a). Computer software for identification of noise source and automatic noise measurement. *Journal of Sound and Vibration*, **277**, 573–582.

- Fujii, K., Hotehama, T., Kato, K., Shimokura, R., Okamoto, Y., Suzumura, Y., and Ando, Y. (2004b). 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. Available at <http://www.jtdweb.org/journal/>.
- Fujii, K., and Ando, Y. Relationship between the visual properties of texture and subjective preference. Unpublished manuscript.
- Gabriel, K. J., Colburn, H. S. (1981). Interaural correlation discrimination: I. Bandwidth and level dependence. *Journal of the Acoustical Society of America*, **69**, 1394–1401.
- Gade, A. C. (1989). Investigations of musicians' room acoustic conditions in concert halls. Part I: Methods and laboratory experiments. *Acustica*, **69**, 193–203.
- Galín, D., and Ellis, R.R. (1975). Asymmetry in evoked potentials as an index of lateralized cognitive processes: Relation to EEG alpha asymmetry. *Neuropsychologia*, **13**, 45–50.
- 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.
- Grantham, D. W. (1986). Detection and discrimination of simulated motion of auditory targets in the horizontal plane. *Journal of the Acoustical Society of America*, **79**, 1939–1949.
- Greenwood, D. D. (1961a). Auditory masking and critical band. *Journal of the Acoustical Society of America*, **33**, 484–502.
- Greenwood, D. D. (1961b). Critical bandwidth and the frequency of the basilar membrane. *Journal of the Acoustical Society of America*, **33**, 1344–1356.
- Gros, B. L., Blake, R., and Hiris, E. (1998). Anisotropies in visual motion perception: A flesh look. *Journal of the Optical Society of America, A*, **15**, 2003–2011.
- Gullikson, H. (1956). A least square solution for paired comparisons with incomplete data. *Psychometrika*, **21**, 125–134.
- Haas, H. (1951). Ueber den Einfluss eines Einfachechos auf die Hoersamkeit von Sprache. *Acustica*, **1**, 49–58.
- Hammett, S. T., and Smith, A. T. (1994). Temporal beats in the human visual system. *Vision Research*, **34**, 2833–2840.
- Hanada, K., Kawai, K., and Ando, Y. (2007). A study of the timbre of an electric guitar sound with distortion. *Journal of South China University of Technology (Natural Science Edition), Proceedings of the 3rd International Symposium on Temporal Design, Guangzhou, November 2007*, pp. 96–99.
- Haralick, R. M. (1979). Statistical and structural approaches to textures. *Proceedings of IEEE*, **67**, 786–804.
- Hargest, T. J., and Pinker, R. A. (1967). The influence of added narrow band noises and tones on the subjective response to shaped white noise. *Journal of the Royal Aeronautical Society*, **71**, 428–430.
- Hase, S., Takatsu, A., Sato, S., Sakai, H., and Ando, Y. (2000). Reverberance of an existing hall in relation to subsequent reverberation time and SPL. *Journal of Sound and Vibration*, **232**, 149–155.
- Hawkins, H. L. (1996) *Auditory Computation*. Springer-Verlag, New York.
- Hecox, K., and Galambos, R. (1974). brainstem auditory evoked responses in human infants and adults. *Archives Otolaryngology*, **99**, 30–33.
- Heeger, D., and Bergen, J. (1995). Pyramid-based analysis/synthesis. *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques*, 229–238.
- Hellman, R. P. (1982). Loudness, annoyance and noisiness produced by single-tone-noise complexes. *Journal of the Acoustical Society of America*, **72**, 62–73.
- Hellman, R. P. (1984). Growth rate of loudness, annoyance and noisiness as a function of tone location within the noise spectrum, *Journal of the Acoustical Society of America*, **75**, 209–218.
- Henning, G. B., Herz, B. G., and Broadbent, D. E. (1975). Some experiments bearing on the hypothesis that the visual system analyzes spatial patterns in independent bands of spatial frequency. *Vision Research*, **15**, 887–897.

- Hidaka, T., Beranek, L. L., and Okano, T. (1995). Interaural crosscorrelation, lateral fraction, and low- and high-frequency sound levels as measures of acoustical quality in concert halls. *Journal of the Acoustical Society of America*, **98**, 988–1007.
- Hinrichs, H., and Machleidt, W. (1992). Basic emotions reflected in EEG-coherences. *International Journal of Psychophysiology*, **13**, 225–232.
- Hirahara, T., Cariani, P., and Delgutte, B. (1996) Representation of low-frequency vowel formants in the auditory nerve. In: *Proceedings, ESCA Research Workshop on the Auditory Basis of Speech Perception*, Keele University, United Kingdom, July 15–19, 1996.
- Hiramatsu, K., Takagi, K., and Yamamoto, T. (1983). The effect of sound duration on annoyance. *Journal of Acoustical Society of Japan*, **32**, 739–750 (with English abstract).
- Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*. The University of Michigan Press, Ann Arbor, MI.
- Hoovey, Z. B., Heinmann, U., and Creutzfeldt, O. D. (1972). Inter-hemispheric “synchrony” of alpha waves. *Electroencephalogram. Electroencephalography and Clinical Neurophysiology*, **32**, 337–347.
- Hotehama, T., Sato, S., and Ando, Y. (2002). Dissimilarity judgments in relation to temporal and spatial factors for the sound fields in an existing hall. *Journal of Sound and Vibration*, **258**, 429–441.
- Houtgast, T., Steeneken, H. J. M., and Plomp, R. (1980). Predicting speech intelligibility in rooms from the modulation transfer function. I. General room acoustics. *Acustica*, **46**, 60–72.
- Inagaki, T., Iizuka, K., Agu, M., Akabane, H., and Abe, N. (2001).  $1/f^n$  fluctuating phenomena in luminous pattern of firefly and its healing effect. *Transactions of the Japan Society of Mechanical Engineers*, **67C**, 365–372 (in Japanese).
- 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.
- Inoye, T., Shinosaki, K., and Yagasaki, A. (1983). The direction of spread of alpha activities over the scalp. *Electroencephalography and Clinical Neurophysiology*, **55**, 290–300.
- Irvine, D. R. F. (1986) *The Auditory Brainstem*. Springer-Verlag, Berlin.
- Jasper, H. H. (1958). The ten-twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology*, **10**, 370–375.
- 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.
- Julesz, B. (1962). Visual pattern discrimination. *IRE Transaction on Information Theory*, **8**, 84–92.
- 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.
- Kaplan, S. (1987). Aesthetics, affect and cognition: environmental preference from an evolutionary perspective. *Environmental Behavior*, **19**, 3–32.
- 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.
- Kato, K., Fujii, K., Hirawa, T., Kawai, K., Yano, T., and Ando, Y. (2006). Investigation of the relation between  $(\tau_e)_{\min}$  and operatic singing with different vibrato styles. *Journal of Temporal Design in Architecture and the Environment*, **6**, 35–48.
- Kato, K., Fujii, K., Hirawa, T., Kawai, K., Yano, T., and Ando, Y. (2007a). Investigation of the relation between minimum effective duration of running autocorrelation function and operatic singing with different interpretation styles. *Acta Acustica Acustica*, **93**, 421–434.

- Kato, K., Tanaka, S., Kawai, K., Yano, T., and Ando, Y. (2007b). Preferred delay time of the single reflection for Japanese traditional "Shigin" singing in relation to the autocorrelation function. *Journal of South China University of Technology (Natural Science Edition)*, Proceedings of the 3rd International Symposium on Temporal Design, Guangzhou, November 2007, pp. 92–95.
- 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. Proceedings 6th International Congress on Acoustics, Tokyo, Paper E-2-4.
- Kelly, D. H. (1961). Flicker fusion and harmonic analysis. *Journal of the Optical Society of America*, **51**, 917–918.
- Kiang, N. Y.-S., Watanabe, T., Thomas, E. C. and Clark, L. F. (1965). *Discharge Pattern of Single Fibers in the Cat's Auditory Nerve*. MIT Press, Cambridge, MA.
- Kimura, D. (1973). The asymmetry of the human brain. *Scientific American*, **228**, 70–78.
- Kinchla, R. A. and Allan, L. G. (1970). Visual movement perception: A comparison of sensitivity to vertical and horizontal movement. *Perception & Psychophysics*, **8**, 399–405.
- Kitamura, T., Shimokura, R., Sato, S., and Ando, Y. (2002). Measurement of temporal and spatial factors of a flushing toilet noise in a downstairs bedroom. *Journal of Temporal Design in Architecture and the Environment*, **2**, 13–19. Available at <http://www.jtdweb.org/>.
- Klumpp, R. G., and Eady, H. R. (1956). Some measurements of interaural time-difference thresholds. *Journal of the Acoustical Society of America*, **28**, 859–860.
- Korenaga, Y. (1997). A new method of calculating speech intelligibility with respect to the delay time of reflections. In: Ando, Y., and Noson, D. (eds.), *Conference Proceedings of MCHA 1995*, Academic Press, London, Chapter 28.
- Kremers, J., Lee, B. B., Pokorny, J., and Smith, V. C. (1993). Response to macaque ganglion cells and human observers to compound periodic waveforms. *Vision Research*, **33**, 1997–2011.
- Kryter, K. D., and Pearsons, K. S. (1965). Judged noisiness of a band of random noise containing an audible pure tone. *Journal of the Acoustical Society of America*, **38**, 106–112.
- Kuttruff, H. (1991). *Room acoustics*, 3rd ed. Elsevier Applied Science, London.
- Kuwano, S., Namba, S., and Fastl, H. (1988). On the judgment of loudness, noisiness and annoyance with actual and artificial noise. *Journal of Sound and Vibration* **127**, 457–465.
- Lauterborn, W. (2003). Personal communication.
- 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.
- Levinson, E., and Sekuler, R. (1980). A two-dimensional analysis of direction-specific adaptation. *Vision Research*, **20**, 103–108.
- Levy, J., and Trevarthen, C. (1976). Metacontrol of hemispheric function in human split-brain patients. *Journal of Experimental Psychology: Human Perception and Performance*, **2**, 299–312.
- Licklider, J. C. R. (1951). A duplex theory of pitch perception, *Experientia*, **VII**, 128–134.
- Licklider, J. C. R. (1959). Three auditory theories. In: Koch, S. (ed.), *Psychology: A Study of a Science*. Study I. Conceptual and Systematic, pp. 41–144. McGraw-Hill, New York.
- Lindsey, D. B. (1952). Psychological phenomena and the electroencephalogram. *Electroencephalography and Clinical Neurophysiology*, **4**, 443–456.
- Liske, E., Hughes, H. M., and Stowe, D. E. (1967). crosscorrelation of human alpha activity: normative data. *Electroencephalogram. Electroencephalography and Clinical Neurophysiology*, **22**, 429–436.
- Liu, F., and Picard, R. W. (1996). Periodicity, directionality, and randomness: World features for image modeling and retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**, 722–733.
- Liu, F. (1997). *Modeling Spatial and Temporal Textures*. PhD Thesis, Massachusetts Institute of Technology.

- Lundeen, C., and Small Jr., A. M. (1984). The influence of temporal cue on the strength of periodicity pitches. *The Journal of the Acoustical Society of America* **75**, 1578–1587.
- Maki, F. (1997). Sound and figure: Concert hall design. In: Ando and Noson (eds.), *Music and Concert Hall Acoustics*, Conference Proceedings from MCHA 1995, Chapter 1.
- Malik, J., and Perona, P. (1990). Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America, A*, **7**, 923–932.
- Mandler, M. B., and Makous, W. (1984). A three channel model of temporal frequency perception. *Vision Research*, **24**, 1881–1887.
- Mao, J., and Jain, A. K. (1992). Texture classification and segmentation using multiresolution simultaneous autoregressive models. *Pattern Recognition*, **25**, 173–188.
- Marshall, A. H., Gottlob, D., and Alrutz, H. (1978). Acoustical conditions preferred for ensemble. *Journal of the Acoustical Society of America*, **64**, 1437–1442.
- 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.
- Mathews, M. V., and Pfafflin, S. M. (1965). Effect of filter type on energy-detection models for auditory signal detection. *Journal of the Acoustical Society of America*, **38**, 1055–1056.
- May, D. N. (1978). Basic subjective responses to noise. In: May, D. N. (ed.), *Handbook of Noise Assessment*. Van Nostrand Reinhold, New York.
- McLachlan, K. R., and Shaw, J. C. (1965). A correlation technique for quantifying the spatial distribution of the electroencephalogram. *Life Sciences*, **4**, 265–270.
- 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.
- Meddis, R., and O'Mard, L. (1997). A unitary model of pitch perception. *Journal of the Acoustical Society of America*, **102**, 1811–1820.
- Mehrgardt, S., and Mellert, V. (1977). Transformation characteristics of the external human ear. *Journal of the Acoustical Society of America*, **61**, 1567–1576.
- Merthayasa, I Gde 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.
- Merthayasa, I Gde N., and Ando, Y. (1996). Variation in the autocorrelation function of narrow band noises: Their effect on loudness judgment. Presented at Japan and Sweden Symposium on Medical Effects of Noise.
- Molino, J. A. (1979). Annoyance and noise. In: Harris, C. M. (ed.), *Handbook of Noise Control*. McGraw-Hill, New York, Chapter 16.
- Moore, B. C. J. (1982). *Introduction to the Psychology of Hearing*, 2nd. ed. Academic, London.
- Moore, B. C. J. (2003). *An Introduction to the Psychology of Hearing*, 5th ed. Academic Press, San Diego.
- Morimoto, M., and Maekawa, Z. (1988). Effects of low frequency components on auditory spaciousness. *Acustica*, **66**, 190–196.
- 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.

- Mouri, K., Fujii, K., Shimokura, R., and Ando, Y. A study on the dynamic properties of auditory interaural crosscorrelation function relating to the moving sound image in the horizontal plane for the band noise (Unpublished manuscript).
- Nääätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilaonen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R. J., Luuk, A., Allik, J., Sinnkonen, J., and Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic responses. *Nature*, **385**, 432–434.
- Nachmias, J., and Rogowitz, B. E. (1983). Masking by spatially-modulated gratings. *Vision Research*, **23**, 1621–1629.
- Nagamatsu, H., Kasai, H., and Ando, Y. (1989). Relation between auditory evoked potential and subjective estimation – effect of sensation level. Report of Meeting of Acoustical Society of Japan, pp. 391–392 (in Japanese).
- 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.
- Nakajima, T., and Ando, Y. (1997). Calculation and measurement of acoustic factors at each seat in the Kirishima International Concert Hall. In: Ando, Y., and Noson, D. (eds.), *Music and Concert Hall Acoustics*, pp. 39–49. Conference Proceedings from MCHA 1995, Chapter 5.
- Nakayama, I. (1984). Preferred time delay of a single reflection for performers. *Acustica* **54**, 217–221.
- Noson, D., Sato, S., Sakai, H., and 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.
- Oertel, D., Fay, R. R., and Popper, A. N. (2002) *Integrative Functions in the Mammalian Auditory Pathway*. Springer, New York.
- Ohgushi, K. (1980). Physical and psychological factors governing timbre of complex tones. *Journal of Acoustical Society of Japan*, **36**, 253–259.
- Okamoto, Y., Soeta, Y., and Ando, Y. (2003). Analysis of EEG relating to subjective preference of visual motion stimuli. *Journal of Temporal Design in Architecture and the Environment*, **3**, 36–42. Available at <http://www.jtdweb.org/journal/>.
- Okamoto, Y., Nakagawa, S., Yano, T., and Ando, Y. (2007). An MEG study of cortical responses in relation to subjective preference for different regularities of a fluctuating light. *Journal of Temporal Design in Architecture and the Environment*, **7**, 10–18. Available at <http://www.jtdweb.org/journal/>.
- Opitz, B., Mecklinger, A., and Friederici, A.D. (2000). Functional asymmetry of human prefrontal cortex: Encoding and retrieval of verbally and nonverbally coded information. *Learning and Memory*, **7**, 85–95.
- Palmer, A. R. (1992) Segregation of the responses to paired vowels in the auditory nerve of the guinea pig using autocorrelation. In: Schouten M. E. H. (ed.), *The Auditory Processing of Speech* pp. 115–124. Mouton de Gruyter, Berlin.
- Palomaki, K., Tiitinen, H., Makinen, V., May, P., and Alku, P. (2002). Cortical processing of speech sounds and their analogues in a spatial auditory environment. *Cognitive Brain Research*, **14**, 294–299.
- Petsche, H. (1996). Approaches to verbal, visual and musical creativity by EEG coherence analysis. *International Journal of Psychophysiology*, **24**, 145–159.
- Popper, A. N., and Fay, R. R. (1992) *The Mammalian Auditory Pathway: Neurophysiology*. Springer-Verlag, New York.
- Portilla, J., and Simoncelli, E. P. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, **40**, 49–71.
- Pressnitzer, D., Patterson, R. D., and Krumboltz, K. (2001). The lower limit of melodic pitch. *Journal of the Acoustical Society of America*, **109**, 2074–2084.

- Probst, Th., Plendl, H., Paulus, W., Wist, E. R., and Scherg, M. (1993). Identification of the visual motion area (area V5) in the human brain by dipole source analysis. *Experimental Brain Research*, **93**, 345–351.
- Raney, J. P., and Cawthon, J. M. (1979). Aircraft noise. In: Harris C.M. (ed.), *Handbook of Noise Control*. McGraw-Hill, New York, Chapter 34.
- Rao, A. R., and Lohse, G. L. (1996). Towards a texture naming system: Identifying relevant dimensions of texture. *Vision Research*, **36**, 1649–1669.
- Rappelsberger, P., and Petsche, H. (1988). Probability mapping: Power and coherence analysis of cognitive processes. *Brain Topography*, **1**, 46–54.
- Raymond, J. E. (1994). Directional anisotropy of motion sensitivity across the visual field. *Vision Research*, **34**, 1029–1038.
- Ritsma, 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: McFadden, D. (ed.), *Neural Mechanisms of Behavior*, pp. 1–33. Springer-Verlag, New York.
- Runderman, D. L., and Bialek, W. (1994). Statistics of natural images: Scaling in the woods. *Physical Review Letters*, **73**, 814–817.
- Sabine, W. C. (1900). *Reverberation. The American Architect and the Engineering Record*, Prefaced by Beranek, L.L. (1992). *Collected Papers on Acoustics*. Peninsula Publishing, Los Altos, CA, Chapter 1.
- 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. Available at <http://www.jtdweb.org/journal/>.
- Sakai, H., Singh, P. K., and Ando, Y. (1997). Inter-individual differences in subjective preference judgments of sound fields. In: Ando, Y., and Noson, D. (eds.), *Music and Concert Hall Acoustics, Conference Proceedings of MCHA 1995*. Academic Press, London, Chapter 13.
- Sakai, H., Sato, S., Prodi, N., and R. Pompoli, R. (2001). Measurement of regional environmental noise by use of a PC-based system. An application to the noise near Airport “G. Marconi” in Bologna. *Journal of Sound and Vibration*, **241**, 57–68.
- Sakai, H., Hotehama, T., Prodi, N., Pompoli, R., and Ando, Y. (2002). Diagnostic system based on human auditory-brain model for measuring environmental noise – An application to railway noise. *Journal of Sound and Vibration*, **250**, 9–21.
- Sakai, K., and Finkel, L. H. (1995). Characterization of the spatial-frequency spectrum in the perception of shape from texture. *Journal of Optical Society of America*, **12**, 1208–1224.
- Sakurai, M., Sakai, H., and Ando, Y. (2001). A computational software for noise measurement and toward its identification. *Journal of Sound and Vibration*, **241**, 19–28.
- Sato, S., and Ando, Y. (1996). Effects of interaural crosscorrelation function on subjective attributes. *Journal of the Acoustical Society of America*, **100(A)**, 2592.
- Sato, S., Mori, Y., and Ando, Y. (1997). The subjective evaluation of source locations on the stage by listeners. In: Ando, Y., and Noson, D. (eds.), *Music and Concert Hall Acoustics*. Academic Press, London, Chapter 12.
- Sato, S., and Ando, Y. (1999). On the apparent source width (ASW) for bandpass noises related to the IACC and the width of the interaural crosscorrelation function ( $W_{IACC}$ ). *Journal of the Acoustical Society of America*, **105**, 1234.
- 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., Sakai, H., and Ando, Y. (2001). The loudness of “complex noise” in relation to the factors extracted from the autocorrelation function. *Journal of Sound and Vibration*, **241**, 97–103.
- Sato, S., Ando, Y., and Mellert, V. (2001). Cues for localization in the median plane extracted from the autocorrelation function. *Journal of Sound and Vibration*, **241**, 53–56.



- Sato, S., Kitamura, T., and Ando, Y. (2002). Loudness of sharply (2068 dB/Octave) filtered noises in relation to the factors extracted from the autocorrelation function. *Journal of Sound and Vibration*, **250**, 47–52.
- Sato, S., 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., Sakai, H., and Prodi, N. (2002). Subjective preference for sound sources located on the stage and in the orchestra pit of an opera house. *Journal of Sound and Vibration*, **258**, 549–561.
- Sato, S., and Ando, Y. (2002). Apparent source width (ASW) of complex noises in relation to the interaural crosscorrelation function. *Journal of Temporal Design in Architecture and the Environment*, **2**, 29–32. Available at <http://www.jtdweb.org/journal/>.
- Sato, S., Nishio, K., and Ando, Y. (2003). Propagation of alpha waves corresponding to subjective preference from the right hemisphere to the left with change in the IACC of a sound field. *Journal of Temporal Design in Architecture and the Environment*, **3**, 60–69. Available at <http://www.jtdweb.org/journal/>.
- Sato, S., Kitamura, T., and Ando, Y. (2004). Annoyance of noise stimuli in relation to the spatial factors extracted from the interaural crosscorrelation function. *Journal of Sound and Vibration*, **277**, 511–521.
- 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. Available at <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.
- Sayles, M., and Winter, I. M. (2008). Reverberation challenges the temporal representation of the pitch of complex sounds. *Neuron*, **58**, 789–801.
- Schaaf, A. van der, and Hateren, J. H. van. (1996). Modeling of the power spectra of natural images statistics and information. *Vision Research*, **36**, 2759–2770.
- Scharf, B. (1962). Loudness summation and spectrum shape. *Journal of the Acoustical Society of America*, **34**, 228–233.
- Schroeder, M. R., Gottlob, D., and Siebrasse, K. F. (1974). Comparative study of European concert halls: Correlation of subjective preference with geometric and acoustic parameters. *Journal of the Acoustical Society of America*, **65**, 958–963.
- 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.
- Seebeck, A. (1844). Über die Definition des Tones. *Annalen der Physik and Chemie* **63**, 353–368.
- Seraphim, H. P. (1961). Ueber die Wahrnehmbarkeit mehrerer Rueckwuerfe von Sprachshall. *Acustica*, **11**, 80–91.
- Shimokura, R., and Ando, Y. (2004). Multiple missing fundamental phenomenon as a monaural-beat perception. *Proceedings of the 17th International Congress on Acoustics*, **II**, 1739–1742, Kyoto.
- Singh, P. K., Ando, Y., and Kurihara, Y. (1994). Individual subjective diffuseness responses of filtered noise sound fields. *Acustica*, **80**, 471–477.
- Slaney, M., and Lyon, R. F. (1993). On the importance of time – A temporal representation of sound. In: Coke, M., Beet, S., and Crawford, M. (eds.), *Visual Representations of Speech Signals*, pp. 95–118. John Wiley & Sons, New York.
- 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., Okamoto, Y., Nakagawa, S., Tonoike, M., and Ando, Y. (2002). Autocorrelation analyses of MEG alpha waves in relation to subjective preference of a flickering light. *NeuroReport*, **13**, 527–533.

- Soeta, Y., Uchida, Y., and Ando, Y. (2001). Matching a tonal tempo with camphor leaves moving in the wind. *Journal of Temporal Design in Architecture and the Environment*, **1**, 21–26.
- Soeta, Y., Uetani, S., and Ando, Y. (2002a). Relationship between subjective preference and alpha wave activity in relation to temporal frequency and mean luminance of a flickering light. *Journal of the Optical Society of America, A*, **19**, 289–294.
- Soeta, Y., Uetani, S., and Ando, Y. (2002b). Propagation of repetitive alpha waves over the scalp in relation to subjective preferences for a flickering light. *International Journal of Psychophysiology*, **46**, 41–52.
- Soeta, Y., Nakagawa, S., Tonoike, M., and Ando, Y. (2002c). 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. Available at <http://www.jtdweb.org/journal/>.
- Soeta, Y., Ohtori, K., and Ando, Y. (2003). Subjective preference for movements of a visual circular stimulus: A case of sinusoidal movement in vertical and horizontal directions. *Journal of Temporal Design in Architecture and the Environment*, **3**, 70–76. Available at <http://www.jtdweb.org/journal/>.
- Soeta, Y., Hotehama, T., Nakagawa, S., Tonoike, M., and Ando, Y. (2004). Auditory evoked magnetic fields in relation to interaural crosscorrelation of band-pass noise. *Hearing Research*, **196**, 109–114.
- 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., Maruo, T., and Ando, Y. (2004). Annoyance of bandpass filtered noises in relation to the factor extracted from autocorrelation function. *Journal of Acoustical Society of America*, **116**, 3275–3278.
- Soeta, Y., Mizuma, K., Okamoto, Y., and Ando, Y. (2005). Effects of the degree of fluctuation on subjective preference for a 1 Hz flickering light. *Perception*, **34**, 587–593.
- Soeta, Y., Nakagawa, S., and Tonoike, M. (2005). Magnetoencephalographic activities related to the magnitude of the interaural crosscorrelation function (IACC) of sound fields. *Journal of Temporal Design in Architecture and the Environment*, **5**, 5–11. Available at <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.
- Sperry, R. W. (1974). Lateral specialization in the surgically separated hemispheres. In: Schmitt, F. O., and Worden, F. C. (eds.), *The Neurosciences: Third Study Program*. MIT Press, Cambridge, MA, Chapter 1.
- Sumioka, T., and Ando, Y. (1996). On the pitch identification of the complex tone by the autocorrelation function (ACF) model. *Journal of the Acoustical Society of America*, **100(A)**, 2720.
- Suzumura, Y., Sakurai, M., Ando, Y., Yamamoto, I., 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. In: Ando, Y., and Noson, D. (eds.), *Music and Concert Hall Acoustics*. Conference Proceedings of MCHA 1995. Academic Press, London, Chapter 23.
- Tamura, H., Mori, S., and Yamawaki, T. (1978). Textural features corresponding to visual perception. *IEEE Transactions Systems, Man and Cybernetics*, **SMC-8**, 460–472.

- Terhardt, E. (1974). Pitch, consonance, and harmony. *The Journal of the Acoustical Society of America*, **55**, 1061–1069.
- Thompson, A. M., and Thompson, G. C. (1988). Neural connections identified with PHA-L anterograde and HRP retrograde tract-tracing techniques. *Journal of Neuroscience Methods*, **25**, 13–17.
- 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.
- Tramo, M. J., Cariani, P. A., Delgutte, B., and Braidá, L. D. (2001) Neurobiology of harmony perception. In: Peretz, I., and Zatorre, R. J. (eds.), *The Cognitive Neuroscience of Music*, pp. 127–151 (Reprint of Tramo et al (2001), *Ann NY Acad Sci*, 2990:2092–2116). Oxford University Press, New York.
- Troland L. T. (1929). *The Principles of Psychophysiology*, Vols. I-III. D. Van Nostrand, New York.
- Tsutsumi, T. (2006). The relationship between music and the concert hall (including the cello performance). *Journal of Temporal Design in Architecture and the Environment*, **6**, 78–81.
- Turner, M. R. (1986). Texture discrimination by Gabor functions. *Biological Cybernetics*, **55**, 71–82.
- Uttal, W. R. (1975). *An Autocorrelation Theory of Form Detection*. Erlbaum, Hillsdale, NJ.
- van de Grind, W. A., Koenderink, J. J., van Doorn, A. J., Milders, M. V., and Voerman, H. (1993). Inhomogeneity and anisotropies for motion detection in the monocular visual field of human observers. *Vision Research*, **33**, 1089–1107.
- van Noorden, L. (1982). Two channel pitch perception. In: Clynes, M. (ed.), *Music, Mind and Brain*, pp. 251–269. Plenum, New York.
- Voss, R. F., and Clarke, J. (1978a). “1/f noise” in music and speech. *Nature*, **258**, 317–318.
- Voss, R. F., and Clarke, J. (1978b). “1/f noise” in music: Music from 1/f noise. *Journal of the Acoustical Society of America*, **63**, 258–263.
- Webster, D. B., Popper, A. N., and Fay, R. R. (1992) *The Mammalian Auditory Pathway: Neuroanatomy*. Springer-Verlag, New York.
- 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.
- Wong, P. C. (2002). Hemispheric specialization of linguistic pitch patterns. *Brain Research Bulletin*, **59**, 83–95.
- Wu, S., Burns, S. A., Reeves, A., and Elsner, A. E. (1996). Flicker brightness enhancement and visual nonlinearity. *Vision Research*, **36**, 1573–1583.
- Yamaguchi, K. (1972). Multivariate analysis of subjective and physical measurements of hall acoustics. *Journal of the Acoustical Society of America*, **52**, 1271–1279.
- 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., and Gourevitch, G. (1987) *Directional Hearing*. Springer-Verlag, New York.
- 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.
- Yost, W. A. (2000). *Fundamentals of Hearing: An Introduction*. Academic Press, San Diego.
- Yrttiäho, S., Tiitien, H., May, P. J. C., and Leino, S. (2008). Cortical sensitivity to periodicity of speech sounds. *Journal of the Acoustical Society of America*, **123**, 2191–2199.
- Zatorre, R., and Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex*, **11**, 946–953.

- Zeki, S., Watson, J. D. G., Lueck, C. J., Friston, K. J., Kennard, C., and Frackowiak, R. S. J. (1991). A direct demonstration of functional specialization in human visual cortex. *Journal of Neuroscience*, **11**, 641–649.
- Zhu, S. C., Wu, Y. N., and Mumford, D. (1998). Filters, random fields and maximum entropy (FRAME) – Towards a unified theory for texture modeling. *International Journal of Computer Vision*, **27**, 1–20.
- 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 Scharf, B. (1965). A model of loudness summation. *Psychological Review*, **72**, 3–26.
- Zwicker, E., and Terhardt, E. (1980). Analytical expressions for critical-band rate and critical bandwidth as a function of frequency. *Journal of the Acoustical Society of America*, **68**, 1523–1525.
- Zwicker, E., and Fastl, H. (1999). *Psychoacoustics*. Springer-Verlag, New York.

# Glossary of Symbols

The number in parentheses indicates the equation in which the symbol is used.

**A** Total amplitude of reflections (3.2)

**A<sub>1</sub>** Pressure amplitude of the single reflection (3.3) or the first reflection

**A<sub>n</sub>** Pressure amplitude of n-th reflection in calculation; it is determined by the (1/r) law; A<sub>0</sub> is the amplitude of the direct sound being unity (2.13), (2.15)

**A<sub>IV,l</sub>, A<sub>IV,r</sub>** Amplitudes of the ABR wave IV obtained from the left and right of auditory pathway, respectively (4.1)

**A<sub>V</sub>** Averaged amplitudes of the ABR wave V, for those of V<sub>l</sub> and V<sub>r</sub>, which are obtained on the left and right sides in auditory pathway, respectively (4.1)

**a, b** Weighting coefficients of the spatial factor extracted from the IACF to calculate the scale value of ASW (7.5)

**a, b, . . . e** Weighting coefficients of the temporal and spatial factors and the orthogonal factors of the sound field to calculate the scale value of dissimilarity (9.17)

**c** Speed of sound in the air (m/s) (2.16)

**c** Constant to calculate the preferred initial time delay gap between the direct sound and the first reflection (3.1b)

**c<sub>l,r</sub>(t)** Impulse responses of the vibration system of the left and right ossicle chains, i.e., from the eardrum to the oval window, including the transformation factor into vibration motion at the eardrum, (Fig. 5.1)

**D<sub>x</sub>** Psychological distance between sound fields of *a* and *b*, *x* being temporal and spatial factors of the sound field (9.1), (9.2), (9.6), (9.8) through (9.15)

**d<sub>0</sub>** Distance between the location of sound source and a listener for the direct sound (2.15)

**d<sub>1</sub>** Distance between the location of sound source and a listener for the path of the first reflection, *n* = 1 in (2.15), (2.16)

**d<sub>k</sub>** Distance between S<sub>k</sub><sup>T</sup> and S<sub>X</sub><sup>SF</sup> (9.1)

**$d_n$**  Distance between the source and the observation point for the  $n$ -th reflection,  $n = 1, 2, 3, \dots$  (2.15)

**$f$**   $10^{-15}$ ; other examples of the unit: m,  $10^{-3}$ ;  $\mu$ ,  $10^{-6}$ ; n,  $10^{-9}$ ; p,  $10^{-12}$

**$f$**  Frequency (Hz)

**Flu\_IACC** Fluctuation of the IACC, which influences annoyance of noise (11.2)

**Flu\_ $\tau$ IACC** Fluctuation of the  $\tau$ IACC, which influences annoyance of noise (11.4)

**$f_{l,r}(t)$**  Sound pressures arriving at the left and right ear entrances (2.12), (2.14)

**F0** Fundamental frequency, the repetition period of a periodic sound.

**F0<sub>a</sub>, f0<sub>b</sub>** Fundamental frequencies of sounds A and B, (Fig. 6.3)

**$f$**  Function of any one-dimensional subjective response in relation to physical factors (3.5), (6.1)

**$g_{l,r}(t)$**  Head-related impulse responses from a point source to two ear entrances in a room (2.13)

**$f(x_i)$**  Scale value as a function of the four orthogonal factors  $x_i$ ,  $i = 1, 2, 3, 4$  (3.6)

**$h(t)$**  Vertical and horizontal movements of the target (14.2)

**$h_{nl,r}(t)$**  Head-related impulse response from a point source to the left- and right-ear entrance of the  $n$ -th reflection (2.13)

**IACC** Magnitude of interaural crosscorrelation, which is the maximum value of the interaural crosscorrelation function (IACF), the spatial factor of the sound field defined by (2.26). One of four orthogonal factors of the sound field.

**$i$**  Alternative period of the mental tasks ( $i = 2n$ ) (11.6)

**$L_{eq}$**  Equivalent sound-pressure level (SPL) measured after obtaining the signal envelope (2.8)

**$I(t)$**  Luminance as a function of time of the stimulus produced by the LED (15.1)

**$L_{Horizontal}$**  Localization in the horizontal plane (7.1)

**$L_{Median}$**  Localization in the median plane (7.2)

**LL** The binaural listening level (2.24). When the A-weighting network as an approximation of the ear sensitivity is used, the unit is given by dBA. One of the four orthogonal factors of the sound field.

**O1, O2** see T3, T4

**P** The normalized ABR magnitude corresponding to the IACC (4.1)

**$P(\omega)$**  Pressure spectrum obtained by the Fourier transform of  $p(t)$  (2.2)

**$P_d(\omega)$**  Power density spectrum (2.1), (2.6)

**p** Significance level

**p(t)** Source signal as a function of time (2.3)

**p'\_{l,r}(t)**  $p'_{l,r}(t) = p(t) * s(t)$ , where  $p_{l,r}(t)$  are the sound pressures arriving at left and right ear entrances, and  $s(t)$  is the ear sensitivity (2.4)

**P\_{IACC}** Period of the fluctuation of  $P_{IACC}$

**P\tau\_{IACC}** Period of the fluctuation of  $\tau_{IACC}$

**r** Correlation coefficient

**r<sub>0</sub>** Location of the sound source,  $r_0 = (x_0, y_0, z_0)$  (2.12)

**r<sub>12</sub>** Correlation coefficient between a sensation  $S_1$  and another sensation  $S_2$  defined by (6.4)

**s** Integer

**s(t)** Ear sensitivity expressed in the time domain (2.4). For the practical application, we can use the impulse response of the A-weighting filter as an approximation.

**S** Scale value of any subjective response (3.5); or the scale value of subjective preference of the sound field (3.7)

**S** Total surface in a room to determine the reverberation time (2.20)

**S<sub>i</sub>** Scale values of preference as a function of the listening level (LL), the initial time delay gap between the direct sound and the first reflection ( $\Delta t_1$ ), the subsequent reverberation time ( $T_{sub}$ ) and the IACC, respectively,  $i = 1, 2, 3, 4$  (3.6), (3.7)

**S<sub>j</sub>** Any primary sensation,  $j = 1, 2, \dots, J < I$  (6.1), (6.2)

**S<sub>L</sub>, S<sub>R</sub>** Temporal and spatial sensations, which are associated with the left and the right hemispheres, respectively (6.5). Overall subjective responses such as preference or annoyance are given by  $S = S_L + S_R$  (6.6).

**S<sub>I<sub>K</sub></sub>** Speech intelligibility of the syllable  $K$  (%) (9.3)

**s(t)** Impulse response of the A-weighting filter, used as an approximation of ear sensitivity (2.22), so that  $f'_{l,r}(t) = f_{l,r}(t) * s(t)$

**T** Time interval (s) (2.4), (2.22)

**T** Tesla: unit of a magnetic flux density,  $1T = 1Wb/m^2$

**(2T)<sub>r</sub>** Recommended temporal duration of signal segments to be used in running autocorrelation analyses, which is related to  $(\tau_e)_{min}$  (5.3)

**T3, T4** Scalp locations for electrode placement according to the 10–20 International System. Temporal scalp locations (T3, T4, T5, T6) mainly reflect activity in auditory cortical regions of either the left (T3, T5) or right (T4, T6) hemisphere. Occipital locations (O1, O2) mainly reflect activity in visual cortical regions, Fig. 15.4., p. 269

**$T_{\text{sub}}$**  Subsequent reverberation time defined by the decay rate to decrease to 60 dB just after early reflections,  $T_{\text{sub}} \cong T_{60}$  (s): corresponding to Sabine's formulas given by (2.20), one of four orthogonal factors of the sound field

**$[T_{\text{sub}}]_p$**  Calculated preferred subsequent reverberation time (s), expressed by the term  $(\tau_e)_{\text{min}}$  (3.3)

**t** Time (s)

**Var\_SPL** Variance  $\sigma^2$  of the SPL (11.1)

**Var\_ $\tau_1$**  Variance  $\sigma^2$  of the  $\tau_1$  (11.1)

**$V_{l,r}(x,\omega)$**  Wave form of the basilar membranes, where  $x$  is the position measured along the left and right basilar membranes from the oval window, Fig. 5.1

**$W_{\text{IACC}}$**  Width of the IACC or width of the IACF at  $\tau_{\text{IACC}}$  (7.3). One of the significant spatial factors related to apparent source width ASW (7.4–7.6), Fig. 2.8

**$w_n(t)$**  Impulse response describing the reflection property of a wall in a room,  $n = 1, 2, 3, \dots$  (2.13), (2.18)

**$x_i$**  The normalized orthogonal factors,  $i = 1, 2, 3, 4$  (3.6) through (3.11)

**$\alpha$**  Alpha wave (8-13 Hz) in EEG and MEG; the value of  $\tau_e$  extracted from its ACF corresponds well with the scale value of subjective preference

**$\alpha(t)$**  Alpha wave as a function of time

**$\alpha_i$**  Weights of the scale value of subjective preference for each orthogonal factor,  $i = 1, 2, 3, 4$  (3.7), Table 3.4

**$\alpha, \beta, \gamma$**  Weighting coefficients of the spatial factor extracted from the IACF to calculate the scale value of ASW (7.6)

**$\beta$**  All of results of the scale value of subjective preference as a function of a normalized factor obtained by the PCT are expressed commonly by the  $(3/2)$  power of the factor, such that , where  $\alpha$  is the weighting coefficient.

(3.7): for normalized four orthogonal factors of the sound field of listeners;

(8.2): for the normalized factor of the sound field  $x = \log \Delta t_1 - \log [\Delta t_1]_p$  of cellists;

(14.1): for the normalized factor  $x = \log \phi_1 - \log [\phi_1]_p$  of the visual field of observers;

(15.2): for the normalized period  $x = \log_{10} T - \log_{10} [T]_p$  of the flickering light of observers;

(7.7): for describing the scale value of subjective diffuseness as a function of the IACC of listeners; and also,



$\Delta t_1 = (d_1 - d_0)/c$  (s); Initial time delay gap between the direct sound and the first reflection; one of four orthogonal factors of the sound field (2.17)

$\Delta t_n$  Delay time of the  $n$ -th reflection relative to the direct sound (s);  $\Delta t_1 = (d_1 - d_0)/c$  (s),  $n = 1$  is one of four orthogonal factors of the sound field (2.17)

$[\Delta t_1]_p$  Calculated preferred initial time delay gap between the direct sound and the first reflection (s), one of the temporal factors of the sound field (3.2), (3.4)

$\delta(t)$  Dirac delta function (2.13)

$\delta_1$  Delay time of the first peak extracted from the spatial ACF of the signal in the spatial domain, corresponding to spatial pitch (m)

$\delta_e$  Effective duration extracted from the spatial ACF of the signal in the spatial domain, defined by the delay time at which the envelope of the normalized ACF becomes and then remains smaller than 0.1 (m)

$\Phi(\Delta x, \Delta y)$  The two-dimensional ACF for describing spatial sensations of texture (13.1)

$\Phi_{lr}(v)$  Auditory mechanism of the interaural crosscorrelation function (IACF), Fig. 5.1

$\Phi_{lr}(\tau)$  Interaural crosscorrelation function (IACF) (2.22), (5.4)

$\Phi_p(\tau)$  The autocorrelation function (ACF) of the signal  $p(t)$  (2.4), (2.5)

$\Phi_p(\tau; \tau, T)$  The short-time moving ACF or the running ACF (2.10)

$\Phi_p(0)$  Sound energy given by the time origin of  $\Phi_p(\tau)$ ,  $\tau = 0$  (2.7), (2.9)

$\Phi_{ll}(0)$ ,  $\Phi_{rr}(0)$  ACF at the time origin, which corresponds to the sound energy at the left and right ear entrances, respectively (2.23), (2.24)

$\Phi_{ll}(\sigma)$ ,  $\Phi_{rr}(\sigma)$  ACF mechanisms in the left and right auditory pathway, respectively, Fig. 5.1

$\Phi_{ll}(\tau)$ ,  $\Phi_{rr}(\tau)$  ACF of the sound signals at the left- and the right-ear entrances, respectively (5.1)

$[\Phi_{ll}(0)\Phi_{rr}(0)]^{1/2}$  Geometric mean of sound energies arriving at the left ear and at the right ear (2.23), (2.24)

$\phi(\Delta x, \Delta y)$  Two-dimensional normalized ACF for describing spatial sensations of texture (13.2)

$\phi_1$  Amplitude at the first major peak at the delay  $\tau_1$  in ACF,  $\phi_p(\tau)$ , corresponding to perceived pitch strength, Fig. 2.1, or signal regularity, a degree of fluctuation of the signal

$\phi_{lr}(\tau)$  Normalized interaural crosscorrelation function (2.23), (5.5)

$\phi_p(\tau)$  Normalized ACF,  $\phi_p(0) = 1$  (2.7), (2.10)

- $\eta$  Elevation angle
- $\nu$  Time delay (s)
- $\sigma$  Time delay (s)
- $\sigma$  Standard deviation
- $\Sigma$  Summation
- $\tau$  Time delay (s)
- $\tau_1$  Delay time of the first peak in the ACF,  $\phi_p(\tau)$ , corresponding to pitch period (s), Fig. 2.1
- $\tau_e$  Effective duration of the ACF, defined by the delay time at which the envelope of the normalized ACF becomes and then remains smaller than 0.1 (the tenth-percentile delay) (s), Fig. 2.2. Effective duration is a measure of the temporal coherence of the signal, i.e. how long a stable temporal pattern faithfully repeats itself.
- $\tau_{\text{IACC}}$  Interaural delay time, at which the maximum IACF (IACC) is observed, Fig. 2.8
- $\tau_e(\text{IACC} = 0.3)$  Value of ACF  $\tau_e$  of the EEG alpha wave obtained by the stimulus with IACC = 0.3 corresponding to a large spatial subjective diffuseness
- $\tau_e(\text{IACC} = 0.95)$  Value of ACF  $\tau_e$  of the EEG alpha wave obtained by the stimulus with IACC = 0.95 corresponding to a narrow subjective diffuseness, Fig. 4.22. The ratio of the values of  $\tau_e$  for the alpha wave changing the IACC, [ $\tau_e(\text{IACC} = 0.3)/\tau_e(\text{IACC} = 0.95)$ ] in the right hemisphere is greater than that in the left hemisphere, Fig. 4.23.
- $(\tau_e)_{\min}$  Minimum value of  $\tau_e$  obtained by analyzing the running ACF on a segment of source signal 2T (s), which is the most active part of the signal. Thus, subjective preference may be critically judged at this particular part in the music signal. For example, echo may be easily perceived at this running part.
- $\tau_p$  Most preferred delay time of the first reflection,  $\tau_e = [\Delta t_1]_p$  (3.3)
- $\omega$  Angular frequency ( $= 2\pi f$ ,  $f$  being frequency [Hz])
- $\xi$  Horizontal angle of sound incidence to a listener ( $^\circ$ )
- $\oplus$  Signals from the left and right auditory pathway to be combined, Fig. 5.1

# Abbreviations

- ABR      The auditory brainstem response an evoked potential, activity from six nuclei in the auditory pathway (the latency is less than 10 ms), The auditory brain response is a stimulus-triggered, averaged, evoked short-latency (0–10 ms) neural electrical gross potential that is generated in response to a train of clicks. The ABR reflects the successive synchronous firings of auditory neurons in the cochlea, brainstem, and midbrain (i.e. roughly the early impulse response of the auditory system), Section 4.1
- ANOVA    Analysis of variance that reveals statistically-significant factors and interference effects between factors
- ASW      Apparent source width, the perceived horizontal size of a sound source, one of the spatial sensations, which is described by spatial factors extracted from the IACF of the sound field, Section 7.2
- ACF      Autocorrelation function. The temporal sensations can be described by temporal factors extracted from the ACF of the sound signal.
- NACF     Normalized ACF, an autocorrelation function normalized by its maximum, zero-delay value so that the function is rendered independent of the absolute amplitude of the signal, Sections 2.2, 5.2, and 5.3
- CCF      crosscorrelation function. The CCF indicates correlations between the values of two signals as a function of relative delay (lead or lag). For example, the CCF between the alpha waves from different electrodes over two cerebral hemispheres, Section 4.4.
- DS       Duration sensation, which is introduced here as one of four temporal sensations, Section 6.5
- EEG      Electroencephalogram, Sections 4.3, 15.1, and 15.3
- FF       Fundamental frequency (Hz), denoted  $F_0$ , the repetition frequency of an acoustic waveform, and the main physical correlate of pitch, (see also  $\tau_1$ ), Sections 6.2 and 6.3

GAs	Genetic algorithms, a class of nonparametric adaptive methods for optimizing combinations of design parameters
HRTF	Head-related transfer function, equivalent to the head-related impulse responses $h_{nl,r}(t)$ , Section 2.2
IACC	Magnitude of the IACF, the maximal value of the IACF, the most significant and a consensus factor in the four orthogonal factors of the sound field, Sections 2.2 and 5.4
IACF	Interaural crosscorrelation function. The spatial sensations of the sound field can be described in terms of the spatial factors extracted from IACF by analyzing sound signals at the two ears arriving at two ear entrances, Sections 2.2 and 5.5.
LED	Light-emitting diode, Sections 13.1, 14.1, and 15.1
LL	Binaural listening level (dBA), or binaural sound-pressure level measured by the geometric mean of $\Phi_{ll}(0)$ and $\Phi_{rr}(0)$ , Section 2.2
MEG	Magnetoencephalogram, Sections 4.4 and 15.2
NI	Nonidentification of speech (%), Section 9.2
PCT	Paired-comparison test (Thurstone, 1927; Gullikson, 1956; Torgerson, 1958): Most of the subjective preference judgment and other subjective responses in this volume were obtained by the PCT. Usually, trials started with a first stimulus, followed by a short blank duration and then a second stimulus. During the subsequent blank duration, the subject judged which stimulus was the subjectively preferred stimulus. The scale value is related to the probability whether stimulus A is preferred to B. For example, if $P(A > B) = 0.84$ , then the value is 1.0. The value, therefore, may be reconfirmed by the goodness of fit (Mosteller, 1951). All data in this volume were reconfirmed by the test. This shows that the model of obtaining the scale value was approved. The scale values of the subjective judgments of each individual subject can also be calculated (Ando and Singh, 1996; Ando, 1998). If the experimental procedure is identical, then the probability data may be integrated over the time and space. Because of its simplicity, and ease of use, it generates reliable and reproducible response data from naive subjects, even children. Experiments that recorded SVR, EEG, and MEG signals were performed in a similar manner to the paired comparison to find the relationship between the factor extracted from the correlation analyses of signal recorded and the scale value of subjective preference judgments.
PET	Positron emission tomography, Section 15.3
PLG	Plethysmogram; a short-term running measure of blood volume and pulse rate that was measured using a fingertip, photoelectric pulse oximeter device. The PLG provides a window on sympathetic and parasympathetic autonomic nervous system influences on peripheral blood vessels associated with stress and relaxation, Section 11.3. The peripheral blood vessels react as a reflection of the autonomic nervous system, which may be observed in the PLG, Section 11.3
SD	Standard deviation

- SI            Speech intelligibility (%), an index that measures the proportion of speech that is audible and correctly recognized by the listener, Section 9.1
- SL            Sensation level (dB), the sound level of a signal in relation to the listener's threshold of audibility, Section 4.2
- SV            Scale value obtained from the PCT. The value given by S may be described by the temporal and spatial factors.
- SVR          The slow vertex response is a stimulus-triggered, averaged, evoked middle- and long-latency (10–500 ms) neural electrical gross potential that is generated in response to a train of clicks and recorded from scalp electrodes. The SVR reflects the successive synchronous firings of auditory neurons primarily in the two hemispheres of the cerebral cortex (i.e. roughly speaking, the SVR can be regarded as the gross impulse response of upper auditory stations), Section 4.2

# Author Index

*NOTE:* The letters ‘n’ denote the note numbers in the text

## A

Alho, K., 89  
Allan, L. G., 262–263  
Alrutz, H., 4, 190  
Amandasun, M., 244  
Ando, D., 3–4, 34, 191  
Ando, Y., 3, 6, 8, 10, 12–13, 20–23, 25–27,  
29–30, 32–36, 34n1, 36–38, 39, 41, 48–50,  
51, 53, 56, 58, 64, 70, 73, 75, 76, 84–87,  
91, 96, 105, 109, 119–120, 121, 125, 130,  
131, 133, 136, 137, 143, 150, 153, 158,  
165, 166, 168, 172, 174, 175, 176, 178,  
179, 185, 189, 190, 196, 199, 205, 209,  
211, 228–230, 231, 248, 261, 263, 265

## B

Badcock, D. R., 246  
Ball, K., 262  
Barlow, J. S., 276  
Bekesy, G., 175  
Belin, P., 88  
Ben-Av, M. B., 245, 251  
Beranek, L. L., 223  
Bergen, J., 244  
Berglund, B., 216  
Bialek, W., 258  
Bowen, R. W., 237  
Braizer, M. A. B., 276  
Brodatz, P., 244, 248  
Buchwald, J. A. S., 42  
Burd, A. N., 12n1, 56

## C

Campbell, F. W., 248  
Cariani P., 6, 78, 95, 104, 235, 259

Cariani, P. A., 6, 78, 79, 95,  
104, 220  
Casby, J. U., 276  
Cawthon, J. M., 205  
Chandler, D. W., 227  
Chen, C., 6, 56, 58, 268  
Chernyak, R. I., 109  
Cherry, E. C., 6, 94  
Cheveigne, A., 95  
Cho, R. Y., 243–244  
Chon, R., 71  
Clarke, J., 258  
Clottes, J., 265  
Cocchi, A., 190  
Colburn, H. S., 227  
Colburn, S., 6  
Courtin, J., 265  
Cross, G. R., 244

## D

Damaske, P., 3, 22, 29, 86, 125, 175  
Davis, A. E., 296  
de Boer, E., 94–95  
de Cheveigne, A., 95  
de Lange, H., 237, 253–254  
de Valois, K. K., 244  
de Valois, R. L., 244  
Delgutte, B., 6, 78, 79, 95, 104, 220  
Derrington, A. M., 246  
Doi, S., 258  
Dubrovskii, N. A., 109

## E

Eady, H. R., 227  
Edward, R. M., 190

Eisner, A., 237  
 Ellis, R. R., 296  
 Eulitz, C., 89, 286

**F**

Fastl, H., 122  
 Finkel, L. H., 249–250  
 Fraisse, P., 242  
 Francos, J. M., 245  
 Fujii, K., 176, 199, 200, 216, 218–219, 235,  
 238, 248–249, 263

**G**

Gabriel, K. J., 227  
 Gade, A. C., 165  
 Galambos, R., 47  
 Galin, D., 296  
 Gottlob, D., 3, 34, 191  
 Grantham, D. W., 227  
 Greenwood, D. D., 109, 213, 216  
 Gros, B. L., 262–263  
 Gullikson, H., 25, 34

**H**

Hammett, S. T., 246  
 Hanada, K., 7  
 Haralick, R. M., 244  
 Hargest, T. J., 216  
 Hecox, K., 47  
 Heeger, D., 244  
 Hellman, R. P., 216  
 Henning, G. B., 238, 246  
 Hewitt, M. J., 95, 220  
 Hidaka, T., 132  
 Hinrichs, H., 268, 280, 287  
 Hiramatsu, K., 217  
 Holland, J. H., 148  
 Hoovey, Z. B., 276  
 Hosaka, I., 6–7, 41  
 Hotehama, T., 7, 158, 190  
 Huang, C. M., 42

**I**

Imamura, M., 34  
 Inagaki, T., 258  
 Inoue, M., 6, 79, 102, 107, 143  
 Inoye, T., 276

**J**

Jain, A. K., 244  
 Jasper, H. H., 49  
 Jeffress, L. A., 6, 94, 246  
 Jewett, D. L., 42  
 Julesz, B., 244

**K**

Kageyama, K., 12, 27, 29, 64  
 Kang, S. H., 26, 228  
 Kaplan, S., 254  
 Kato, K., 31, 173n1, 176  
 Katsuki, Y., 70, 77, 109  
 Keet, M. V., 130, 131, 135  
 Kiang, N. Y. -S., 77  
 Kimura, D., 53, 89  
 Kinchla, R. A., 262–263  
 King, R., 244  
 Kitamura, T., 207  
 Klumpp, R. G., 227  
 Korenaga, Y., 181–182  
 Kryter, K. D., 216  
 Kurihara, Y., 6–7, 136  
 Kuttruff, H., 19  
 Kuwano, S., 216

**L**

Lev, A., 42  
 Levinson, E., 262  
 Levy, J., 296  
 Licklider, J. C. R., 6, 78, 94, 246  
 Lindsey, D. B., 268  
 Liske, E., 276  
 Liu, F., 245, 250  
 Lohse, G. L., 244  
 Lundeen, C., 97

**M**

Machleidt, W., 268, 280, 287  
 Maekawa, Z., 132  
 Maki, F., 153–154, 156  
 Makous, W., 253–254  
 Malik, J., 244  
 Mandler, M. B., 253–254  
 Mao, J., 244  
 Marshall, A. H., 165  
 Martens, W. L., 122  
 Marui, A., 122  
 Mathews, M. V., 109  
 May, D. N., 221–222

McLachlan, K. R., 276  
 Meddis, R., 6, 78, 95, 220  
 Mehrgardt, S., 126–127  
 Mellert, V., 126–127  
 Merthayasa, I. G. N., 112, 209  
 Molino, J. A., 222  
 Moore, B. C. J., 95  
 Morimoto, M., 132  
 Morioka, K., 26, 34, 36–37  
 Mosteller, F., 25, 34  
 Mouri, K., 31, 85, 87, 165, 190

**N**

Näätänen, R., 89  
 Nachmias, J., 238  
 Nagamatsu, H., 51  
 Nakagawa, S., 55, 82, 88, 301n3  
 Nakajima, T., 153–154, 158  
 Nakayama, I., 32, 165–166  
 Noson, D., 158, 166, 175

**O**

O'Mard, L., 6  
 Ohgushi, K., 124  
 Okamoto, Y., 88, 288–289, 295  
 Opitz, B., 89

**P**

Palomaki, K., 82n4, 301n4  
 Pearsons, K. S., 216  
 Perona, P., 244  
 Petsche, H., 268, 280,  
 286–287  
 Pfafflin, S. M., 109  
 Picard, R. W., 245  
 Pinker, R. A., 216  
 Pompoli, R., 199  
 Portilla, J., 245  
 Pressnitzer, D., 104  
 Probst, Th., 292

**R**

Raney, J. P., 205  
 Rao, A. R., 244  
 Rappelsberger, P., 280,  
 286–287  
 Raymond, J. E., 262–263  
 Ritsma, R. J., 95  
 Robson, J. G., 248

Rogowitz, B. E., 238  
 Rose, J. E., 94  
 Runderman, D. L., 258

**S**

Sabine, W. C., 19  
 Sagi, D., 245, 251  
 Saifuddin, K., 6, 119  
 Sakai, H., 101, 153, 199, 205–206  
 Sakai, K., 249  
 Sakamoto, M., 34, 205  
 Sato, S., 6–7, 32, 38, 60, 88, 114, 126, 130,  
 131, 133, 148, 166, 174, 175, 190, 214,  
 216, 222, 223, 276, 299n3  
 Sayers, B. M. A., 6, 94  
 Sayles, M., 91, 298n3  
 Scharf, B., 213, 216  
 Schroeder, M. R., 3, 29, 254  
 Searle, C. L., 6, 78  
 Secker-Walker, H. E., 6, 78  
 Seebeck, A., 94  
 Sekuler, R., 262–263  
 Shaw, J. C., 276  
 Shimokura, R., 87, 105, 158  
 Siebrasse, K. F., 3  
 Simoncelli, E. P., 245  
 Singh, P. K., 121, 136, 153, 158, 168, 269  
 Small Jr., A. M., 97  
 Smetana, B., 174–175  
 Smith, A. T., 246  
 Soeta, Y., 6, 8, 48, 54–55, 63, 66, 68, 70, 72,  
 82n3, 88, 112, 214, 235, 253–254, 256,  
 259, 263, 268, 276, 283, 301n3  
 Sohmer, H., 42  
 Sperry, R. W., 53, 89, 296  
 Sumioka, T., 96  
 Suzumura, Y., 156, 158

**T**

Taguti, T., 13, 31, 176  
 Tamura, H., 244, 248–249  
 Terhardt, E., 95, 214  
 Thompson, A. M., 47  
 Thompson, G. C., 47  
 Thurstone, L. L., 25, 34  
 Torgerson, W. S., 25, 34, 190  
 Trevarthen, C., 296  
 Tsutsumi, T., 177  
 Turner, M. R., 244



**U**

Uttal, W. R., 248, 251

**V**

van de Grind, W. A., 262

van Noorden, L., 95

Voss, R. F., 258

**W**

Wada, J. A., 296

Wever, E. G., 94

Wightman, F. L., 95

Winter, I. M., 91, 298n3

Wong, P. C., 88

Wu, S., 254

**Y**

Yamaguchi, K., 190

Yost, W. A., 75, 95, 96, 213

Yrttiaho, S., 104

**Z**

Zatorre, R., 88

Zeki, S., 292

Zhu, S. C., 244

Zwicker, E., 109, 122, 213–214, 216

# Subject Index

## A

- Aircraft noise measurement, 200–206
  - measured factors extracted from ACF, 203–204
  - $W_{IACC}$  value, 205
- Alpha wave in EEG
  - ACF analysis, 267–275
  - CCF analysis, 275–282
- Anisotropic texture, 246
- Annoyance
  - alpha wave corresponding with, 68–72
    - ACF analyses, 68–72
    - CCF analyses, 68–72
  - MEG response corresponding with, 63–72
  - noise effects on, 301–303
  - scale value of, 71–72
- Annoyance of noise, 213–232
  - in relation to spatial factors, 223–228
    - both SPL and IACC, 223–225
    - both SPL and  $\tau_{IACC}$ , 225–228
  - in relation to the temporal factors, 213–222
    - band-pass filtered noise, 213–218, *See also individual entry*
    - traffic noise, 218–222, *See also Traffic noise annoyance*
- See also under Children*
- Apparent source width (ASW), 91, 127–136
  - of band-pass filtered noise, 130–136
    - NACF of, 127
    - scale value, 130–136
  - of complex noise, 132–134
  - in concert hall sound field, spatial expression due to, 175
- Auditory brainstem responses (ABRs)
  - examination, 6, 39, 40–48
  - ABR amplitude
    - corresponding with IACC, 44–46
    - corresponding with sound pressures, 44–46
    - from the left and right auditory pathway, 74
    - neuronal correlates of horizontal sound direction, 40–44
      - averaged amplitudes, 42
      - averaged latencies, 44
      - left and right ABRs, 41, 47
      - neural signals flow, 43
      - relative latencies of ABR peaks, 44
- Auditory sensations, 297–301
  - ACF, 298
  - IACF, 299
  - noise effects on tasks and annoyance, 301–303
  - spatial, 299–300
  - subjective preference, 300–301
  - temporal, 298–299
- Auditory spatial sensations, 299–300
  - IACF, 299
- Auditory system model for temporal and spatial information processing, 4–8
- Auditory temporal sensations, 298–299
- Auditory-temporal window  $2T$ , 84–85
- Autocorrelation (ACF) mechanism, 5–8
- Autocorrelation function (ACF) model, 45
  - in central auditory signal processing pathways, 81–83
  - autocorrelation histograms, 80
  - autocorrelogram, 80
  - See also under Temporal factors*
- for pitch calculation, 98–100
  - frequency limits of, 101–105
  - in-phase waveforms, 96
  - normalized autocorrelation function (NACF), 97
  - pitch-matching test, 98–99
  - random-phase waveforms, 96
- of source signal, 10–13
  - effective duration, 11
  - music source signals, 12

- Autocorrelation function (ACF) model (*cont.*)  
 normalized ACF, 10  
 running ACF, 13–18  
 significant items extracted from, 10  
 speech source signals, 12  
 $\tau_1$ , temporal factor, 11  
 $\phi_1$ , temporal factor, 11
- B**
- Band-pass filtered noise annoyance  
 measured ACF factors, 215  
 as a function of bandwidth, 215  
 scale value and filter bandwidths,  
 relationship, 217–218  
 in relation to the temporal factors, 213–218  
 loudness, 213–214, 216
- Band-pass filtered noise, ASW of, 130–136  
*See also under* Apparent source width (ASW)
- Bandwidth and loudness, 111–113
- Beat induced by dual missing fundamentals,  
 105–108  
 in-phase condition, 106  
 random condition, 106
- Binaural signals, spatial sensations of, 125–141
- C**
- Central auditory system, signal processing in,  
 39–72, 73–89  
 ABRs examination, 40  
 ACF processor in, 81  
 alpha wave range signal analysis, 40  
 human aspect of EEG, 39–72  
 human aspect of MEG, 39–72  
 interaural crosscorrelation mechanism,  
 86–87  
 left and right auditory pathway  
 ABRs from, 74  
 EEG on, 74–75  
 MEG on, 75  
 SVR on, 74  
 neural evidence, 73–75  
 physical system, 73  
 pooled autocorrelation histograms, 80  
 pooled autocorrelogram, 80  
 for subjective responses, 76  
 SVRs examination, 39  
*See also* Auditory brainstem  
 responses (ABRs) examination;  
 Electroencephalography (EEG);  
 Magnetoencephalography (MEG);  
 Slow vertex responses (SVRs);  
 Spatial factors
- Cerebral hemispheres specialization  
 for sound field, 87–89  
 for visual field, 295–296
- Children  
 noise and music effects on  
 hemispheric tasks, 229–232  
 V-type relaxed children during adding  
 task, 229  
 V-type relaxed children during search  
 task, 230
- Coarseness, 243–245
- Complex noise, ASW of, 132–134  
 scale values, 134
- Complex tones and noise, pitches of, 93–105  
*See also under* Loudness; Pitches of  
 complex tones and complex noise
- Composition, at concert hall, 173–174
- Concert hall acoustics, 143–178  
 concert hall as a musical instrument,  
 172–175  
 composition, 172–174  
 matching temporal factor of sound field,  
 172–174  
 performance, 173–174  
 program music, 173–174  
 spatial expression due to ASW in sound  
 field, 174–175  
 musical performance blending the sound  
 field, 175–178  
*See also under* Performance, at concert  
 hall  
 as a public space, 148–158  
 actual design studies, 153–158  
 Kirishima International Concert Hall,  
 153–158  
 sound field design studies, 148–158,  
*See also* Genetic algorithms (GAs),  
 in concert halls design studies  
 Tsuyama Music Cultural Hall, 156  
 variation–evaluation–selection  
 cycle, 149  
 subjective preference of cellists  
 for delay time of single reflection,  $\Delta t_1$ ,  
 165–172  
*See also* Piano sound, pitch of; Seat  
 selection in concert hall
- Contrast, 243–245
- Correlation noise measurement method, 200  
 envelope-based and correlation, compari-  
 son, 200  
 physical factors, 200

- spatial sensations, 200
- temporal sensations, 200
- Crosscorrelation function (CCF) analysis, 40
  - of alpha wave
    - in EEG, 275–281
    - over the scalp, 88

**D**

- Delay time, preferred, of single reflection, 26–29
  - maximum score, 26
  - normalized scores, 26
  - preference scores as a function of delay, 27
  - preferred delay and ACF duration, relationship between, 29
  - subjective attributes before and after, 28
  - total score, 26
- Design studies in concert halls, 148–158
  - See also under* Concert hall acoustics; Seat selection in concert hall
- Diffuseness, 91
- Dissimilarity judgment of sound fields in real hall, 189–197
  - distance due to spatial factors, 195–197
    - correlation coefficients, 196
    - partial regression coefficients, 196
  - distance due to temporal factors, 194–195
  - listening position, 190
  - maximum-length sequence (MLS) signal, 190
  - scale values, 192
- Duplex model, 246
- Duration sensation (DS), temporal, 119–120
  - scale values obtained by PCT, 120

**E**

- Early reflections after the direct sound ( $\Delta t_1$ ), 31
- Electric guitar sound with distortion, timbre of, 120–123
- Electrical responses in central auditory system, 39–72
  - See also* Auditory brainstem responses (ABRs) examination; Electroencephalography (EEG); Magnetoencephalography (MEG); Slow vertex responses (SVRs)
- Electroencephalography (EEG), 39
  - on the left and right hemispheres, 74–75
  - response with subjective preference, 58–60
    - ANOVA results, 59

- averaged value of ACF  $\tau_e$ , 58
  - in response to change of  $\Delta t_1$ , 55–58
  - in response to change of IACC, 60–63
  - in response to change of  $T_{sub}$ , 58–60
- Envelope-based noise measurement method, 199–200
  - and correlation, comparison, 200
  - physical factors, 200
  - spatial sensations, 200
  - temporal sensations, 200
- Environmental noises evaluation, 223–228
  - See also* Annoyance of noise
- Equivalent current dipole (ECD), 54–55
- Existing concert hall evaluation
  - of sound field subjective preference based on temporal and spatial factors, 37–38

**F**

- Flickering light, subjective preference of, 253–259
  - EEG responses, 267–282
    - ACF analysis of the alpha wave, 267–275
    - ANOVA, 270–273
    - average scale values, 269
    - CCF analysis of alpha wave, 275–282
    - $\tau_e$  values, 274
    - $\phi$  values, 274
  - MEG responses, 282–288
    - for sinusoidal flickering light, 282–288, *See also under* Sinusoidal flickering light, MEG responses for preferred period of, 253
    - scale value of preference for, 256
    - $\tau_1$ , 254–255
    - $\phi_1$ , 254–255
- Flushing toilet noise measurement, 207–211
  - measured factors extracted from ACF, 208
  - plans of upstairs and downstairs in an apartment, 207
- Fourier transformation, 9–10
- Frequency and loudness, 109
- Frequency limits of ACF model for pitch calculation, 101–105

**G**

- Genetic algorithms (GAs), in concert halls
  - design studies, 148–158
  - scale value behavior, 150
  - shape, optimizing, 148–158

Genetic algorithms (GAs), in concert (*cont.*)  
 leaf-shape plan of, 156  
 parent chromosomes, 150  
 shoebox form, 151

## H

Harmonic complex tone, 93  
 Head-related impulse responses, 20  
 Head-related transfer function (HRTF), 126  
 Hemispheric response difference in change of  
 $\Delta t_1$ , 56–57  
 IACC, 56  
 LL (listening level), 56–57  
 Horizontal direction to a listener, preferred, of  
 single reflection, 29  
 Horizontal movement of single target  
 subjective preference for, 259–263  
 EEG responses, 289–295  
 two-way ANOVA, 290–292  
 Horizontal plane, cues of localization  
 in, 125  
 Horizontal sound direction, neuronal correlates  
 of, 40–44

## I

Individual subjective preference, in concert  
 hall, 158–161  
*See also under* Seat selection in concert  
 hall  
 In-phase waveforms, in ACF model for pitch  
 calculation, 98  
 Interaural crosscorrelation function (IACF)  
 mechanism, 5–8, 20, 75–76  
 magnitude, 21  
 normalized, 20–21  
 for signals arriving at left- and right-ear  
 entrances  
 spatial factors extracted from, 81–82  
 Interaural level difference (IALD), 20  
 Interaural time deference (IATD), 20  
 Isotropic texture, 246

## L

Listening level (LL), binaural, 25, 30  
 Localization, sound, 125–127  
 cues of localization  
 in the horizontal plane, 125  
 in the median plane, 126–127  
 Loudness, 108–118  
 bandwidth and, 108

of complex noise, 114–118  
 scale values, 114  
 spectrally complex noise, 114, 117  
 frequency and, 109  
 scale value of, 113  
 of sharply filtered noise, 108–113  
 Low pitch perception of complex tones,  
 93–100

## M

Magnetic responses in the central auditory  
 system, 39–72  
*See also* Auditory brainstem responses  
 (ABRs) examination; Elec-  
 troencephalography (EEG);  
 Magnetoencephalography (MEG);  
 Slow vertex responses (SVRs)  
 Magnetoencephalography (MEG), 7, 39,  
 63–72  
 EEG *versus*, 63  
 on the left and right hemispheres, 75  
 response corresponding with annoyance,  
 63–72  
 response corresponding with subjective  
 preference, 63–72  
 ACF analysis of alpha wave, 66–68  
 Magnitude of the IACF (IACC), 25, 33–34  
 ABR amplitude corresponding with, 46  
 Maximum-length sequence (MLS) signal, 190  
 Median plane, cues of localization in, 126–127  
 Missing fundamental phenomenon, 237–243  
 in-phase stimuli, 239–243  
 random-phase stimuli, 239–243  
 Missing reflection, 168

## N

Neuronal correlates of horizontal sound  
 direction, 40–44  
 Neuronal response correlates in change of  $\Delta t_1$ ,  
 48–50  
 $\Delta t_1$ , neuronal response, 25, 48–50  
 EEG response, 55–58  
 Noise/noise measurement methods, 199–211  
 correlation method, 200  
 effects on tasks and annoyance, 301–303  
 envelope-based and correlation, compari-  
 son, 200  
 physical factors, 200  
 spatial sensations, 200  
 temporal sensations, 200  
 envelope-based method, 199–200

*See also* Aircraft noise measurement;  
Annoyance of noise; Flushing toilet  
noise measurement  
Normalized autocorrelation function (NACF)  
of complex tones, 97

**O**

Optimal conditions maximizing subjective  
preference, 30–34  
IACC, 25, 33–34  
LL, 25, 30  
 $T_{\text{sub}}$ , 25, 33  
 $\Delta t_1$ , 25, 31  
Orthogonal factors of sound field, 25  
LL (Binaural and spatial factor), 25

**P**

Paired-comparison test (PCT) of subjective  
preference, 25–38  
Paired stimuli method, 49  
Perception models, 91–93  
psychoneural models of, 92  
psychophysical models of, 92  
Perceptual texture, 244  
Performance, at concert hall, 175–176  
sound field blending, 177–178  
control of source strength, 173  
position, selection, 178  
reverberation time, 177  
temporal factor of sound field, 177  
Periodic structure in texture modeling and  
synthesis, 245  
Physical attributes of subjective sensations,  
91–93  
individual differences, 93  
*See also* Perception models  
Physical factors of sound field, 18–23  
impulse responses of reflecting walls, 19  
n-th reflection, 19  
source signal, 20  
temporal-monaural factors, 19–20  
transmission from a point source to ear  
entrances in a room, 18  
impulse responses, 18  
*See also* Spatial-binaural factors  
Piano sound, pitch of, 143–147  
ACF analysis, 145–147  
88-note signals, 146–147  
source signal, 145  
Pitch of piano sound, 143–147  
*See also* Piano sound, pitch of

Pitches of complex tones and complex noise,  
93–105  
ACF model, 95–96  
pitch-matching test, 96  
autocorrelation model, 94–95  
beat induced by dual missing fundamentals,  
105–108  
binaural crosscorrelation, 94  
interspike interval information, 95  
low pitch perception of, 93–100  
NACF, 96–97  
time-domain cancellation models, 95  
Power spectrum of source signal, 9–10  
Preferred regularity of texture, 263–264  
scale values of, 264  
two-dimensional spatial textures, 264

**Q**

Quasi-Newton numerical method, 168, 255

**R**

Random-phase waveforms, in ACF model for  
pitch calculation, 96  
Regularity, 243–244  
Reverberation time after early reflections  
( $T_{\text{sub}}$ ), 33

**S**

Sabine's formula, 19  
Scale value  
of annoyance, 70–72  
of loudness, 109  
of subjective preference, 34–37  
contour lines, 37  
Seamless correlation model, 251  
Seat selection in concert hall  
enhancing individual preference,  
158–165  
individual subjective preference,  
158–161  
seat selection system, 158  
preferred conditions for each individual,  
158–165  
cumulative frequency of preferred LL,  
163  
cumulative frequency of preferred  
subsequent reverberation time, 163  
cumulative frequency of the preferred  
initial time delay gap, 163  
seating area, 161–162

- Sharply filtered noise, loudness of, 108–113  
*See also under Loudness*
- Simulation of sound field, 23–24
- Single reflection, sound field with, preferred conditions of, 26–29  
 preferred delay time, 26–29  
 maximum score, 26  
 normalized scores, 26  
 preference scores as a function of delay, 27  
 preferred delay and ACF duration, relationship between, 28  
 subjective attributes before and after, 28  
 total score, 26  
 preferred horizontal direction to a listener, 29
- Sinusoidal flickering light, MEG responses for, 282–288  
 for different degree of fluctuation of, 288–289  
 preferred periods of, 283  
 scale values, 283  
 $\tau_e$  values, 285  
 $\phi$  values, 285
- Slow vertex responses (SVRs)  
 of central auditory system, examination, 39, 48–55  
 differences in response latency, 53–55, *See also under Subjective preference*  
 direct stimulus sound field, 49  
 neuronal response correlates in change of  $\Delta t_1$ , 48–50  
 paired stimuli method, 49  
 reflection stimulus sound field, 50  
*See also Hemispheric response*  
 difference in change of  
 on left and right hemispheres, 74
- Sound coordinator as a specialist in concert hall, 176
- Sound field, 25–38, 87–89  
 spatial factors of, analysis, 9–24  
*See also under Physical factors of sound field*  
*See also under Single reflection, sound field with, preferred conditions of; Subjective preference*
- Sound pressure level (SPL), 223–228  
*See also under Annoyance of noise*
- Sound pressures, ABR amplitude corresponding with, 44–46
- Source signal, analyses, 9–18  
 power spectrum of, 9–10  
 temporal factors, 9–24  
*See also Autocorrelation function (ACF) model*
- Spatial auditory factors, 3–4  
*See also Temporal and spatial auditory factors*
- Spatial-binaural factors, 20–23  
 head-related impulse responses, 20  
 IACC, 23  
 IACF, 20  
 IALD, 20  
 IATD, 20  
 normalized interaural crosscorrelation function, 20–21  
 simulation system, 23–24  
 $W_{IACC}$ , 23  
 $\tau_{IACC}$ , 23
- Spatial expression in concert hall sound field due to ASW, 174–175  
 due to strength of music, 178
- Spatial factors  
 effect on speech nonidentification, 185–189  
 cocktail party effects, 185  
 distance calculated, 188  
 weighting coefficients, 188  
 extracted from IACF signals arriving at left- and right-ear entrances, 86–87  
 interaural delay time, 86  
 magnitude of interaural crosscorrelation, 86  
 width of IACF, 86  
*See also under Annoyance of noise*
- Spatial sensations  
 of binaural signals, 125–141  
*See also Apparent source width (ASW); Localization, sound; Subjective diffuseness*  
 of sound signal, 91–93  
 formulation of, 94  
*See also Visual spatial sensations*
- Spatial visual sensations, 299–300  
 ACF, 300  
 IACF, 300
- Speech identification in sound fields, 179–197  
 temporal factors effects on, 179–185  
 Japanese syllables, 181–182  
 single syllable, 182  
 unvoiced consonant, 182–184  
 voiced consonant, 182–184  
*See also under Spatial factors*  
*See also Dissimilarity judgment*
- SQUID (superconducting quantum interference device), 63

- Subjective diffuseness, 136–141  
 by IACC, 136–137  
   as a function of horizontal angle of incidence to a listener, 138  
   scale value of, 137
- Subjective preference, 267–296  
 auditory sensations, 298  
 cerebral hemispheres specialization for visual field, 295–296  
 differences in response latency corresponding with, 53–55  
   test sound field and reference sound field, 54
- EEG responses, 267–296  
*See also under* Flickering light, subjective preference of; Horizontal movement of single target
- of flickering light, 253–259
- MEG responses, 282–289  
*See also under* Flickering light, subjective preference of
- in relation to temporal and spatial factors, 253–265
- for temporal and spatial auditory factors, 3–7  
   applied to individual preference, 8  
   of texture, 263–265  
   for vertical and horizontal movements of target, 253  
   visual sensations, 297–298  
*See also under* Electroencephalography (EEG); Magnetoencephalography (MEG)
- Subjective preference of sound field, 25–38  
 existing concert hall evaluation  
   based on temporal and spatial factors, 37–38  
 optimal conditions maximizing, 30–34  
    $\Delta t_1$ , 31  
   IACC, 33–34  
   LL, 30  
    $T_{sub}$ , 33
- orthogonal factors of, 25  
   IACC (Binaural and spatial factor), 25  
   LL (Binaural and spatial factor), 25  
    $T_{sub}$  (Monaural and temporal factor), 25  
    $\Delta t_1$  (Monaural and temporal factor), 25
- theory of, 34–37  
   scale values, 35–36  
*See also* Paired-comparison test (PCT) of subjective preference; Single reflection, sound field with, preferred conditions of

## T

- Tasks, noise effects on, 301–303
- Temporal and spatial auditory factors, 3–4  
 brain and environment interaction created by, 4  
 information processing, auditory system model for, 4–8  
 auditory brainstem responses (ABRs), 6  
 binaural crosscorrelations, 6  
 magnetoencephalography (MEG), 7  
 slow vertex responses (SVRs), 7  
 spatial sensations and preference, 6  
 temporal sensations and preference, 6
- Temporal factors, 213  
 extracted from ACF of sound signal, 83–84  
    $\Delta t_1$ , 82  
   hemispheric specializations, 82  
   temporal window for ACF processing, 84–85  
    $T_{sub}$ , 82  
 of sound field, in concert hall acoustics, 177  
*See also under* Annoyance of noise; Speech identification in sound fields
- Temporal-monaural factors, 19–20
- Temporal sensations  
 of sound signal, 91–124  
 formulation of, 91–93  
*See also* Duration sensation (DS), temporal; Loudness; Pitches of complex tones and complex noise  
*See also* Visual temporal sensations
- Temporal visual sensations, 304  
 ACF, 300  
 IACF, 300
- Temporalwindow for IACF processing, 87
- Textural models, 243–251  
*See also under* Visual spatial sensations
- Texture, subjective preference of, 263–265  
 preferred regularity of texture, 263–264  
 spatial ‘vibrato’ in a drawing, 264–265
- Timbre of electric guitar sound with distortion, 120–124
- Time-domain cancellation models, pitches, 95
- Traffic noise annoyance  
 in relation to temporal factors, 218–222  
   measured ACF factors, 219, 221  
   scale value, 213  
   tonal and un-tonal noises, 220
- Transmission of sound from a point source to ear entrances in a room, 18



**V**

Vertical movements of target, subjective preference for, 259–263

Visual sensations, 304–306  
  spatial, 304

  subjective preference, 305–306  
  temporal, 304

Visual spatial sensations, 235–236, 243–251

  ACF model

    anisotropic texture, 246

    disadvantage, 244

    isotropic texture, 246

    one-dimensional spatial ACF, 247

    for perceptual grouping, 245

    scale value, 248–249

    of two-dimensional texture pattern, 243, 247

  coarseness, 243

  contrast, 243

  and preference, 236

  regularity, 243

  textural models/features, 243–244

    periodic structure in, 245

Visual temporal sensations, 237–251

  and preference, 236

*See also* Missing fundamental phenomenon