Perception and psychological evaluation for visual and auditory environment based on the correlation mechanisms

Fujii, Kenji

(Degree)
博士（学術）

(Date of Degree)
2002-03-31

(Date of Publication)
2009-03-17

(Resource Type)
doctoral thesis

(Report Number)
甲2548

(URL)
https://hdl.handle.net/20.500.14094/D1002548

Perception and psychological evaluation for visual and auditory environment based on the correlation mechanisms

Kenji Fujii

Submitted to the Division of Environmental Planning
Department of Global Development Science
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
at Kobe University

January 2002

Abstract

Modeling of the visual and auditory processing and understanding the relationship between physical and subjective property is the main topic in this dissertation. This research was motivated by the desire to understand primal attributes in vision and auditory sensation gathering toward complex psychological responses and evaluations in the real environment. As psychological responses mentioned here, subjective preference and subjective impressions of sounds, visual properties of landscape and interior space are supposed. To evaluate qualities of such materials from psychological or affective viewpoint, relationships between physical property and subjective evaluation are investigated.

The problem in understanding such psychological responses by a scientific manner is that this kind of information is difficult to express in language, therefore to definition and quantification. It is easy to understand as we experience it everyday in sights and sounds, but this is rather vague to describe these properties. The position I take in this dissertation is that such vague and complicated sensibilities occur as assemblies of subjective attributes processed in vision, audition, and so on. These attributes include color, shape, spatial and temporal characteristics of surface, or loudness, pitch, and timbre of sound. We need to know the way that human process information about these subjective attributes.

In this dissertation, I introduce the correlation mechanism for modeling the process in the visual and auditory perception. The signal processing using the autocorrelation and cross-correlation described in this dissertation is effective for describing periodical structure of the signal. The background of this correlation analysis is that periodicity is a salient cue in human perception. Information about the nature of a periodic phenomenon, such as its structure, strength, and frequency, is important for our understanding of the environment. The fundamental result that I will present is a demonstration that it is possible to apply the correlation mechanism to the process in vision and auditory system. Results are summarized below. The kinds of processing by visual and auditory sense have two aspects: temporal processing and spatial processing. It is generally believed that vision is especially good at spatial processing and audition

is good at temporal processing. But temporal information is important for visual processing and spatial information is important for auditory processing. In visual processing, detection of the moving objects is crucial for our life. Information of rhythm or tempo is a cue for perception of periodical events. As for auditory processing, information of direction and distance of sound source is important cue for us to navigate in the field. Spatial properties of diffuseness and source width are also important to define the quality of a sound field such as auditorium.

In Chapter 2, **Spatial vision**, I present a series of studies on physical properties and psychological evaluations of two-dimensional spatial pattern. I show that the autocorrelation function (ACF) analysis provides useful measures for representing three salient perceptual properties of texture, namely, contrast, coarseness, and regularity. Another experiment showed that the degree of regularity is a salient cue for texture preference judgment. Described ACF model offers the advantage of extracting perceptual properties and evaluating subjective reaction in texture perception.

In Chapter 3, **Temporal vision**, I discuss the underlying mechanism for the temporal perception in vision. To address the problem, I focus on the fundamental properties of the temporal vision mechanism. Psychophysical experiment was performed on subjective flicker rates for complex waveforms. Results showed that human observers perceived a rate at the fundamental frequency, although the energy at this frequency was not included in the signals. It implies the existence of correlation mechanism in temporal vision.

In Chapter 4, **Audition**, I present a series of studies on physical properties and psychological evaluations of sound. I used the ACF in analyzing a sound signal and the interaural cross-correlation function (IACF) to characterize the spatial properties of sound field. It was found that the acoustical properties are well represented by the factors extracted from the ACF and the IACF, and that the subjective evaluations for sound signal are explained by the combinations of the ACF factors.

As applications of the correlation model for signal processing, I propose two examples. One is an analyzing method of textural features for image retrieval and pattern recognition. Another is sound feature extraction for speech recognition technology. These techniques are using our ability for detecting periodical structure in visual and auditory signals.

Acknowledgment

I take this opportunity to express my thanks to the many people who have helped me, in various ways, to attain my objective of preparing this dissertation. Firstly I would like to express my gratitude for the considerable guidance and encouragement received from Professor Yoichi Ando, Kobe University, whose contribution to this research has been substantial. I also wish to express my appreciation to Professor Takaji Matsushima and Professor Shinichi Murakami, Kobe University.

I would also like to specially mention Associate Professor Shin-ichi Kita, Faculty of Letter at Kobe University. Work on temporal vision described in Chapter 3 is the result of collaborative effort with him. His remarkable insight into the experimental psychology and human visual system, and his willingness to speculate, discuss, and cooperation in paper writing with me, have enriched and my thesis.

I would like to express my thanks to Dr. Shin-ichi Sato, Dr. Hiroyuki Sakai for their many supports and invaluable advice especially about the acoustical measurement described in Chapter 4. I also acknowledge Masanobu Inoue, Ryota Shimokura, Toshihiro Kitamura, and Shoji Uetani for their assistance with the measurement. The data in Chapter 2 is the result of the effort by Shinofu Sugi and Toshiyuki Itano.

I would like to thank all members of Ando Laboratory, past and present. Especially I thank Junko Atagi, Kazi Saifuddin, Yoshiharu Soeta, Takashi Higano, and Takuya Hotehama for many productive suggestions and discussions. I also thank those who kindly participated in the experiments as subject.

Kenji Fujii

Contents

1. Introduction

Preface

Periodicity is common in the natural world. It is also a salient cue in human perception. Information about the nature of a periodic phenomenon, such as its structure, strength, and frequency, is important for our understanding of the environment. Humans are very good at identifying complex patterns. The auditory system easily senses complex periodicities such as the rhythm and tempo that occur usually in music and speech, the visual system readily grasps the symmetries and repetitions inherent in tile patterns, and the mind searches for simple regularities to explain phenomena that appear complex and irregular. Undoubtedly humans have the mechanism detecting periodic phenomenon. This mechanism is probably common in all the sense we have. Among our five senses, it is generally believed that vision is especially good at spatial processing and audition is good at temporal processing. But temporal information is important for visual processing and spatial information is important for auditory processing. In this dissertation I explore the underlying mechanism for periodicity detection in vision and audition involving space and time. I show that a periodic phenomenon is one of the bases of more complex subjective attributes in visual and auditory sensations.

1.1 Motivation

I started this research project motivated by the desire to understand primal attributes in vision and auditory sensation gathering toward complex psychological responses and evaluations in the real environment. As psychological responses mentioned here, subjective preference and subjective impressions of sounds, visual properties of landscape and interior space are supposed. To evaluate qualities of such materials from psychological or affective viewpoint, relationship between physical property and subjective evaluation need to be cleared. Related to these topics, there is an emerging research area called "KANSEI Information Processing" from Japan. KANSEI is a Japanese word that does not have a direct counterpart in Western languages. The concept of KANSEI is strongly tied to the concept of personality and affective sensibility. KANSEI is an ability that allows humans to solve problems and process

information in personal way. It can be related to emotion, but also refers to the human ability of information processing in ways not just logical.

The problem in understanding KANSEI by a scientific manner is that this kind of information is difficult to express in language, therefore to definition and quantification. It is easy to understand as we experience it everyday in sights and sounds, but this is rather vague to describe these properties. The position I take in this dissertation is that such vague and complicated sensibilities occur as assemblies of subjective attributes processed in vision, audition, and so on. These attributes include color, shape, and texture of the objects, or loudness, pitch, and timbre of sound. Thus we need to know the way that human process information about these subjective attributes. Modeling of the visual and auditory processing and understanding the relationship between physical and subjective property is therefore the main topic in this dissertation.

We also need to know what is a salient cue for subjective evaluations among the physical and subjective primal attributes. As Gestalists revealed, there are wholes, the behavior of which is not determined by that of their individual elements, but where the part-processes are themselves determined by the intrinsic nature of the whole. Our sensation and subjective evaluation might be the Gestalt itself. It is possible that our visual and auditory system grasp the whole information of sight and sound roughly by characterizing only several important properties. I explore what are salient cues in perception of complex visual pattern and acoustic signals, and how do human process such complex patterns.

1.2 Organization

This dissertation is organized by following five chapters. In Chapter 2, 3, and 4, I discuss three different properties of human perception in vision and audition. As shown in Table 1.1, the kinds of processing by visual and auditory sense have two aspects: temporal processing and spatial processing. It is generally believed that vision is especially good at spatial processing and audition is good at temporal processing. But temporal information is important for visual processing and spatial information is important for auditory processing. In visual processing, detection of the moving objects is crucial for our life. Information of rhythm or tempo is a cue for perception of periodical events. As for auditory processing, information of direction and distance of sound source is important cue for us to navigate in the field. Spatial properties of

diffuseness and source width are also important to define the quality of a sound field such as auditorium. To assess the psychological evaluation of sound in the real environment, it is necessary to investigate such spatial sensations as well as temporally processed primary sensations. Contents in each chapter are summarized as follows.

Table 1.1 Spatial and temporal processing in vision and audition.

	Spatial	Temporal
Vision	Shape	**Tempo**, Rhythm
	Depth	Motion
	Pattern	
Audition	**Location**	**Loudness, Pitch, Timbre**
	Spaciousness	Tempo, Rhythm

Chapter 2, **Spatial vision**, discusses about spatial properties of visual perception. I present a series of experiments on physical properties and psychological evaluations of two-dimensional spatial pattern. A simple autocorrelation model that provides useful measures for representing salient perceptual properties of texture is introduced. Then I discuss the supposed correlation mechanism of the visual system for texture perception.

In Chapter 3, **Temporal vision**, I discuss the underlying mechanism for the temporal perception in vision. Temporal processing in visual system has long been neglected in spite of its importance in human perception. To focus on the fundamental properties of the temporal vision mechanism, the subjective flicker rate of complex waveforms is examined. Later I discuss the possibility of correlation mechanism in temporal vision.

In Chapter 4, **Audition**, I present a series of studies on physical properties and psychological evaluations of sound. I use a method of autocorrelation and inter-aural cross-correlation in analyzing a sound signal. These studies are based on the previous knowledge on correlation mechanism in the auditory system. I try to characterize the primal acoustical properties of sound field and to understand the mechanism of perception about such properties.

In Chapter 5, **Applications**, image processing and speech recognition techniques are proposed. Proposed methods are application of human ability to extract

signal characteristics in vision and audition.

Finally in Chapter 6, **Conclusions**, summary of this dissertation, contributions and future directions are described.

2. Spatial vision

2.1 Introduction

In this chapter, I discuss about spatial properties of visual perception. For many practical reasons, a large part of research in spatial vision is performed with black & white or simple gray scale patterns. Although they can shed light on certain aspects of early vision mechanism, these are highly impoverished stimuli compared to the natural visual environment. To explore more sophisticated mechanism of the visual system, I focus my research interest on perception of complex spatial pattern, texture.

The term "texture" is used in a variety of meaning and context. Not only as a visual texture, but also textile surface touch, sound quality, and food material refer to texture. There are many definitions even in a visual texture. Common nature of texture according to these descriptions is that texture is intuitive to understand, but somewhat difficult to define. The only ground truth is what human observers perceive to be the kind of textural properties of the image content of that signal. This places texture and spatial pattern into the domain of perceptual attributes of image. A pressing question for psychological and image-processing researchers is to discover physical properties that correlate with this perceptual attribute.

Also, texture is one of the most important cues for spatial perception such as depth and location of objects in the space. Since Gibson (1950), many efforts have given into texture perception to understand the mechanism of spatial perception. But how does our visual system extract and characterize textural information for further process of spatial perception is still an open problem. To understand the underlying mechanism for textural properties therefore contributes to the investigation of the high-level mechanism of the visual system.

The purpose of the study in this chapter is two folds. One is to extract the physical features corresponding to visual perception. This includes primal features of texture and psychological evaluation of texture. Another is to understand the mechanism of texture perception in our visual system. The chapter is organized as follows. In Section 2.2, I will briefly review a number of texture related studies in the field of psychophysics, engineering, and psychology. Then I will present an approach of texture

analysis based on the correlation model. Finally, I will discuss the supposed mechanism of the visual system for texture perception.

2.2 Related work

In this section I review previous works, which are related to this chapter. This contains the study of the low-level and high-level mechanism of the visual system, texture modeling in the image processing, and psychological dimension of texture perception.

2.2.1 Pattern perception and model of spatial vision

Psychophysical experiments have shown that certain texture types can be discriminated and segmented preattentively. Preattentive discrimination and segmentation often occur in textures with different first-order statistics (Figure 2.1). An analysis of a variety of textures led Julesz to believe that textures with the same "global" second-order statistics are indistinguishable from one another (Julesz 1962).

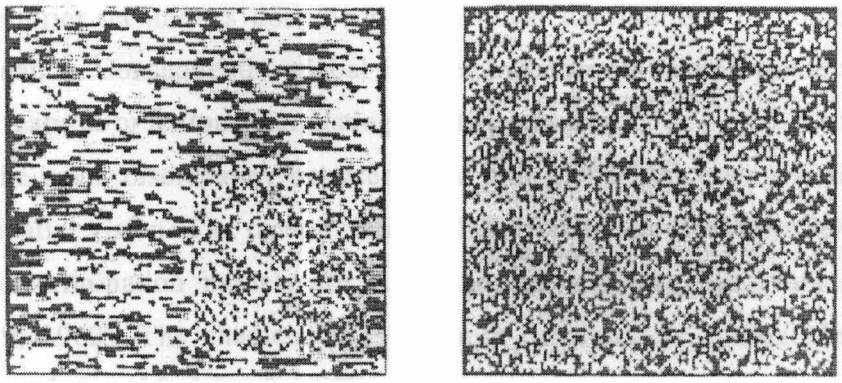

Figure 2.1 Julesz's random dot (Julesz, 1962). Two regions are preattentively discriminable when they have different second-order statistics (a), but not discriminable when they have same second-order statistics (b).

This finding became the basis of texture analysis in engineering. Later, he discovered evidence to the contrary (Julesz 1980). These were texture pairs with the same global second-order statistics, which remained easily discriminable. Each pair appeared to have some "local" differences in the conspicuous features, such as elongated blobs, orientation, length, width and line crossing (e.g., Figure 2.2). These features were called textons (Julesz 1983).

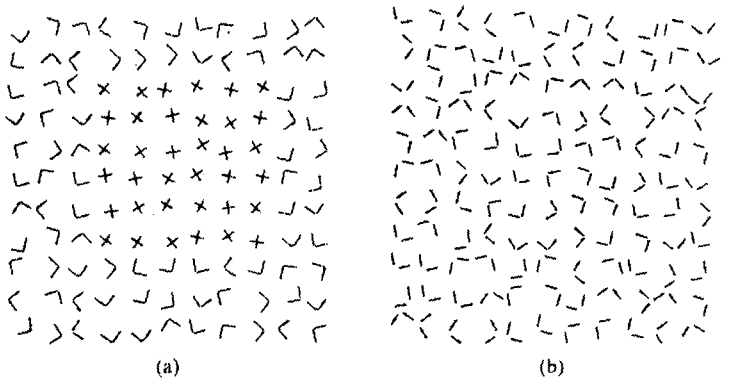

Figure 2.2 Example of texton (Julesz, 1983). (a) Texture pair of which elements having different textons (line crossing vs. not-crossing segments) is distinguishable, whereas (b) of which elements have identical textons yields indistinguishable texture pair.

Preattentive texture discrimination, in this case, occurs as a result of differences in texton types or some other first-order statistics between local features such as their density or standard deviation, rather than as global second-order differences. Analysis by global Fourier techniques does no t reveal local distributions or combinations of image luminance. Consequently, global Fourier transform is inappropriate for analysis of the local features or textons that distinguish two textures.

The same discrepancy between local and global processing is found in the models of spatial vision. The receptive field characteristics most frequently mentioned in texture discrimination models are elongated bar and edge detectors (Hubel and Wiesel 1959, 1962). The characteristics of these detectors are reasonable to explain Julesz' texton theory. Models of a different type claim that visual stimuli are processed in parallel by a number of independent spatial frequency channels (Campbell and Robson 1968). Extensive experimental evidence suggests that some form of spatial frequency analysis is performed in the striate cortex (DeValois et al. 1979). These findings lead to the view of the cortex as a kind of spatial Fourier analyzer.

These problems can be avoided by applying Fourier transforms to the windowed local regions corresponding to a specific spatial range within the visual field. Local distributions, such as textons, are revealed by the Fourier transforms in small regions, and global second-order statistics are characterized by the Fourier transforms in a relatively large area. A spatial/spectral analysis such as Wavelet or Gabor filtering, is a method for determining the optimal window in both spatial and spatial-frequency

domains (Bovik et al., 1990; Jain & Farrokhnia, 1991). The spatial/spectral analysis has proved to be effective in explaining to some extent the human texture discrimination (Figure 2.3). It was found that two textures are often difficult to discriminate when they produce a similar distribution of responses in a bank of spatial-frequency and orientation selective linear filters (Turner, 1986; Malik & Perona, 1990).

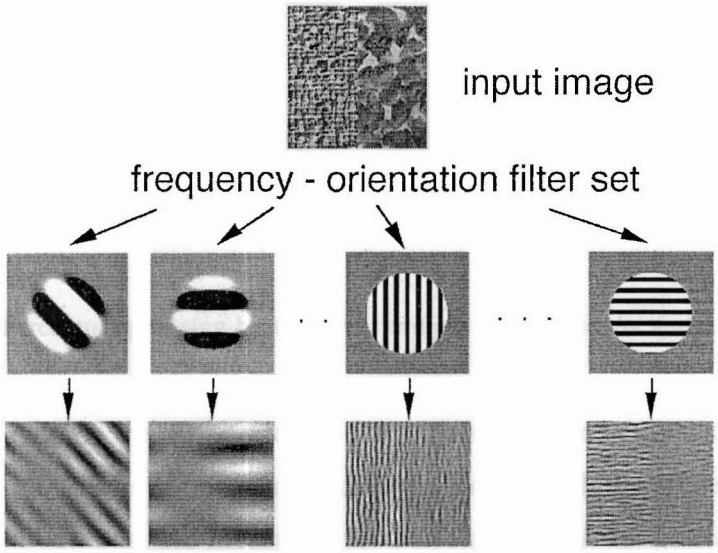

Figure 2.3 Example of a frequency and orientation filter bank. The difference of the distributions in each filter output is used for discrimination of two textured areas.

Perceptual grouping is a process in chunking of visual information and in image segmentation, consisting an important aspect of visual processing. However, the rules that govern grouping lack a quantitative formulation. According to the Gestalt theory of grouping, simple rules such as similarity of elements, proximity, good continuation, common fate and connectedness dominate perceptual grouping by segmenting a visual scene into regions having some internal consistency (Koffka, 1935). It is still unclear how to define shape and similarity and how to deal with multiple cues. An approach has been taken to model perceptual grouping on the basis of a similarity metric (Beck, 1966). That is, by grouping together elements that share common features. Ben-Av and Sagi (1995) presented a quantitative model for perceptual grouping, which was based on the intensity autocorrelation (Figure 2.4). The model performance was successfully compared with data from psychophysical experiments. Results suggested that at least some of the Gestalt rules of grouping (i.e. similarity and proximity) could

be formalized in terms of directionally estimated spatial correlations. They also proposed a possibility that oriental filters (e.g. Gabor filters) with long-range interactions between them can be used for estimations of directional autocorrelations.

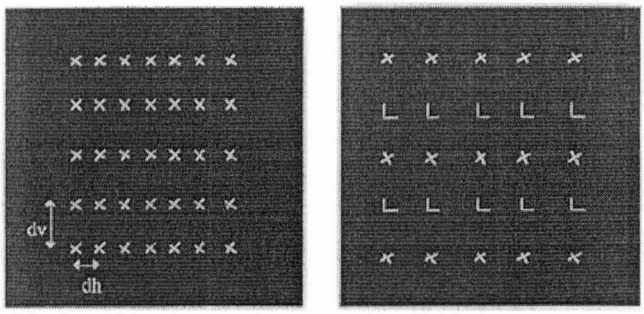

Figure 2.4 Perceptual grouping by similarity and proximity, Ben-Av and Sagi (1995).

Importance of the autocorrelation function has been emphasized by Uttal in his book (Uttal, 1975) on form perception. He conducted the experiment of form detection with dot patterns. He measured the detectability of dot patterns, which were masked by the noise patterns (Figure 2.5). It was found that the regularity of dots in a pattern increased the detectability of the pattern. In the autocorrelation function, the periodicity in the signal is emphasized and the random noise is diminished. Consequently, only the periodical signal is extracted and then perceived as a form. The autocorrelation model well explained the large body of the psychophysical data.

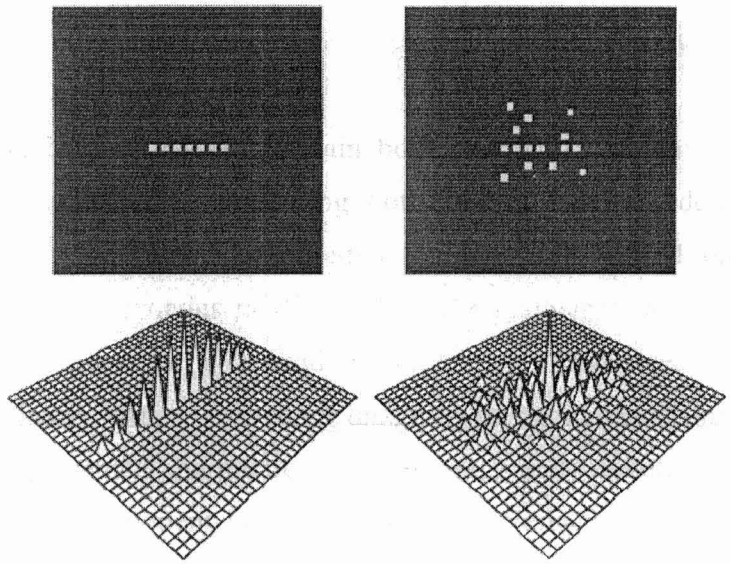

Figure 2.5 Uttal (1975), form perception by dot pattern and autocorrelation functions.

2.2.2 Texture modeling in engineering

Describing a texture by a small set of parameters is useful for effective compression and transmission of an image, because a textured region exhibit certain degree of homogeneity and can be regarded as containing a same or near same information over neighborhood. Historically, texture models are categorized into statistical, structural models (Haralick, 1979). Recently, new approach such as statistical/structural hybrid models and spatial/frequency (wavelet) models are studied (for review, see Wechsler, 1980).

The statistical approach focuses on the statistical properties of image patterns. This is the most natural and basic approach, because texture is not usually associated with identifiable objects. A texture pattern is characterized either by statistics of image pixel gray levels or by a stochastic model. Early methods include histogram, co-occurrence matrix and run length (Haralick et al, 1973). In the 1980's, a large number of literature appeared in the area of texture modeling using random field statistics (Cross & Jain, 1983; Mao & Jain, 1992). Models like the Markov random field or autoreggression models treat texture as a probability field. These models can describe texture by using a small number of parameters, but they are not based on the mechanism of the visual system and can hardly represent multiple textural properties corresponding to visual perception. The structural methods represent a texture pattern by its primitives and their

spatial placement rules (Haralick, 1979, Matsuyama et al, 1983). The main deficiency of the structural methods is that they are not capable of capturing the randomness that natural textures always exhibit.

Natural textures usually contain both structural and statistical components. Texture models capable of representing both structure and randomness have been studied. Francos et al. (1993) proposed a texture model based on the 2-D Wold decomposition of homogeneous random fields. The mathematical foundation of Wold-based texture modeling is the 2-D Wold decomposition of homogeneous random fields. The 2-D Wold theory allows a textured image to be decomposed into three mutually orthogonal components: harmonic, evanescent, and indeterministic (random) components. These component images can be characterized separately. Francos et al. applied Wold-based models to image coding and reconstruction. It was shown that a handful of model parameters could reconstruct natural textures that are visually indistinguishable from the originals (Figure 2.6). Liu and Picard (1996) constructed an image retrieval system based on the Wold theory. Compared to other models, the Wold model appeared to offer perceptually more satisfying results in the image retrieval experiments.

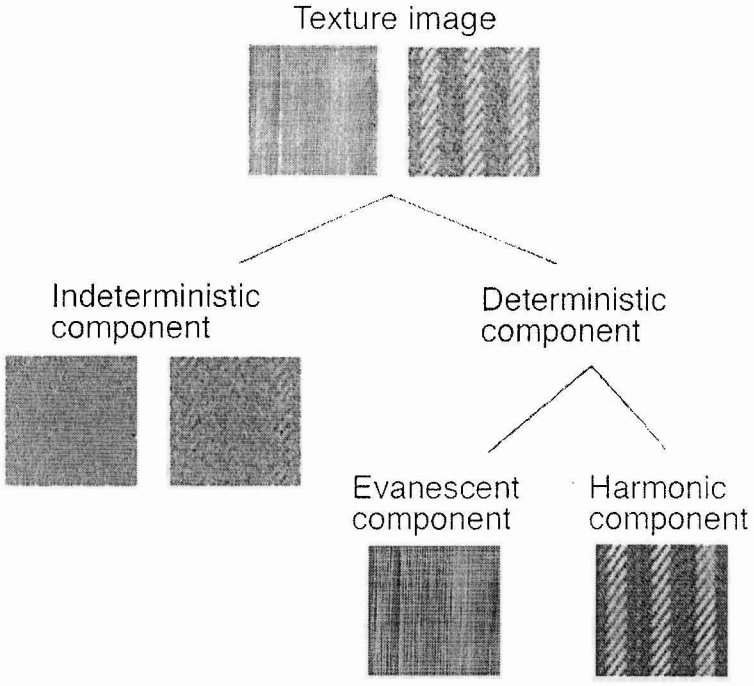

Figure 2.6 Examples of Wold decomposition by Liu and Picard (1996).

2.2.3 Perceptual properties and affective evaluation of texture

For constructing mature machine vision and computer interface, it is important that the computational measurements of texture correspond to the multidimensional perceptual properties well. Here I review some studies on perceptual properties of texture in the psychological framework.

From the descriptions seen in the past literature and from observation of textures in Brodatz's photographic album (Brodatz, 1966), Tamura et al. chose six properties of visual texture: coarseness, contrast, directionality, line-likeness, regularity, and roughness, to model computational features of textured image (Tamura et al., 1978). These properties were supposed to be common to all visual textures and have both extremes in the concept of each feature, e.g., coarse versus fine for coarseness and regular versus irregular for regularity, and so on. Then psychological experiments were conducted to establish the ordering of 16 texture samples based on the six textural properties. To improvise computational features for the six properties, they tested and modified heuristic features proposed in the literature as well as composing new ones. The final computational features were chosen as that provided the highest correlation between the computer and human ordering. In line with the direction of Tamura et al., Amandasun and King (1989) proposed five computational features corresponding to textural properties of coarseness, contrast, busyness, complexity, and texture strength. They also composed computational feature heuristically from the absolute differences between the gray scale value of each pixel and the averaged gray scale value in a neighborhood surrounding area.

One common problem in Tamura et al.'s and Amadasun and King's work is that no strong reasons were given why these properties were chosen. Although their proposed features seem to be characteristic of natural textures, it is not clear what the relative importance of there features are and how they span the perceptual space of human texture perception. Another problem is that the computational features were improvised heuristically for the individual properties. It is not clear how these features can be used together to represent a texture pattern. As for the former problem, Rao and Lohse (1995) conducted a psychological study to identify the relevant dimensions of human texture perception. In their experiment, twenty subjects rated 56 pictures from the Brodatz album on 9-point scales labeled by twelve adjectives such as repetitive, directional, random, granular, uniform, regular, etc. Then the subjects were asked to sort

the pictures into groups of similar ones. The rating data was analyzed by using classification and regression tree analysis, discriminant analysis, and principle component analysis. Also, the grouping data was analyzed by using hierarchical cluster analysis and non-parametric multidimensional scaling (MDS). Combining the analysis results of both rating and grouping data, the top three dimensions of texture perception were identified. These dimensions are shown in Figure 2.7. They presented the possibility that any kind of natural texture could be characterized by the combinations of features in their 3D space.

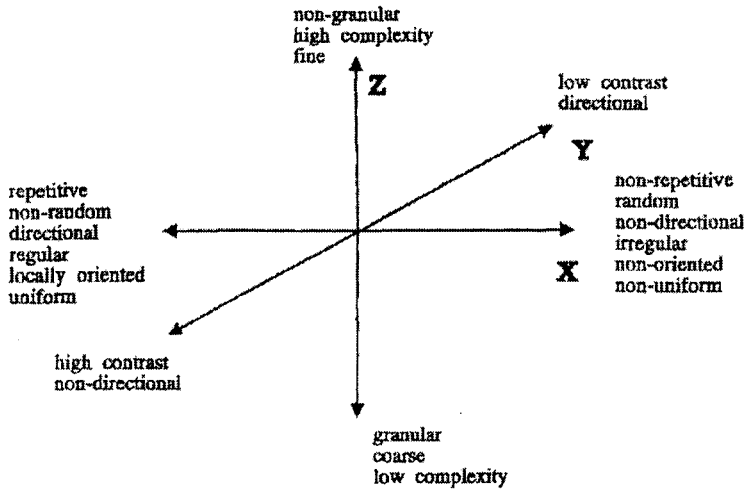

Figure 2.7 Rao and Lohse (1996), the three dimensions of texture perception.

2.3 Textural properties corresponding to visual perception

Summary

I show that the autocorrelation function (ACF) analysis provides useful measures for representing three salient perceptual properties of texture, namely, contrast, coarseness, and regularity. The validity of the ACF analysis was examined by comparing the calculated factors to the subjective scores collected for various kinds of natural textures. The effectiveness of the analysis depends on the structure of the estimated ACF. When a texture has a harmonic structure, the estimated ACF has periodical peaks corresponding to the periods of the texture. Both perceived coarseness and regularity are strongly related to these peaks in the ACF. As for the random texture, however, the estimated ACF does not have a periodical structure. In this case, the decay rate of the ACF can represent the texture coarseness and regularity.

2.3.1 Introduction

In this section, I present a new approach toward measuring textural properties corresponding to visual perception. Previously a number of studies have focused on textural properties through both perception and computational models. But the relationship between perceptual and computational properties remains unclear. To be used in machine vision or a computer interface, the computational properties of texture must correspond to the perceptual properties well. To understand the mechanism of our visual system, it is important to know how we extract and characterize the information to perceive texture. Therefore, I believe that this study contributes both to the research of image processing and human vision.

Tamura et al. (1978) described six features corresponding to visual perception, namely, contrast, coarseness, regularity, roughness, directionality, and line-likeness. They measured human perception in terms of these features and compared them with their computational results. Although their study significantly contributed to the texture analysis, their approach had drawbacks. They used already developed features only modifying a given feature and combining 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. Amadasun and King (1989) carried out a similar study, but their computational features had similar drawbacks.

A number of studies have attempted to tackle the problem from a different perspective focusing on the dimensionality of texture perception. Rao and Lohse (1996) developed a classification method for visual texture. Based on psychological similarity judgments, they constructed a three-dimensional space for texture classification. The three orthogonal dimensions they identified were repetitive vs. non-repetitive; high-contrast and non-directional vs. low-contrast and directional; and granular, coarse, and low-complexity vs. non-granular, fine, and high-complexity. An experiment conducted by Cho et al. (2000) suggests that the dimensionality of perceptual texture space is at least four. They described four orthogonal attributes, namely, coarseness, regularity, contrast, and lightness. These studies suggest that the important properties of perceptual texture can be described by using three or four independent factors.

I analyzed textural properties based on the correlation mechanism in the visual system. Uttal (1975) conducted an experiment of form detection with dot patterns. He measured 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 the random noise is minimized. As a result, only the periodical signal is extracted and then perceived as a form. Ben-AV and Sagi (1995) developed an autocorrelation function (ACF) model for perceptual grouping. They used matrix patterns to quantify the grouping law. The ACF was calculated in two directions (vertical and horizontal). The grouping occurred in the direction with a higher correlation. Their results imply that there is a correlation mechanism involved in the process of perceptual grouping. They suggest that the early local spatial filtering stage (tuned to specific frequency and orientation) may be followed by a long-range interaction between the filters to estimate the ACF of the input image.

The ACF is mathematically equivalent to the power spectrum through Fourier transforms, which means that we can derive the same information from both the ACF and power spectrum. For example, a vertical stripe pattern has its fundamental Fourier components along the horizontal axes. The ACF of the same pattern has a periodicity corresponding to the reciprocal of its fundamental frequency. However, there is a certain case, in which there is a discrepancy between the spectrum and the ACF. If a pattern does not have a fundamental component (missing fundamental: MF), we can also see the period of the fundamental (Henning et al., 1975). This suggests that our visual system can detect the spatial periodicity of patterns without Fourier components. A spectrum analysis failed to explain this perceptual phenomenon because the fundamental component could not be detected. Instead, the ACF of this MF pattern was found to have a periodicity of a fundamental component. Although the perceptual mechanism of the MF pattern is still under discussion (Badcock & Derrington, 1989; Hammett & Smith, 1994), this phenomenon implies that there may be a correlation mechanism in our visual system.

In this study, I tried to develop a set of measures for texture properties that corresponding to visual perception. Motivated by the discussion described above, I assumed that an ACF mechanism does exist in the visual system. I analyzed natural textures and extracted several features by using the ACF analysis. Then I compared our psychological judgments and computational measurements to extract and quantify the textural properties of contrast, coarseness, and regularity. These three properties were

chosen because of their importance to the study of texture perception described above.

2.3.2 Autocorrelation analysis of natural textures

In this section, we describe mathematical formulation of the ACF and extracted factors, which we compared with human perception. To show the validity of our analysis, previous psychological findings are also considered.

Definition of ACF

The ACF of a two-dimensional texture pattern is defined as

$$\Phi(\Delta x,\Delta y) = \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} p(x,y)p(x+\Delta x,y+\Delta y)/MN, \tag{2.1}$$

where p(x, y) is the input signal and p(x+Δx, y+Δy) is the shifted version of the input. The analyzed image was zero-meaned before the calculation to remove the DC components. "M" and "N" refer to the signal size in horizontal and vertical directions. In this study, M = N = 256. Equation (1) is the convolution of the signal itself. For computational efficiency, the ACF was computed as an inverse FFT of the image power spectrum. Usually, the ACF is normalized as

$$\phi(\Delta x,\Delta y) = \Phi(\Delta x,\Delta y)/(\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} p(x,y)^2/MN). \tag{2.2}$$

The denominator in equation (2) is the maximum value of the ACF and is defined as Φ(0). After this normalization, the ACF takes the maximum value of 1 at the origin.

Factors extracted from the ACF

From the calculated ACF, we extracted four typical features. Here we only considered a part of one quadrant because the ACF is symmetrical. As shown in Figure 2.8, the ACF decays from the origin to the outwards. We assumed that the properties of the two-dimensional ACF are held in the one-dimensional ACFs. To simplify the calculation algorithm, we considered two one-dimensional ACFs for the x and y directions from the origin. When the ACF had a directionality, we chose the periodical one, and when both directions had a periodicity, we chose the one in which the maximum peak had a higher amplitude. By this manipulation, the periodical nature of the ACF was extracted.

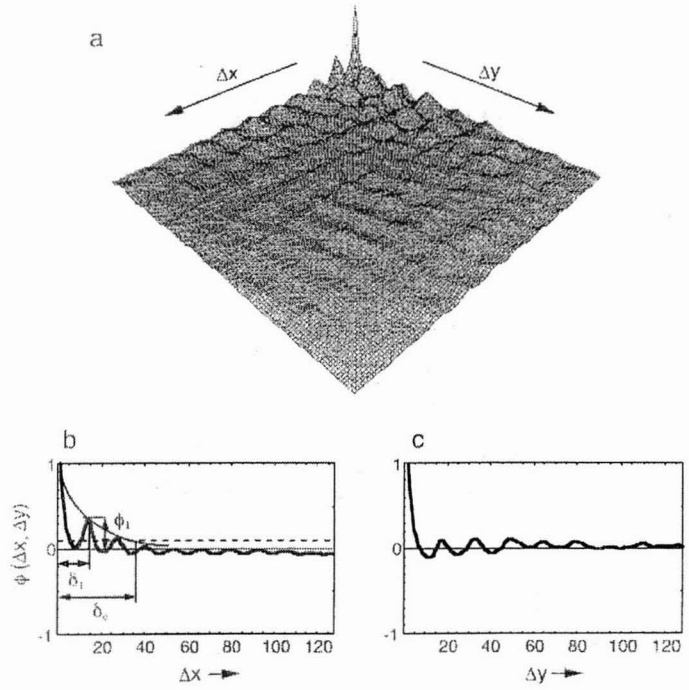

Figure 2.8 A calculated ACF and factor definitions. (a) Two-dimensional ACF of texture D3 in Figure 2.9. (b) and (c) One-dimensional ACF for the x and y directions. By comparing the heights of the maximum peaks in both directions, we extracted three factors from the ACF for the x direction.

Below, we define and describe the four factors. (1) $\Phi(0)$: the autocorrelation at shift value $\Delta x = \Delta y = 0$. Because the ACF was calculated from the zero-meaned image, $\Phi(0)$ corresponds to the root-mean-square (RMS) of the signal. Thus, it is assumed that $\Phi(0)$ corresponds to the perceived contrast. Only this factor was calculated from the two-dimensional ACF, and it is shown on the dB scale ($10 \log_{10}\Phi(0)$). (2) δ_1: the displacement of the maximum peak and (3) ϕ_1: the amplitude of the maximum peak in the normalized ACF. These two factors are related to the periodicity of the image. The value of δ_1 is a reciprocal of the fundamental frequency, and ϕ_1 is the strength of the harmonic components. It is possible that the coarseness and regularity can be represented by δ_1 and ϕ_1 respectively. (4) δ_e: the effective range of the ACF. We defined this value as a displacement at which the normalized ACF decayed below 0.1. Because its value is related to the period and height of the peaks, δ_e may be related to the perceived coarseness and regularity. The values of δ_1 and δ_e were normalized based on the image size. The visual angle (degree) of the values is calculated by applying image width (cm) and dividing by viewing distance (cm).

2.3.3 Preliminary experiment

Data collection

As a preliminary experiment, we analyzed 16 textures from the "Brodatz texture database" (Brodatz, 1966), that were used by Tamura et al. (1978). This texture set includes a variety of natural and man-made objects (Figure 2.9). The analyzed data consists of 256 × 256 images with 8-bit (256) gray levels. The ACF was calculated and four factors were extracted from each image. As mentioned before, Tamura et al. measured human perception about six textural features including contrast, coarseness, and regularity, the features we are interested in. We compared their results with our computed results. A psychological data set was obtained from the figure in Tamura et al.'s article by using an image scanner, and their subjects' average scores were re-scaled between 0 and 1. The subjective scores for the contrast, coarseness, and regularity were correlated with the factors extracted from the ACF, namely, $\Phi(0)$, δ_1, and ϕ_1.

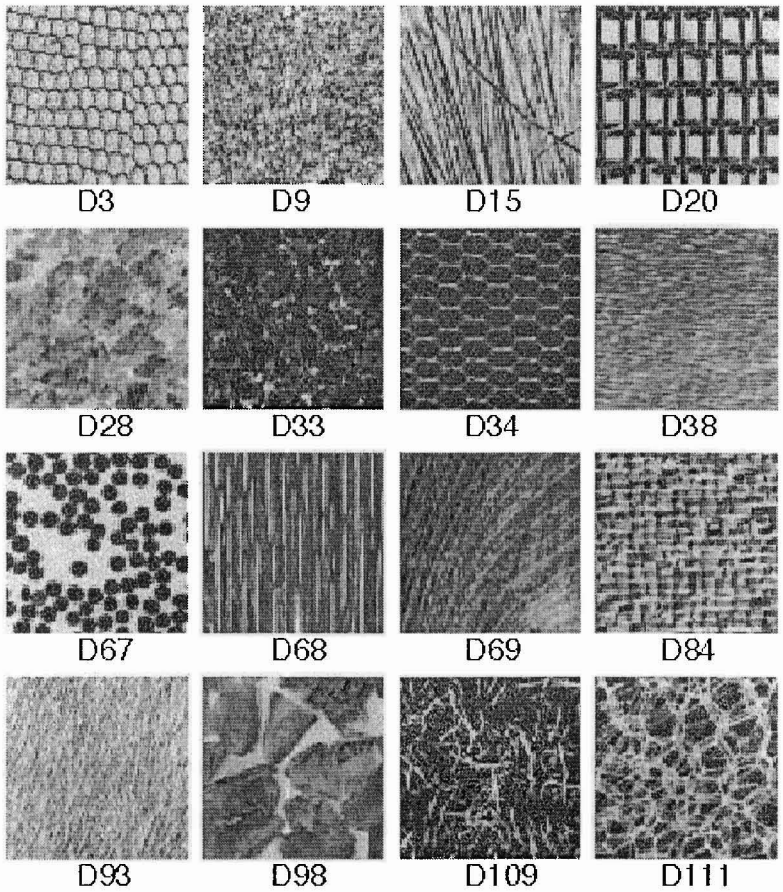

Figure 2.9 Texture set used by Tamura et al., (1978).

Comparing psychological measurements and ACF factors

Figure 2.10 shows the relationship between the ACF factors and subjective scores. The ACF factors are shown on a logarithmic scale because of its goodness for linear regression. We can see that our ACF factors are in good agreement with the subjective scores. For contrast, in particular, correlation coefficient was very high ($r = 0.89$, $p < 0.01$) and we concluded that these factors correspond to the perceptual property of contrast well. As described above, ACF factor $\Phi(0)$ corresponds to the RMS of the image. This is consistent with the previous studies that found that textures with the same RMS are perceived as the same contrast (Mayhew & Frisby, 1978; Tiippana et al., 1994).

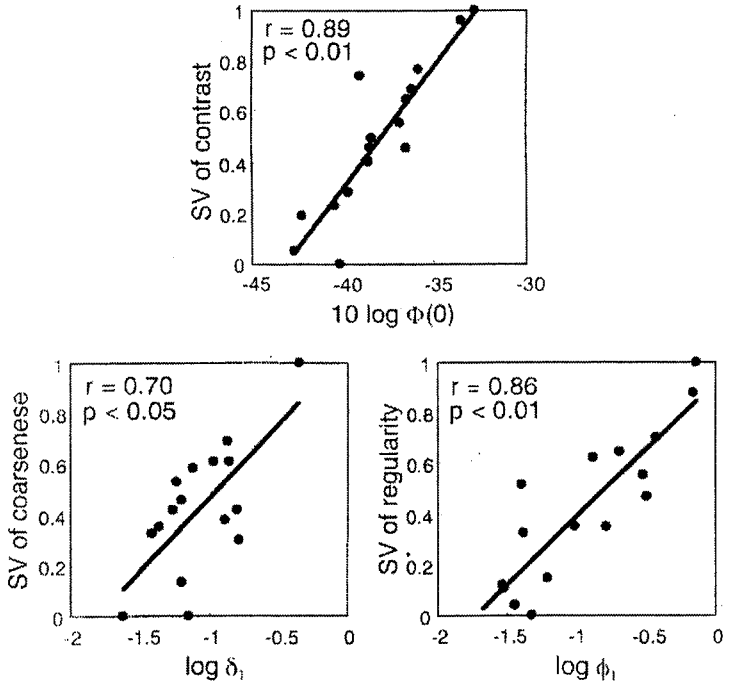

Figure 2.10 Relationships between the scale values of textural properties and corresponding ACF factors. Each dot represents a scale value for each texture, and the straight lines express linear regression.

The correlation between the SV of regularity and the value of ϕ_1 was also high ($r = 0.86$, $p < 0.01$). Particularly for the high value of ϕ_1, perceived regularity was well represented. Some of the textures with values of log (ϕ_1) below -1 (0.1 in real value) showed some discrepancies. A low value of ϕ_1 could mean an irregularity or a randomness of the texture. But the correlation only in this range was very low ($r = 0.35$).

This result implies that regular-random dimension can hardly be represented by the value of ϕ_1 alone, especially in random texture.

The correlation between coarseness and δ_1 was 0.70, which is worse than the previous two results. There was a discrepancy for some textures in which ϕ_1 had a low value. This is because δ_1 is strongly affected by the value of ϕ_1. When the ACF contains high peaks because of a high regularity of the image, ϕ_1 and δ_1 can be easily determined and the resultant value of δ_1 accurately represents the dominant period of the image. But when the image is random, the ACF decays immediately with no particular peaks. As a result, the determined value of δ_1 might not reflect the period of the image.

We showed that the ACF factors could represent the perceptual properties of contrast, coarseness, and regularity. But there remains some variance that cannot be explained by single factors alone. A portion of this unexplained variance might be due to the properties not being truly one-dimensional. Actually some correlation between the perceived coarseness and contrast was found by Tamura et al. We used multiple regression analysis to examine whether the plural factors affect the perceptual properties.

Multiple regression analysis

The aim of the multiple regression analysis is to predict the behavior of a dependent variable by using a weighted sum of independent predictor variables. In our case, we tried to predict the subjective scores by using a linear combination of four ACF factors. To obtain an optimal model for any given number of predictors, all 11 possible regression equations were examined. The correlation coefficients and significance levels were used to determine the goodness of fit.

For contrast and coarseness, the best predictors were $\Phi(0)$ and δ_1. Although the other factors also had some use, they were not significant. Comparing the results, we can see that the variance unexplained by single factor shown in Figure 2.10 became small in Figure 2.11. The partial regression coefficients of $\Phi(0)$ and δ_1 were 0.98 and 0.36 for the contrast, 0.50 and 0.79 for the coarseness. It means that the perceived contrast and coarseness are represented mainly by $\Phi(0)$ and δ_1, and may be affected by another factor. This result is consistent with the fact that there was a correlation between perceived contrast and coarseness. For regularity, only single factor ϕ_1 was significant in all combinations. Even though the other factors also had some contribution, the resultant correlation little gained with these factors. Therefore, we concluded that in this

experiment, single factor ϕ_1 was sufficient to represent the perceived regularity.

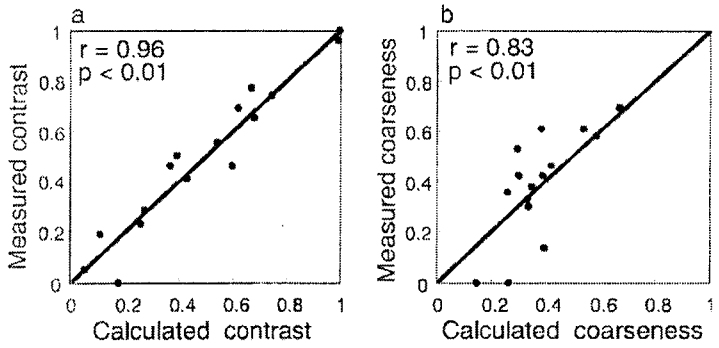

Figure 2.11 Relationships between calculated SVs by means of multiple regression analysis and measured SVs of contrast and coarseness.

Discussion

Based on the results described above, we believe that the ACF factors can be used to represent the perceptual properties of contrast, coarseness, and regularity. The correlation coefficients showed similar or higher levels of correspondence compared with those found in Tamura et al.'s original results. It is quite surprising that these simple measures account so well for the subjective attributes of complex natural texture. The remaining problem is how to improve the correspondence of regularity and coarseness for random texture.

As described above, the ACF of random texture does not have a periodical property. Consequently, it is difficult to extract information related to regularity and coarseness only from the maximum peak of the ACF. A recent study reported that human visual system might use different cues in texture perception in proportion to the strength of harmonic components (Sakai & Finkel, 1995, 1997). In the case of random textures, the value of δ_e (the effective range of the ACF, defined as a ten-percent decay of the ACF) may represent the perceived regularity and coarseness. If an image does not contain any harmonic components, the value of δ_e corresponds to the slope of its power spectrum. When the spectrum is flat (white noise), the value of δ_e becomes very small (theoretically, $\delta_e \rightarrow 0$). When the spectrum has a slope (correlated noise), the value of δ_e becomes larger in proportion to the value of the slope (informal observation). It is possible that the more the noise-like images correlate, the more we can perceive the

regularity of the images. In coarse texture, δ_e is large, whereas in fine texture, δ_e is small. A detailed discussion about this issue is presented in the next section on the psychological experiment. In our experiment, we used a set of random-looking textures to examine whether the value of δ_e could represent the perceived coarseness and regularity.

2.3.4 Psychological experiment

Methods

Ten subjects (5 males and 5 females) participated in the experiment. Their ages ranged from 22 to 26. All had normal or corrected-to-normal visual acuity. Except for two of the authors (FK and SS), the rest of the subjects were unaware of the purpose of the study.

The stimuli were presented on a CRT display under normal indoor lighting conditions. The stimuli consisted of a series of two texture pairs located horizontally. The viewing distance was approx. 100 cm, which result in each texture being subtended at a visual angle of 6 × 6 degree. The sample texture set used in the experiment consisted of 12 botanical textures as shown in Figure 2.12. The pictures were taken by a digital camera and were transformed into 256 × 256 images with 8-bit gray levels. All the samples were then analyzed by using the ACF in the same way as described in the previous section, and four factors were extracted. Compared with Tamura et al.'s texture set, our samples had a higher degree of similarity to one another. Consequently, the extracted factors fell into a narrow range (Table 2.1).

Table 2.1. ACF factors for textures used in Tamura's experiment and ours. Mean values are shown with standard deviation.

Sample set	$\Phi(0)$	ϕ_1	δ_1	δ_c
Tamura et. al	-37.99 (2.78)	0.21 (0.23)	0.11 (0.1)	0.19 (0.36)
Present study	-39.08 (1.61)	0.09 (0.05)	0.03 (0.01)	0.02 (0.01)

Furthermore, as we can see from Table 2.1, the values of ϕ_1 were much lower for our stimuli. This means that the structure of our sample textures was not harmonic but more random. The reason why we chose such a texture set is twofold. On the one hand, we wanted to examine whether the human subjects could detect small differences

in the properties and whether the results of the ACF analysis corresponded to the subjects' judgments. On the other hand, we wanted to examine whether the ACF factors are useful in analyzing textures with less harmonic structures.

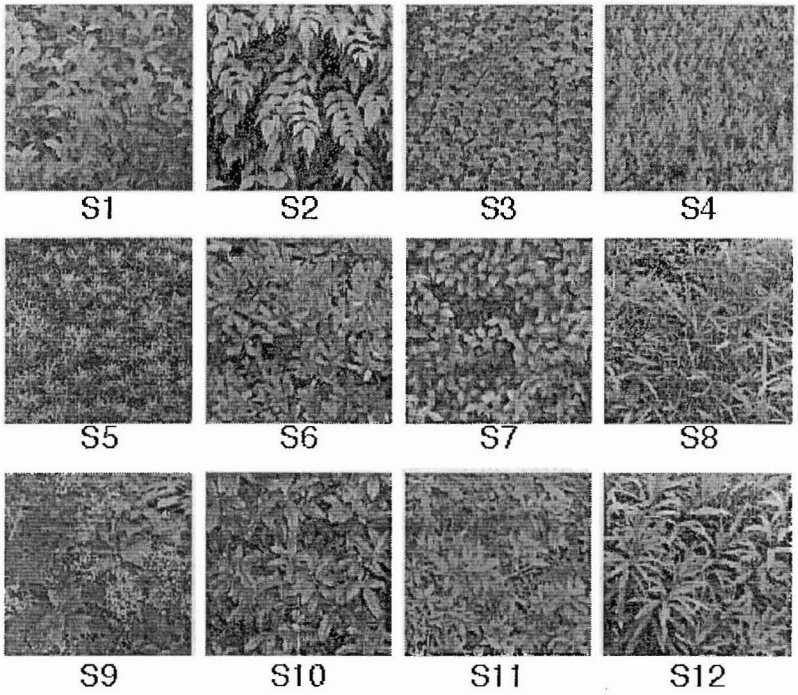

Figure 2.12 Stimulus set used in the psychological experiment.

Before the experiment, the subjects were given a brief explanation of the basic concept of texture and the three features described above (the explanations were written in Japanese, except for the names of features such as "coarseness; coarse *vs.* fine", which were given in English). The method of judgment in our study was a paired comparison. All possible pairs from the twelve images (66 pairs) were presented to the human subjects in a random order in one session. During a presentation of 10 s, they had to make their decisions about the three features, i.e., to choose from each pair the pattern that was coarser, had a higher contrast, and was more regular. All subjects had eight series of sessions, giving a total of 528 comparisons for each feature.

Data analysis and Results

We processed the experimental results by applying "the law of comparative judgment" (case V; Thurstone, 1927). This law was used to produce a one-dimensional scale value (SV) for each stimulus from the total matrix of superiorities collected from the paired comparisons. The results were reconfirmed by the goodness of fit (Mosteller, 1951), and

the agreement of all subjects' judgments was tested by the chi-square test ($p < 0.05$). In Table 2.2, we show the correlation coefficients between the subjective scores for each feature. Based on the results, we can make the following observations.

Table 2.2. Correlation matrix of the scale values of contrast, coarseness, and regularity.

Property	contrast	coarseness	regularity
contrast	-		
coarseness	0.672*	-	
regularity	0.017	-0.363	-

*$p < 0.05$

a) For all perceptual properties tested in the experiment, the subjects' judgments are reliable and there are certain underlying criteria upon which the tested subjects agreed. This does not mean that the same data were obtained for all human subjects in the experiment. Individual differences must be considered.

b) Especially in the regularity score, a relatively high inter-subject variance was observed. There might be several reasons for this variance. First, some of our subjects may have interpreted "regularity" differently, or the concept of regularity itself might have been difficult to define for the tested textures. Second, because there was little difference in material or appearance between the sample textures, the subjects may have had a difficulty in comparing regularity. Finally each subject might have used different cues for the judgment of regularity. We specified regularity as a property of the placement rules, but the variation of elements, especially in the case of natural textures, may reduce the regularity as a whole. As a result, it is assumed that the subjects perceived a high regularity of those textures for which we can describe texture elements easily. Additionally, fine texture tends to be perceived as regular (there is a slight inverse correlation between regularity and coarseness, see Table 2.2).

c) There was a high correlation between the SVs of contrast and coarseness ($p < 0.05$). This means that more coarse texture is perceived to have a higher contrast. The coefficient value was consistent with Tamura et al.'s results.

From the above observations, we conclude that our experiment was valid for measuring human perception of textural properties, even though there were some individual differences. The psychological data and computed features of the textures are

compared described as described below.

Comparing psychological measurements and ACF factors

The relationship between each SV and its corresponding ACF factor is shown in Figure 2.13. There is a high correlation ($r = 0.81$, $p < 0.05$) between the SV of contrast and $\Phi(0)$. This result is consistent with that of the preliminary experiment and is satisfactory.

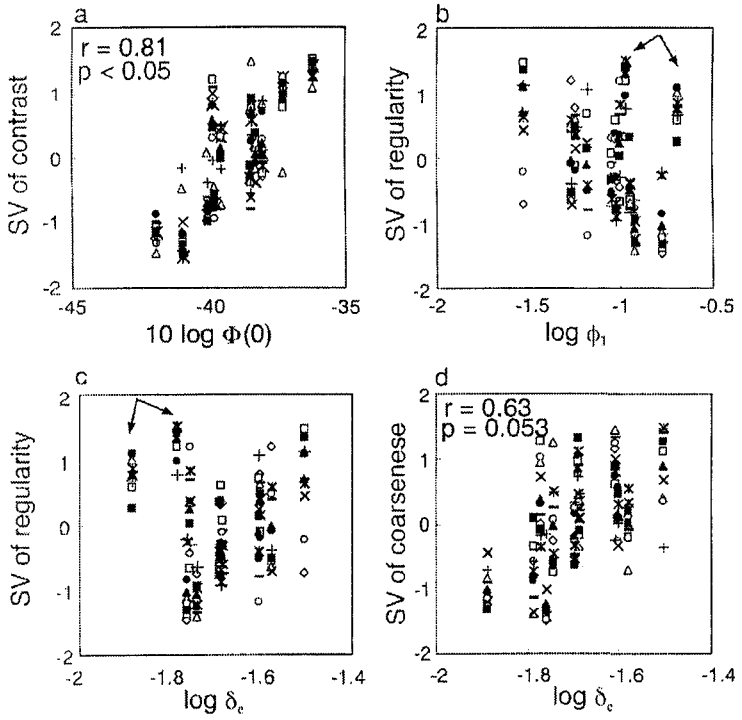

Figure 2.13 Relationships between the scale values and ACF factors. (a) SV of contrast and $\Phi(0)$. (b) and (c) SV of regularity and ϕ_1, and δ_c, respectively. (d) SV of coarseness and δ_c. Arrows in (b) and (c) indicate two textures mentioned in the text.

As for regularity, our result was very different from that obtained in the preliminary observation. In that experiment, we hypothesized that regular textures had high-amplitude periodical peaks in the ACF and that the height of the maximum peak corresponded to the perceived regularity. Our preliminary observations confirmed this hypothesis clearly. In our psychological experiment, however, we could not obtain similar results. As shown in Figure 2.13 (b), the relationship between the SV of regularity and the value of ϕ_1 is unexpected. For the range of log ϕ_1 below -1 (0.1 in real

value), there appeared to be a tendency toward a negative correlation. This result is perhaps related to the fact that most of the textures in our stimuli had a low value of ϕ_1. As a result, the maximum peak in the ACF might not represent the periodicity of the image, resulting in unexpected values of ϕ_1. The reason for this negative correlation is not clear. To represent the perceived regularity for these textures, we can use the value of δ_e. We can see that there is a normal correlation between the regularity and δ_e with two exceptional cases (Figure 2.13 (c)). These two textures are also indicated in Figure 2.13 (b), which shows that these textures have large values of ϕ_1. This result can be interpreted as follows. When the value of ϕ_1 is low, δ_e reflects the perceived regularity. When ϕ_1 exceeds a certain level, however, the perceived regularity might be affected by the harmonic components included in the texture. It is possible that our subjects perceived the regularity in proportion to the value of δ_e for the textures with less harmonic structures, and to the value of ϕ_1 for the textures with more harmonic structures.

In Figure 2.13 (d), the result for coarseness is shown in relation to δ_e. There is a moderate correlation even allowing some variance ($r = 0.63$, $p = 0.053$). We assumed that the perceived coarseness is related to the value of δ_1, but the correlation between δ_1 and SV was not significant ($r = 0.56$, $p = 0.28$). As described earlier, most of our stimuli did not have a harmonic structure. As a result, the determined value of δ_1 may not necessarily reflect the periodicity of the image. The value of δ_e appears more suitable to represent the perceived coarseness.

Multiple regression analysis

As in our preliminary experiment, we found a correlation between the SVs (see Table 2.2). Plural factors may be suitable to represent perceptual properties. We conducted a multiple regression analysis again for the psychological data we obtained.

We found that the factors of $\Phi(0)$ and δ_e were significant predictors of the SVs of contrast and coarseness (Figure 2.14). The regression coefficients showed that the SV of contrast is represented mainly by $\Phi(0)$. In our experiment, predictor δ_1 in the preliminary experiment was replaced by δ_e, which is a factor representing the perceived coarseness. Considering the two experimental results together, it becomes obvious that the values of the perceived contrast and coarseness affect each other. Although there are some individual differences, as shown in Table 2.3, the dominance of predictors is interchanged by $\Phi(0)$ and δ_e each other. For regularity, we could not find any significant

combination of predictors to fit the SV except for a single factor of δ_e. It means that the other ACF factors are not related to the perceived regularity. At this point, we think that the best predictor of perceived regularity is δ_e.

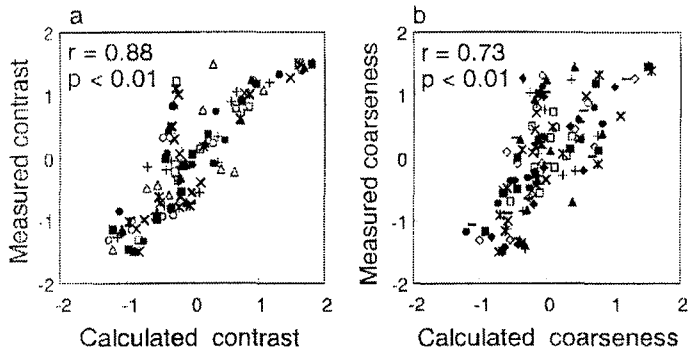

Figure 2.14 Relationships between predicted SVs by means of multiple regression analysis and measured SVs of contrast and coarseness. Each plot represents results for each subject.

Table 2.3. Regression coefficients of $\Phi(0)$ and δ_e and correlation coefficients of regression equations for SVs of contrast and coarseness.

	Contrast				Coarseness		
Subject	$\Phi(0)$	δ_e	r	Subject	$\Phi(0)$	δ_e	r
A	0.60**	0.50*	0.85**	A	0.26	0.71**	0.80**
B	0.55*	0.48*	0.80*	B	0.10	0.55*	0.58
C	0.79**	-0.02	0.78*	C	0.23	0.29	0.40
D	0.56*	0.47	0.79*	D	0.49*	0.40	0.68*
E	0.57*	0.44	0.78*	E	0.36	0.63*	0.78*
F	0.63**	0.44*	0.83**	F	0.16	0.81**	0.85**
G	0.77**	0.32	0.89**	G	0.09	0.43	0.46
H	0.86**	0.08	0.88**	H	-0.49	0.36	0.55
I	0.74**	0.21	0.80**	I	0.05	0.75**	0.76*
J	0.73**	0.43*	0.91**	J	0.15	0.70*	0.75*

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

Discussion

Based on the observations above, we believe that the ACF analysis is still useful for random textures. The value of $\Phi(0)$ corresponded well to the perceived contrast

similarly to that in regular textures. The values of $\Phi(0)$ in our experiment were centered in a narrow range, but were found to represent the perceived contrast well. These results mean that the RMS of the image is consistent with the perceived contrast.

As for regularity and coarseness, the value of δ_e corresponded to that in the subjective scores. This result is consistent with our assumption. When a texture has a harmonic structure, the estimated ACF has periodical peaks corresponding to the periods of the texture. Both the perceived coarseness and regularity are strongly related to these peaks in the ACF. For the textures with less harmonic structures, like those we used, the initial slope of the ACF is important. The value of δ_e shows a degree of correlation within the image. The more the textures correlate, the more the subjects perceived the coarse and regular properties. The boundary of these two aspects for the perception of texture periodicity was the value of $\phi_1 \approx 0.1$ in this study. Further research is needed to clarify this mechanism to develop a unified model of periodicity perception for any kind of texture.

The multiple regression analysis revealed that the perceived contrast and coarseness could be represented by a linear combination of the ACF factors with certain weightings. Of all the possible combinations, the most significant predictors were found to be $\Phi(0)$ and δ_e. Comparing this with the result of the preliminary experiment, we believe that the perceived contrast and coarseness may affect each other. Because the experiment was based on paired comparisons within a texture set, it was difficult to determine a single measure of perceived properties for both texture sets. Further research is needed using more textures or different texture sets to extend our method to be applied to any texture images.

2.3.5 Conclusion

We showed that the ACF analysis provides useful measures for representing the textural properties of contrast, coarseness, and regularity. This study differs from the previous works reviewed in Introduction mainly in the following two aspects. First, the examined textural properties in this study were chosen based on a study of the dimensionality of texture perception. The three properties examined here were found earlier to be salient in texture discrimination. Therefore, analyzing these properties for texture modeling makes sense. Second, the proposed analysis is based on a plausible mechanism in the human visual system. The computed factors were not composed heuristically. Consequently, the ACF model can provide important information for texture

representation.

The validity of the ACF analysis was examined by comparing the calculated factors to the subjective scores collected for various kinds of natural textures. The applicability of the ACF in analyzing random textures was also examined. The effectiveness of the analysis depends on the structures of the estimated ACF. When a texture has a harmonic structure, the estimated ACF has periodical peaks, which correspond to the periods of the texture. Both the perceived coarseness and regularity are strongly related to these peaks in the ACF. As for the random texture, however, the estimated ACF does not have a periodical structure. In this case, the effective range of the ACF is important to representing the texture coarseness and regularity.

Another contribution of this study is to describe a new approach toward understanding a mechanism after the relatively low-level and simple linear filtering stage. It is known that there are many local spatial filters in the visual system, that are tuned to particular spatial frequencies and orientations. However, it is difficult to characterize multiple perceptual properties of texture by means of this multi-filter mechanism alone. The described ACF model offers the advantage of extracting perceptual properties. Here we did not need to make assumptions about this early filtering stage, but considering the operation of local filters with the ACF model might give a full account of texture perception.

2.4 Texture preference

Summary

The purpose of the present study is to quantify the perceptual properties of texture and evaluate subjective preference of texture. Particularly, comparison between naturally occurred and man-made textures was attempted in terms of perceived regularity. Texture could be classified into two categories the degree of regularity, which is estimated by the amplitude of the maximum peak in the autocorrelation function (ACF). Psychological experiment showed that the degree of regularity is a salient cue for texture preference judgment. Human subjects preferred texture with mixed regularity and randomness, which arouse from fluctuations in object shape, brightness, color, and structural pattern.

2.4.1 Introduction

Background

In the field of architectural, landscape, or the industrial design, texture has an important role as design material so as to affect the subjective evaluation. Texture causes emotional or affective reactions such as impression and preference. However, expression of textural properties in an objective manner is difficult because texture is the mixture of many sensations and the basic visual properties of texture are not understood quantitatively. If the visual features of texture are handled in a numerical definition, it will be possible to develop suitable textures for individual designer and consumer by separating or blending the features. Measurement of textural properties corresponding to visual perception would be useful for the construction of objective scale for texture.

The word "texture" expresses the characteristic attribute of materials such as "wooden texture" or "stone texture", in architectural design. But we perceive texture for material surface in a various way depending on the observing conditions. We can perceive different texture in a different distance for a same material, and we can perceive texture for an assembly of objects such as pebbles or leaves. Therefore, it is considered that texture perception is caused not by a material itself but by an optical pattern, which is reflected by the material surface. In this paper, we treat texture as an optical pattern and try to describe its visual features by measurable physical features.

Previous studies

Texture has been studied on two levels in pattern recognition: statistical and structural (e.g., Tomita et al., 1982; Haralick, 1979). On the statistical level, texture is regarded as defined by a set of statistics. Even simple statistics such as the average, variance, and histogram of the gray level, can be used for classifying a limited class of textures. Humans are sensitive to second-order statistics, as pointed out by Julesz (1973). The gray level co-occurrence matrix, the Fourier power spectrum, the autoregression model, and the autocorrelation function are second-order statistics. Roughly speaking, these statistical methods are useful for random pattern textures. To describe textures of more complex structures such as brick wall and lattice pattern, structural analysis is useful. On the structural level, a texture is defined by elements (primitive), which occur repeatedly according to their placement rules.

Textural properties used in pattern recognition are defined for computational convenience. To evaluate the subjective response of texture, physical description of texture should correspond to the human visual perception. Tamura et al. (1978) proposed six textural features corresponding to visual perception: coarseness, contrast, regularity, directionality, line-likeness, and roughness. As similar directions, Amandasun (1989), Cho (2000), and Rao & Lohse (1996) investigated textural properties for discrimination and dissimilarity judgment of texture. According to these studies, regularity is the most important properties. They suggested that three or four dimensions are sufficient to describe texture as the combination of regularity and other properties.

Purpose of the study

The purpose of the present study is to quantify the perceptual properties of texture and evaluate subjective preference of texture. Particularly, comparison between naturally occurred and man-made textures was attempted in terms of perceived regularity. Psychological experiment was conducted for human subjects by means of paired comparison method.

2.4.2 Analysis

Material and procedure

As described before, perception of texture is dependent of observed distance. Texture of brick or tile pattern is considered to have the following two levels: (1) texture of material, and (2) placement pattern of materials. These two levels could be considered as statistical texture and structural texture. Generally, texture preference is considered as preference of material itself. However, in the case of wall texture, which consists of an assembly of objects, psychological evaluation is affected by both levels. Furthermore, considering the fact that the whole feature is understood before the partial feature (global precedence phenomenon) as Navon (1977) pointed out, the structural feature is more important. In this study, therefore, I treated architectural wall texture as the whole pattern rather than the material texture. Later, I mean the word texture as structural texture unless exceptional description.

Following is the flow of data analysis: 1) photographing texture in the natural environment, 2) digitizing and editing texture images, 3) pre-processing to the images, and 4) calculation of the ACF. Details of each stage are described below.

As analyzed material, man-made texture such as lattice, brick, and tile pattern, and natural textures such as leaves and pebbles was photographed in the real environment (Figure 2.15). Pictures were digitized by an image scanner and edited to 256 × 256 pixels, 8-bit gray scale images. In the psychophysical experiment by Ohno, it was found that we perceive "random dot pattern" as texture when the spatial frequency is roughly between 5 and 30 cycle/deg (Ohno, 1989). In the real visual environment, however, the spatial frequency distribution is more coarse and complex to examine the precise range of texture perception. To keep the digitized image within the texture perceived range, resolution of the image scanner was decided such that the size of each primitive of the image was between 10 and 30 pixels (0.1 and 0.3 degree at 1.5 m, giving spatial frequencies of 10 and 3.3 cycle/deg).

Signal whose mean and variance are not constant over space and time is called "non-stationary" signal. Many non-stationarities could be observed in the collected texture images. Especially, the lighting conditions are not uniform across the image. For example, there can be an illumination gradient across the image. Such non-stationary signal is not suitable for the ACF analysis, because a slow periodicity much longer than the integration range cannot be directory captured by the ACF. To remove such effect, low frequency components were cut out by applying high-pass filter.

After pre-processing, the ACF was calculated for the objective evaluation of texture. As described in previous section, textural properties of contrast, coarseness, and regularity could be quantified by the ACF factors. Following four factors were extracted from calculated two-dimensional ACF. 1) $\Phi(0)$: the average power of the image, which corresponds to perceived contrast. 2) δ_1: the displacement of the maximum peak and (3) ϕ_1: the amplitude of the maximum peak in the normalized ACF. These two factors correspond to perceived coarseness and regularity, respectively. 4) δ_e: the effective range of the ACF. For texture without harmonic structure, the value of δ_e corresponds to coarseness and regularity of texture.

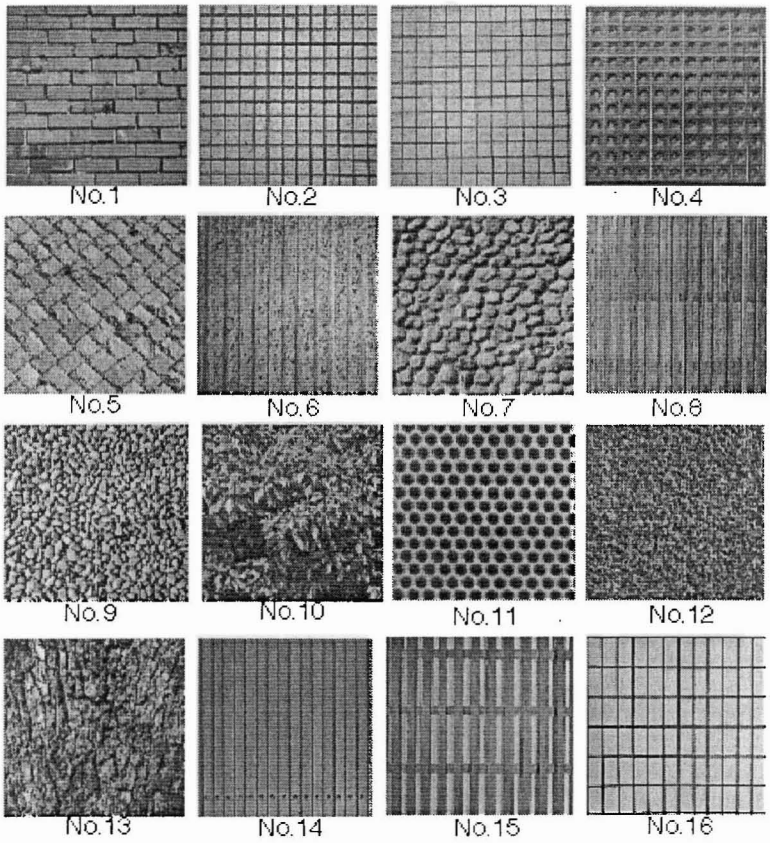

Figure 2.15 Man-made and naturally occurring textures photographed in the real environment.

Results

The ACF of regular texture has higher degree of periodicity than of random texture. To compare the degree of periodicity, we consider the amplitude of the maximum peak in the ACF, ϕ_1. As described in previous section, rating of the ϕ_1 value could represents perceived regularity of texture. Figure 2.16 tells us the relationship between the value of ϕ_1 and perceived regularity well. Natural occurring texture or texture of natural material such as stone has lower value of ϕ_1 than man-made texture. Man-made texture is a regularly structured pattern with same shape and size, whereas various size and shape objects are placed randomly in natural texture. This difference is well represented in the periodicity of the calculated ACF.

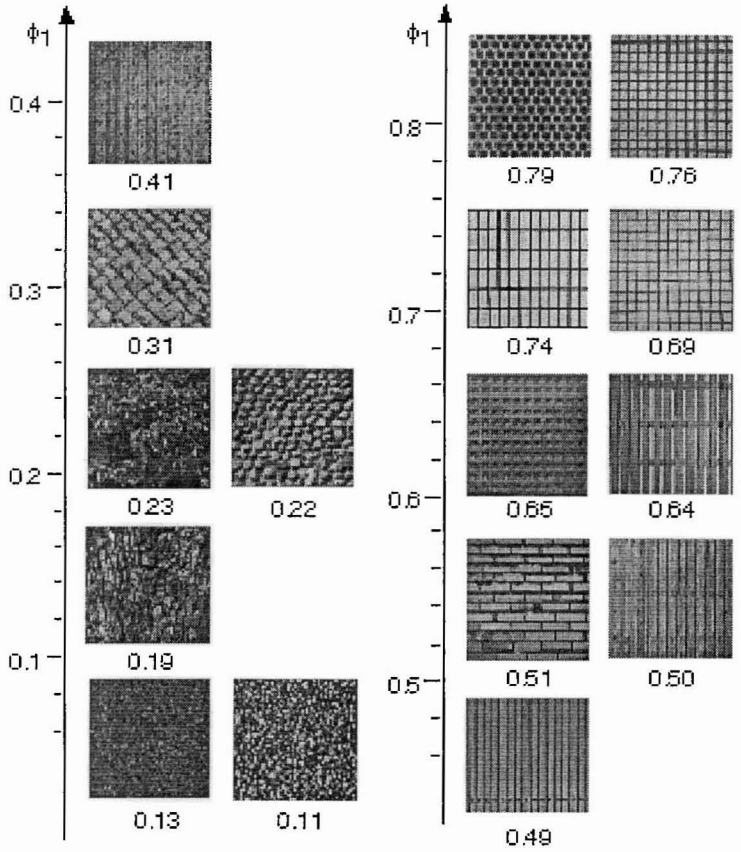

Figure 2.16 Ranking of ϕ_1 for 16 textures. It is clear that perceived regularity is described by the ϕ_1 values.

Wall pattern of brick and tile is generally made of standardized manufactured materials. If the size and shape of materials, and spacing between the objects in pattern is completely equal, the calculated ACF does not decay. It means that such texture is perfectly regular in theoretically. But practically, there exists some kind of fluctuation, because of a variety of small error in object size and spacing, and un-uniformity in light reflection. Consequently, perceived texture is not perfectly regular. In this case the ACF gradually decays. Therefore, the value of ϕ_1, which is the measure of perceived regularity, is also considered as the measure for degree of fluctuation in texture.

Natural texture, in contrast, generally has randomness because of a variety of size and shape of objects and placement rules. Therefore the calculated ACF does not have periodical component and decays steeply (Figure 2.17). But they are not completely random. They also have some degree of regularity as described by the ACF factor ϕ_1. Figure 2.16 shows that the more organized the shape and size of component is, the larger the ϕ_1 value is.

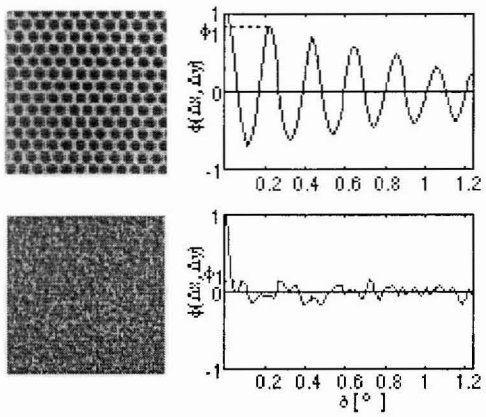

Figure 2.17 Calculated ACFs for regular texture (top) and random texture (bottom). For random texture, the ACF does not have periodical peaks and decays steeply.

To summarize the result of analysis, we can say that natural texture consists of a balance of regularity and randomness. Man-made texture has high degree of regularity, and naturally occurred texture has high degree of randomness. Human preference for texture is assumed to be related to this balance of regularity and randomness. Extremely regular or random texture may not be preferred. It could be assumed that the preferred texture has moderate balance of regularity and randomness. In the following experiment, this assumption is examined.

2.4.3 Psychological experiment

To examine how much regularity is preferred, we conducted a psychological experiment. Human subjects judged five samples from the analyzed texture set by means of paired comparison test. It is hypothesized that the most preferred texture had moderate degree of regularity. We tried to formulate the subjective score of preference by the calculated ACF factors.

Method

Ten male students participated in the experiment. Their ages range from 22 to 24. All had normal or corrected to normal visual acuity. They were naive as to the purpose of the study.

Stimuli were presented on a CRT display under a dark surrounding. The display was set at a distance of 1.5 m from the subjects. Stimulus was 256 × 256, 8-bit gray images extending about 2.5 degree of the visual angle. Stimuli were chosen from

the texture set described in previous subsection. Stimulus textures have different values of ϕ_1, which means each texture have different degree of regularity. Other two properties of contrast and coarseness were kept constant by following manipulations. Contrast of stimulus images was adjusted by dynamic range of gray scales, and coarseness of the images was adjusted by modifying resolution.

Subjects were presented pairs of two stimuli and asked to judge whether they preferred. Two stimuli lasted 3 s with a blank of 1 s between them. All possible pairs from five stimuli (10 pairs) were presented in a random order in one session. All subjects conducted ten series of sessions, giving totally 100 judgments.

Data analysis and results

Collected data was analyzed by applying "the law of comparative judgment" (case V; Thurstone, 1927). This was used to produce a one-dimensional scale value (SV) for each stimulus from the total matrix of superiorities collected from the paired comparisons. The results were reconfirmed by the goodness of fit (Mosteller, 1951), and the agreement of all subjects' judgments was tested by the chi-square test ($p < 0.05$).

Results of all subjects are shown in Figure 2.18, top. SV of preference has single peak value for each subject, even allowing some individual differences. The most preferred range might exist in the value of ϕ_1 for each subject. Subjects did not prefer texture of which high and low value of ϕ_1. By averaging the scores of all subjects, it was found that $\phi_1 \approx 0.5$ was most preferred value for texture regularity (Figure 2.18, bottom). Considering these results and previous section together, it is considered that the degree of regularity is a visual property affecting subjective preference for texture.

Comparing the result of individual subject, evaluations for random texture of which the ϕ_1 value is low had large variance. It might be other factor than degree of regularity for subjects' preference judgment. In the present experiment, we used various textures in natural environment for stimuli. Although we choose stimuli in accordance with their ϕ_1 values with other two variables (i.e. contrast and coarseness) constant, there might be other variables exist. One of such variables, we can consider the effect of difference in material texture. As described before, texture perception is decided by global feature and local feature. In the present experiment, the subjects' judgment might be affected by both of global degree of regularity and local textural pattern of material. About this point, experimental method could be examined such that the stimuli are changed to the synthesized artificial texture with material texture excluded.

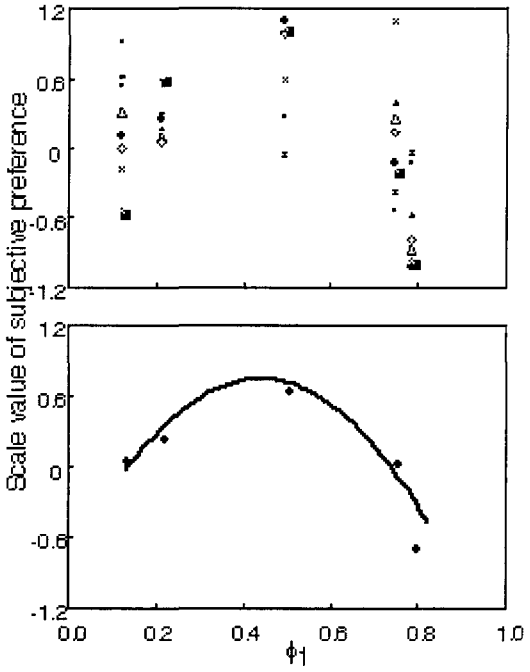

Figure 2.18 Scale values (SV) of preference for all subjects (top), and average SV (bottom). For average SV, fitting curve is also shown.

2.4.4 Discussion

As I described first in this section, this study treated texture as an assembly of the primitive objects. Therefore, I concerned with the size, shape and placement pattern of the component materials. To describe physical properties of texture in building surface, the ACF analysis was applied. When the same components are placed in regular pattern, the ACF does not decay theoretically. Consequently, the maximum peak in the ACF always becomes $\phi_1 = 1.0$. But practically, for the man-made texture in the real world, the ACF decays gradually to zero and the ACF peak takes value $\phi_1 < 1.0$. This is because of the noise component due to the error in spacing between the material and fluctuation in light reflections. The present experiment showed that texture with such noise components are preferred to texture with high degree of regularity. Preferred texture needs to have a balance of regularity and randomness. Therefore, when we make a wall texture by assembling manufactured materials like tile and brick, we should use a set of materials in which some degree of variance in color and shapes contained. Along with this discussion, natural material has variety of shape and size, and its surface is uneven, irregular, and rugged so as to introduce a fluctuation of light reflection. This

might be one of the reasons that such natural materials are preferred to the manufactured materials.

For naturally occurring texture such as leaves and pebbles, texture image inherently include higher degree of randomness, because of a variety of size and shape of objects and placement rules. But they are not completely random. They also have some degree of regularity as described by the ACF factor ϕ_1. Our analysis showed that the more regular the shape and size of component is, the larger the ϕ_1 value is. Our subjects did not prefer texture with lower value of ϕ_1. We can say again that the preferred texture needs to have a balance of regularity and randomness.

As I discussed above, this study has shown that the ACF analysis is useful to describe textural property of contrast, coarseness, and regularity. Also, the degree of regularity in texture is an affecting factor in human evaluation for texture preference. As an example of application of the results gained, we can consider a texture of tile pattern with some degree of randomness introduced by manipulating the placement pattern or color variation. We can examine the most preferred degree of randomness by computer simulation as shown in Figure 2.19.

Finally, of course, there might be another factor other than regularity, which is not examined in this study. Preference judgment of texture may not be as simple as that we could explain only by one-dimensional analysis. Thus we need to find such a factor and analyze the texture preference multi-dimensionally.

2.5 Summary

In this chapter, I presented a series of studies on physical properties and psychological evaluations of two-dimensional spatial pattern. I showed that the autocorrelation function (ACF) analysis provides useful measures for representing three salient perceptual properties of texture, namely, contrast, coarseness, and regularity. Another experiment showed that the degree of regularity is a salient cue for texture preference judgment. Described ACF model offered the advantage of extracting perceptual properties and evaluating subjective reaction in texture perception.

3. Temporal vision

3.1 Introduction

In this chapter, I discuss the underlying mechanism for the temporal perception in vision. In the field of vision science, temporal processing in visual system has long been neglected. Because the temporal resolution of the visual system is much poor in contrast with its high spatial resolution, we are apt to think that the important information is received mostly by spatial vision. However, the temporal information is very important for perception in the real environment. Detection of the moving objects is crucial for our life. Information of rhythm or tempo is a cue for perception of periodical motion such as *human motion* and other *biological activity*, and natural behaviors like *waving water*, *fire*, and *leaves moving in the wind.*

When we perceive tempo or rhythm for a motion or some behavior, it is considered that we detect a kind of periodicity contained in that signal. When we could detect a periodicity and that periodicity is in the limited time range, we could perceive a sense of tempo. Rhythmic perception occurs when the sequences of successive events have more complex pattern, which are perceived as groupings of two, three or four elements. It should be noted that subjective rhythm is created by the repetition of stimulus and give rise to the repetition of a pattern in the time sequence. In both cases of tempo and rhythm perception, we need to detect a kind of periodicity in the signal. Then, is there a mechanism for detecting periodicity involved in the visual system? It is a main question posed in this chapter. It is recognized that temporal processing such as periodicity detection is an important strategy in signal processing in the auditory system. Perceived pitch of a sound is decided as the most salient periodicity in the signal by means of correlation mechanism. Is there a correlation mechanism also in the temporal vision mechanism? To address the problem, I focus on the fundamental properties of the temporal vision mechanism, namely the subjective flicker rate of complex waveforms. After a brief review of relevant field, I present an experimental result, which implies the existence of correlation mechanism in temporal vision (Fujii et al., 2000).

3.2 Related work

In this section I will give a brief review of previous works on response to flickering light, mechanism of temporal vision, and temporal perception.

3.2.1 Studies on flicker response

A number of studies have dealt with the human response to flickering light. They mostly concerned the sensitivity to or detectability of flicker with sinusoidal modulation. Temporal sensitivity refers to our response to timing of visual signals. Consider a light that is pulsing, such as the blinking cursor on a computer screen. It is easy to discern that the light is blinking because of the low frequency, or rate at which it flashes. Imagine gradually increasing the frequency. As the light blinks faster and faster, it would eventually reach a rate where it would no longer be possible to detect that the light is flashing, it would look like a "solid", or continuous light. We say that our visual system has fused the flicker.

The frequency at which that occurs is called the critical flicker frequency (CFF) or flicker fusion frequency (FFF). Refresh rates of cinematography (48 Hz) and television (60 Hz) are carefully chosen to prevent flicker. But now it is known that CFF depends on a variety of physical and psychological factors. Hecht and Verrijp (1933) showed that CFF increases as luminance is increased. More precisely, de Lange (1958) demonstrated that for a visual stimulus with a fixed temporal frequency, sensitivity to flicker increases as the amplitude of the luminance modulation is increased. Flicker perception depends on the depth of modulation (modulation amplitude) and not only on time-averaged luminance. Temporal frequency characteristics of visual flicker sensitivity have been found by Kelly (1961, 1969, 1971). Figure 3.1 shows the flicker threshold measured with five levels of background (mean) luminance. The figure tells that flicker sensitivity generally has band-passed characteristic (peaks around 10-20 Hz) and the mean luminance increases CFF.

Another physical factors that determine the rate at which flicker is fused is size of the stimulus and location of the stimulus within the visual field. At frequencies higher than about 20 Hz, CFF is known to increase with increasing target stimulus size. Below 20 Hz, sensitivity decreases with increasing target size (Keesey, 1972). According to the Granit-Harper Law, CFF increases linearly with the logarithm of the retinal area subtended by the target (Granit and Harper, 1930). For small targets flicker is more easily perceived when the image falls on fovea, rather than in the periphery.

This may appear to suggest that the fovea is more sensitive to flicker. However, the periphery is more sensitive to flicker when larger targets (such as flickering fields or CRT screens) are used. Roehrig (1959) best demonstrated this by showing that CFF increases only when the outer circumference of a visual target is increased, while CFF remains the same when the inner circumference of a ring-shaped target of the same size is decreased.

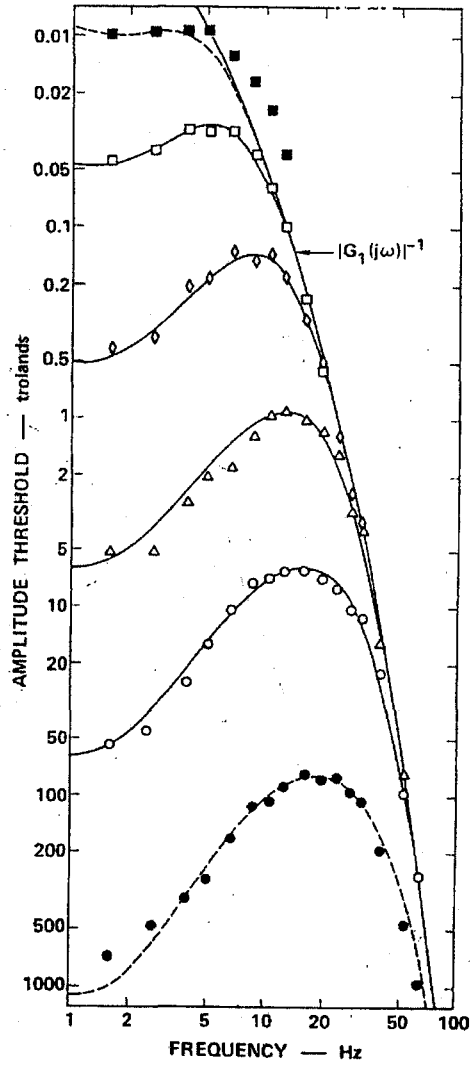

Figure 3.1 Temporal frequency characteristics of flicker sensitivity, measured by Kelly (1961).

3.2.2 Model of temporal vision

Studies of temporal sensitivity in the visual system have been dominated by linear

system theory. Most of psychophysical data is economically described within this framework. The underlying assumption is that the visual response to any periodic waveforms could be analyzed by decomposing the waveform into its Fourier series and then evaluating response to its components separately. One of an example is a study of de Lange (1952). He investigated the sensitivity to a variety of different flicker like square and saw-tooth waveforms. Square waves and saw-tooth waves each can be regarded as sums of a non-flickering component (i.e., average luminance or dc component), the fundamental component (i.e., the first harmonic), and a series of higher-harmonic components. He showed that above 10 Hz, sensitivity to a variety of different periodic waveforms could be predicted from sensitivity to the first harmonic component. This implies that the eyes respond to any periodic flicker as a linear system.

The visual system, however, does not respond perfectly linearly. Especially for flicker above threshold (i.e., at high contrast level), the visual system behaves nonlinearly. For example, it is known that flickering lights appear brighter than steady lights of equal mean luminance. This is an effect named brightness enhancement or the Brücke-Bartley effect (e.g., Bartley et al, 1957). The effect of brightness enhancement could not be explained by a linear system. It requires nonlinearity somewhere in the visual system. Wu et al. (1996) investigated the mechanism of brightness enhancement by using amplitude-modulated flicker (Figure 3.2).

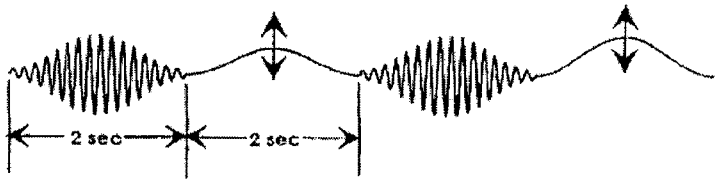

Figure 3.2 Brightness enhancement caused by the amplitude-modulated flicker, Wu et al. (1996).

The results could be modeled by a broad temporal filter followed by a single accelerating nonlinearity (Figure 3.3). But they also suggested that the nonlinearity is more complex than a single, static one. As possible alternatives, compressive nonlinearity, asymmetric detector (Kelly & Savoie, 1978), and dual-pathway (ON and OFF) model (Krauskopf, 1980; Schiller et al., 1986) could be considered (Figure 3.4).

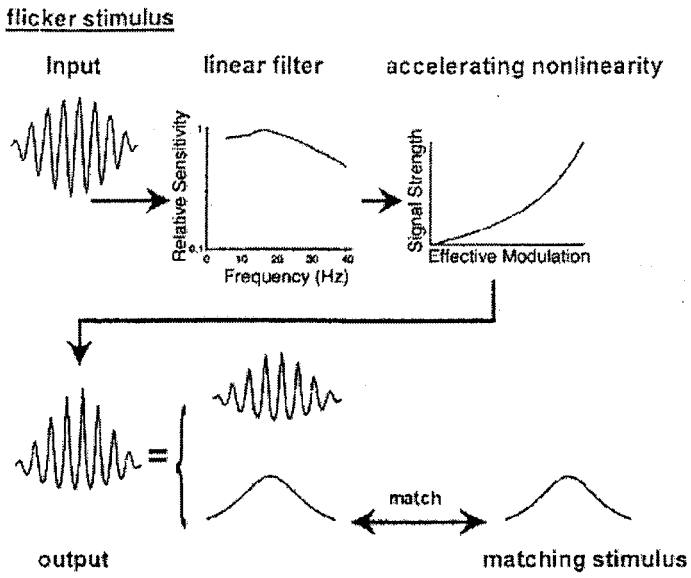

Figure 3.3 Nonlinear model proposed by Wu et al. (1996).

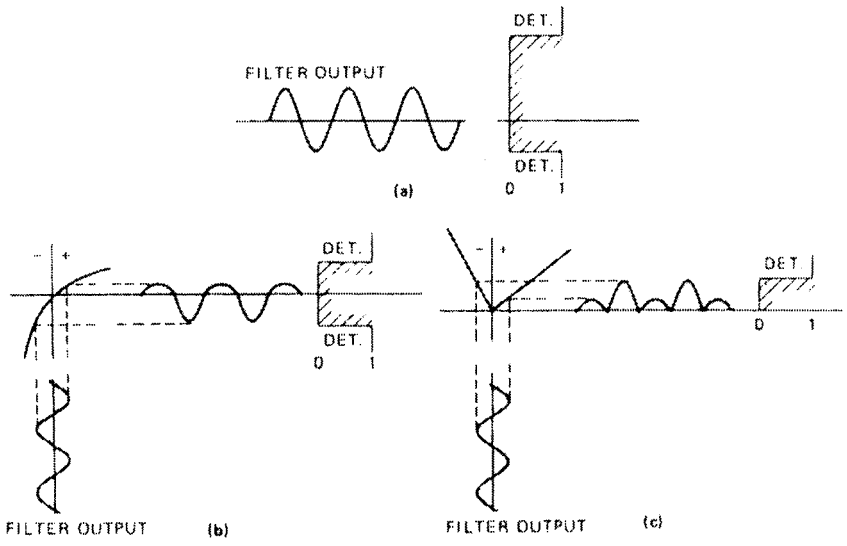

Figure 3.4 Three types of asymmetric nonlinearities in temporal vision proposed by Kelly (1978) (a) An offset detector, in which the positive limit is 2.7 times the negative one, with no other intervening nonlinearity. (b) A strongly compressional nonlinearity, followed by a symmetrical, two-sided detector. (c) As asymmetric rectifier, with a negative slope 2.7 times its positive slope, followed by a one-sided detector.

Most of studies mentioned above assumed a single pathway in the visual system so as to simplify the problem. However, a fundamental property of the visual

system is multiplicity of parallel pathways. Since the properties of these initial pathways places fundamental constraints on subsequent stages of visual processing, it is important to understand the individual properties of these pathways and how they are organized. It is now accepted that the initial stage of visual processing involves an array of spatio-temporal filters implemented by the characteristics of receptive fields. A collection of receptive fields with the same filtering properties is also called "channel". Although we have a substantial amount of information about the spatial properties of these filters, investigation into their temporal properties has been neglected. From psychophysical data of adaptation (Mandler and Bowker, 1980), and other evidence, the channels that are tuned to different temporal frequencies were believed to be few in number and broad in frequency bandwidth. Data on temporal frequency discrimination (Mandler, 1984; Mandler and Makous, 1984) required at least three channels. Hess and Snowden (1992) found evidence for three temporal channels - a low pass channel, a band-pass channel peaked at around 10Hz and another peaked at around 18 Hz (Figure 3.5). Properties of these channels are used in the model of motion perception (Johnston and Clifford, 1995).

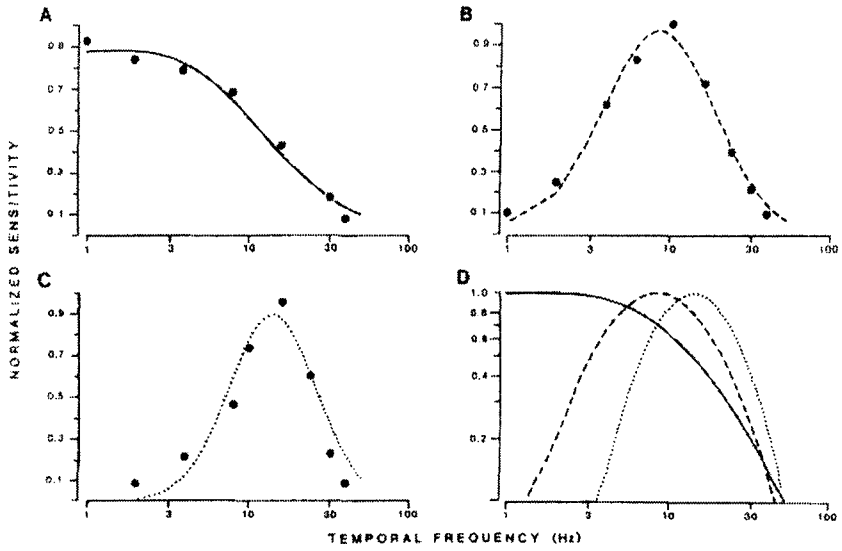

Figure 3.5 Hess and Snowden (1992), three temporal filters.

3.2.3 Studies on time perception

Time perception is formed when the stimuli are presented both visually and auditory. Although the duration of perceived time is different to some extent, they are not

perceived as different kind of time. Many researchers accepted the idea that, to perceive time is in fact a general expression that means that succession and duration of the event are both perceived. This idea means that the results of the processing information in physical stimuli give the cue for time perception. Perception of duration refers to the ability to comprehend successive events as perceptually simultaneous, within the framework of the psychological present. Psychological present has no fixed duration. It is based on what is perceived and refers to our capacity to apprehend the series of events. It has an upper limit about 5 s, and has an average value of 2 to 3 s (Fraisse, 1984).

To speak of rhythm is to speak of ordering in temporal succession. A series of identical sounds (separated by equal intervals) heard for a certain time is spontaneously perceived as grouping of two, three, or four elements. This phenomenon is also observed with visual stimuli (Handel & Yoder, 1975). Such subjective rhythms are created by the repetition of stimulus and give rise to the repetition of a pattern in the course of time. This aspect seems to be a gestalt in the time domain. We perceive tempo when the equal or almost equal length of duration is presented successively. The succession of shorter time duration is perceived as faster rate. We can make estimation for tempo as fast or late, and we can produce spontaneous tempo by tapping of fingers. For both tempos perceived moderate and produced spontaneously, duration between the events is several hundred ms. This duration has a range with an upper limit about 2 s and lower limit about 100 ms. Mishima (1951, 1956) investigated such a "mental tempo" by using the flickering light and metronome. He found consistent tempo between different modalities.

3.3 Missing fundamental phenomenon in temporal vision

Summary

Subjective flicker rates were measured for compound waveforms consisting of five harmonics without fundamental component. It is found that observers perceived a rate at the fundamental frequency, although the energy at this frequency was not included in the signals. It is called the missing fundamental phenomenon in auditory pitch sensation, and an analogous finding is known to occur in spatial vision. Moreover, observers perceived the rates at fundamental even in the random-phase conditions, in which real waveforms the period of the fundamental is unclear. The results indicate that the perceived flicker rates are not detected from the temporal waveforms *per se*. One

possible operation to extract such a periodicity in the signal is autocorrelation function to the nonlinearly distorted temporal waveforms.

3.3.1 Introduction

This section describes a phenomenon that shows an analogy between visual and auditory system. As described later, it is called "missing fundamental", which is known in the auditory pitch sensation. When the signal contains only a number of harmonics without fundamental frequency, we hear the fundamental frequency as a pitch. The missing fundamental phenomenon found in temporal vision is presented here.

We investigated how humans process or perceive flickers with compound waveforms. Although a number of studies have dealt with the human response to flickers, they are mostly concerned with the sensitivity to flickering stimuli. In contrast, this paper is concerned with the temporal aspect of visual perception. Our interest is what humans perceive in the temporal dynamics of the visual stimulus. Some studies were related to compound waveforms (de Lange, 1952; Bowen, Pokorny & Smith, 1989; Bowen, Pokorny, Smith & Fowler, 1992; Kremers, Lee, Pokorny & Smith, 1993; Eisner, 1995). In particular, square or sawtooth waveforms were commonly used in comparison with sinusoidal waves. Square and sawtooth waveforms each consist of the fundamental frequency (F_0) and a series of sinusoidal components (harmonics). For sufficiently high temporal frequencies the sensitivity to a variety of different periodic waveforms could be predicted from sensitivity to the F_0 component. At lower frequencies, the sensitivities to the compound waveforms were affected by higher-order harmonics (de Lange, 1954). On the other hand, we know of no previous studies that dealt with a compound waveform without the F_0 component. The effect of F_0, which is not contained in the waveform is known in the spatial vision (Henning, Herz & Broadbent, 1975; Nachmias & Rogowitz, 1983) and in the auditory pitch sensation (Wightman, 1973a,b).

Henning et al. (1975) reported that in their experiment on the simultaneous masking of vision, the F_0 component not being contained in the masking stimulus affected the detection of the test stimulus. They used 1.9 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 4-th, 5-th, and 6-th 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 perception of pitch in the auditory system, it is known that the perceived pitch for a complex tone consists of some harmonics (e.g., 600, 800, 1000,1200, and 1400 Hz) corresponds to the F_0 (200 Hz). Such a pitch is called a residue, low pitch or virtual pitch. Wightman (1973a) found that the residue is not affected by the relative phase of the components in the stimulus. It means that the pitches of complex tones do not depend on the stimulus waveforms. According to the pattern-transformation model (Wightman, 1973b), the auditory system detects the pitch not from the temporal waveform *per se* but by means of peaks in the autocorrelation function of the stimulus. Autocorrelation function provides information about pitch, which is not dependent on the stimulus waveforms.

The two major questions we wished to address were, firstly, whether the F_0 component was perceived, and secondly whether the subjective rate is affected by the relative phase of the components in the stimulus. In the following experiment, we measured the subjective flicker rates for compound waveforms consisting of harmonic components (without F_0). Since the components are combined linearly, there is no Fourier energy at F_0. To examine the second question, we used two kinds of waveforms as the stimulus. One was an in-phase, and the other was a random-phase waveform. If the perceived rates are based on the temporal waveform itself, observers could not detect the rates for random-phase stimuli, because the relative phase of components makes the periodicity of waveforms unclear. If, on the other hand, the F_0 components are perceived for both in- and random-phase stimuli, there should be a mechanism to detect it.

3.3.2 Method

Observers

Four human observers, who were males and 23-26 years old, participated in the experiment. All had normal or corrected-to-normal vision. They were allowed sufficient practice before starting the experiment, because they had never participated in such an experiment before. They dark-adapted for about 1 minute before all sessions.

Apparatus

The light source was a 7-mm-diam green LED, set at a distance of 0.8 m from the observer in dark surrounding. The LED stimulus field was spatially uniform and the

size of it corresponded to 0.5 deg. Stimulus waveforms were generated by the computer with a 16-bit digital-to-analog converter. The mean luminance was set to 20 cd/m^2 and kept constant during the sessions. To prove the linearity of the apparatus, we measured the luminance waveforms of the stimuli with a luminance meter (TOPCON BM-8) with response time of 1 ms. The LED output was not linear at very low light level, but such nonlinear components were sufficiently smaller than signal components (-30 dB) for our experiment.

Stimuli

Stimuli in the present study were compound waveforms consisting of five components. The frequency of each component corresponded to the n-th harmonic of the common F_0. We selected eight combinations of components in terms of frequency range and F_0. The F_0 for stimuli 1, 2, 3, and 4 was 1 Hz. Stimulus 1 consisted of 3, 4, 5, 6, and 7 Hz, and the ranges 11-15, 21-25, and 31-35 Hz were selected for stimuli 2, 3, and 4, respectively. For stimuli 5, 6, 7, and 8, components were selected in the frequency range between 30 and 40 Hz. Stimulus 5 consisted of 30, 30.75, 31.5, 32.25, and 33 Hz, harmonics of the 0.75 Hz F_0. The F_0s of stimuli 6, 7, and 8 were 2, 2.5, and 3 Hz, and the components were 30-38, 30-40, and 27-39 Hz, respectively. These components had equal amplitude and were compounded to make two kinds of waveforms for each stimulus. One was an "in-phase ", and the other was a "random-phase" waveform. The waveforms of the signals used in the experiment are illustrated in Figure 3.6. It shows that the temporal waveforms of stimuli were affected by the phase of components, so that the in-phase and random-phase stimuli had different waveforms. The in-phase waveforms 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 peak. In the experiment, four "random-phase" waveforms were presented to all observers.

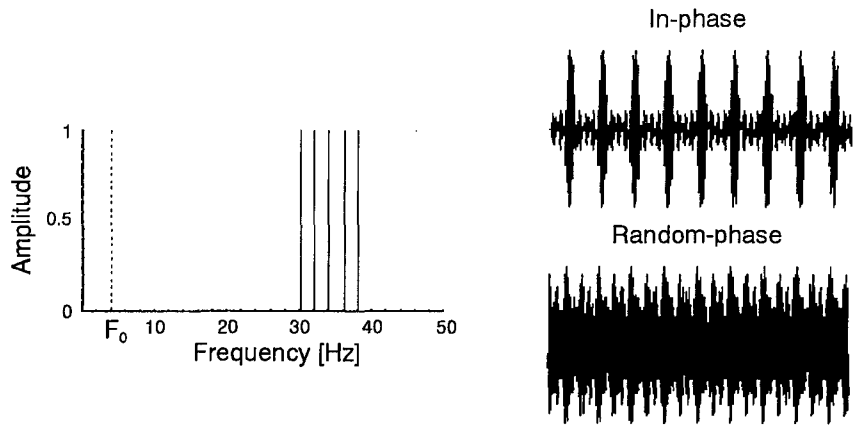

Figure 3.6 Left: the spectrum of the compound waves used in the experiment (components are 30, 32, 34, 36, and 38 Hz, for example) does not include the energy at fundamental frequency F0 = 2 Hz). Right: real waveforms show that the temporal waveforms are affected by the phase of components. The in-phase waveform has remarkable peaks corresponding to the F0, and the random-phase waveform has several peaks, of which periodicity is not clear.

Procedure

The subjective flicker rate of the stimulus was measured by means of the "method of limits". The flicker with compound waveforms was used as the test stimulus, and sinusoidal flicker was used as a comparison stimulus. These two stimuli were presented in pairs with a blank interval. The observers' task was to judge which of these two stimuli seemed to flicker at the faster rate. As the comparison stimulus, we used ascending and descending series. That is, 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' response reversed. The mean of the two values before and after reversal of the observers' response 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. It 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 above 3 Hz. In the descending series, trials started from a value a few Hz 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 condition four trials were repeated. Thus, 16 matched frequencies were obtained for each test stimulus.

3.3.3 Results

Results are shown in Figure 3.7 and 3.8 as histograms. The abscissa represents the frequency of the comparison stimulus matched with each test stimulus. On the ordinate, the probability of the response is represented in percentiles. We conducted the experiment for several frequency ranges and found that sensitivity decreased for the high frequency components in the compound waveforms. There were limit frequencies, above which no response was seen to the component frequencies. These values allowed some individual differences between 5 Hz to 20 Hz. For the stimuli with components of low frequencies below such limits, observers were able to match test stimuli with component frequencies. For the in-phase stimuli, observers perceived the rates at F_0. But in this case, this frequency is easily detected, because it is consistent with the time interval between the periodic peaks appearing in the temporal waveforms as shown in Figure 3.6. For the random-phase stimuli, matched frequencies were comparable to the several aperiodic peaks, which correspond to the component frequencies. It is possible that the human visual system could detect the flicker rates from local peaks in the waveforms in this low frequency range. In the high frequency range above the limits described, however, the F_0 components were perceived most frequently for both in- and random-phase stimuli, even allowing some exception such as certain multiples of F_0 (Figure 3.7, right).

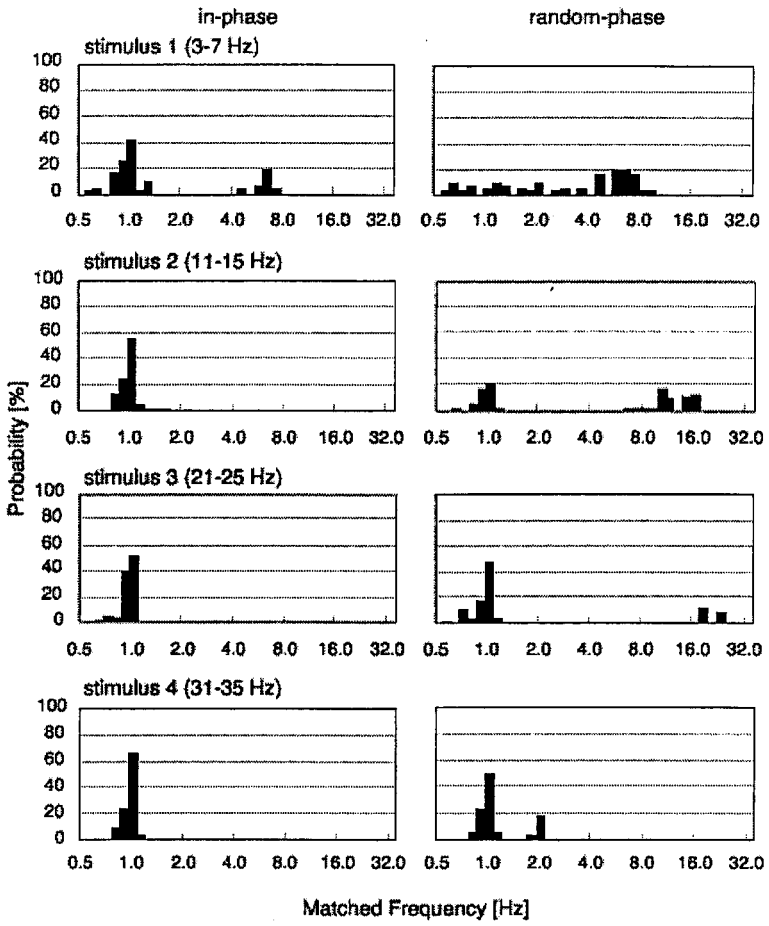

Figure 3.7 Results of four observers were summed. The abscissa represents the frequency of the comparison stimulus matched with each test stimulus. On the ordinate, the probability of the response is represented in percentiles. For stimuli 1-4, F0 was set as 1 Hz. Component frequencies were chosen between 3-7 Hz to 31-35 Hz, respectively.

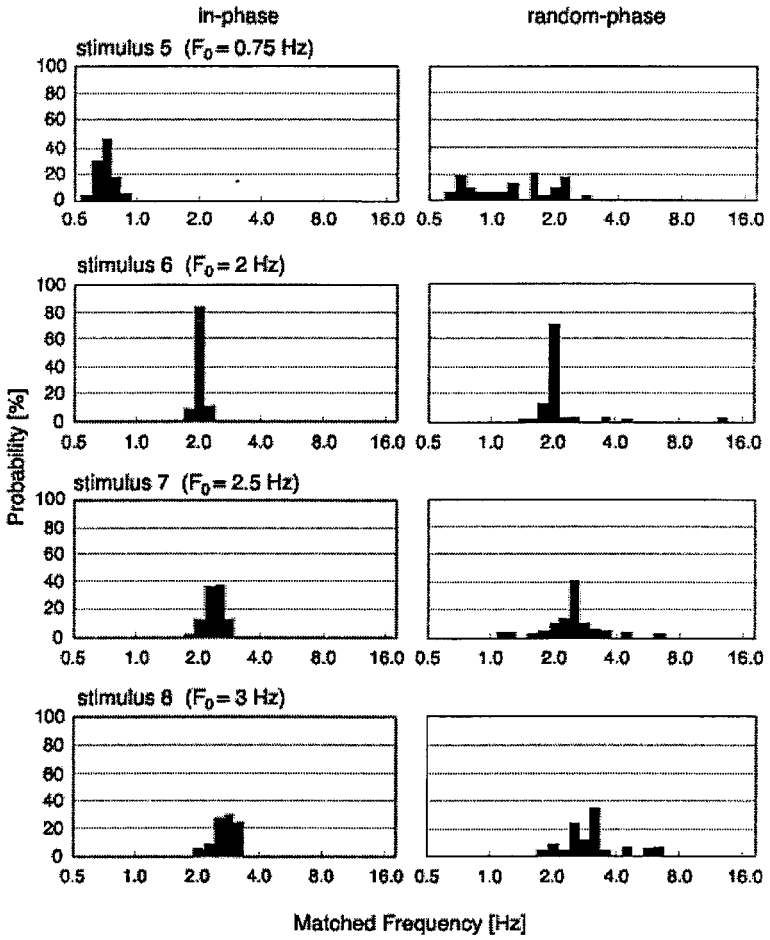

Figure 3.8 Results of four observers were summed. The abscissa represents the frequency of the comparison stimulus matched with each test stimulus. On the ordinate, the probability of the response is represented in percentiles. The F0s for stimuli 5-8 were varied between 0.75 Hz and 3 Hz. Component frequencies were selected between 30-40 Hz.

Figure 3.9 shows the relationship between the observers' response and F_0. Both curves have a similar profile except that the probability was about 10 % higher for in-phase condition. Although probability was affected by phase, the most frequently perceived rates were F_0 component in all cases. The highest probability is seen at 2 and 2.5 Hz for random- and in-phase condition, respectively. These values correspond to the periods of 500 ms and 400 ms, which periods 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. Since the observers were told to match to the most prominent component, a

missing fundamental that was visible but was not the strongest component is missed in the data. Had the observers been told to match to all visible frequencies, they might have given more response to F_0 even out of the sensitive range.

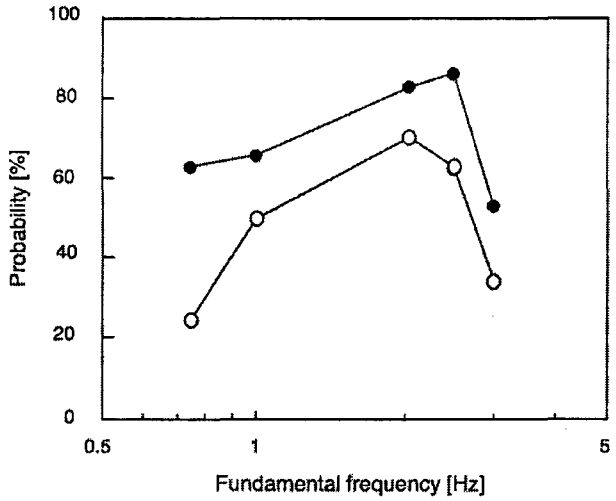

Figure 3.9 Relationship between the response probability and F0. The data for five F0s (stimulus 4, 5, 6, 7 and 8) were gathered and plotted in terms of response probability in the F0 ± 10 % frequency range. Filled circles and open circles represent in- and random-phase conditions, respectively.

3.3.4 Discussion

The purpose of this study was to investigate the missing fundamental phenomenon in the judgment of flicker rates by using the compound waveforms. In the high frequency range above 20 Hz, we found a phenomenon of missing fundamental. Observers detected the rates at fundamental frequency, which were not included in the stimuli. The results indicate that the perceived flicker rates are not detected from the temporal waveforms *per se*, because the period of F_0 is not clear in the temporal waveforms for the random-phase stimuli. It is necessary to suppose another mechanism in the visual system to extract the periodicity in the temporal variation of flickering stimuli.

It is known that the visual system behaves nonlinearly for flickering stimuli that are above threshold, such as the stimuli we used (Wu, Burns, Reeves, & Elsner, 1996; Macleod, Williams, & Makous, 1992; Eisner, Shapiro, & Middleton, 1998). Suprathreshold flicker is accompanied by changes in color and changes in brightness. Such nonlinear distortions produce real sinusoidal components at the fundamental frequency and some multiples of it in the compound waveforms. We have examined

whether the nonlinear distortion affected the observers' responses in our experiments. The amplitude of the distortion depends on the attenuation by the temporal filter preceding the nonlinearity and the form of nonlinearity. For the present purpose, this can be represented simply by a power series in the temporal waveforms of the stimuli, f(t), and we only considered here the quadratic terms. The characteristic of the temporal filter was assumed to be flat. The nonlinear response then becomes $f^2(t)$.

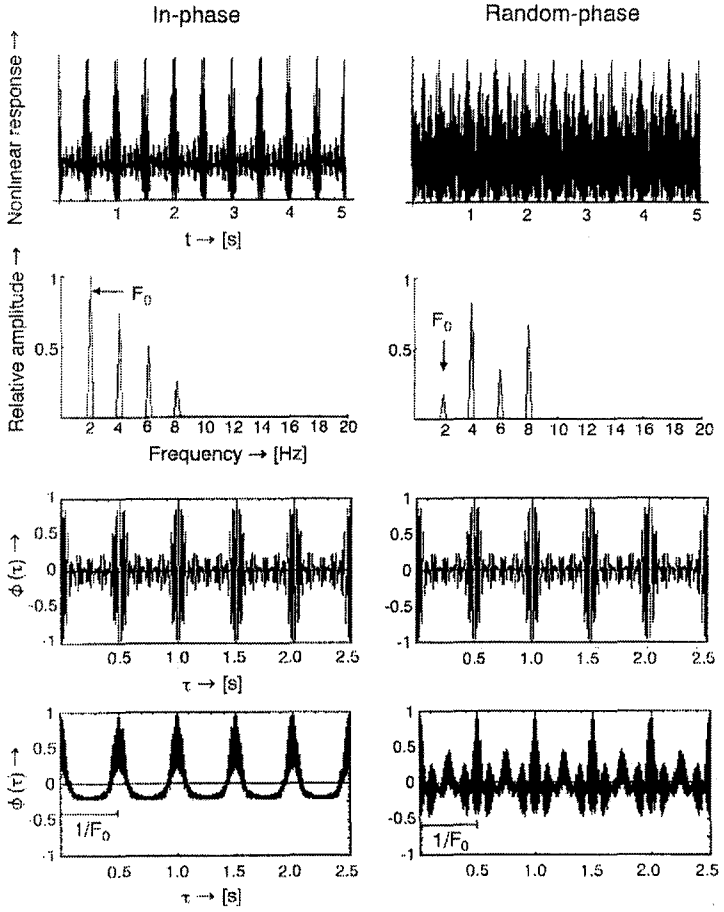

Figure 3.10 Top: nonlinear responses for the stimuli (see Figure 3.6) show that the nonlinearity expands the upper portion of the wave, creating distortion products at the low frequencies. Second row: spectra of the nonlinear responses contain F0 and multiples of it. Third row: the autocorrelation functions of the real stimulus waveforms are identical for different phase condition and have periodical peaks at F0. Bottom: the autocorrelation functions of the nonlinear responses are not identical, but have identical periodical peaks at F0.

The accelerating nonlinearity expands the upper portion of the wave (Figure 3.10, top), creating distortion products at low frequencies (Figure 3.10, second row). The relative size of low frequency distortion products depends on the relative phase of the compound waveforms. When the waveforms are "in-phase", the strongest low frequency distortion will produce a component at F_0 (Figure 3.10, second row, left). It is therefore no surprise that observers see F_0 most strongly with in-phase compound stimuli. The situation for the "random-phase" condition is more complicated because phase modulation itself adds sidebands or other frequencies to the stimulus. Fourier transform of the resulting signal confirms the severe reduction of the F_0 component (Figure 3.10, second row, right). In some extreme cases the signal lacks the F_0 component. Nonlinear distortion itself could not explain our results because the observers perceived F_0 most strongly for random-phase stimuli. Since the observers' task is to set the most prominent frequency, information about other frequencies is lost. Therefore it is not clear whether the observers detected distortion at several multiples of F_0. However, the important thing here is that they preferentially perceived F_0 as apparent rate of flicker with compound stimuli.

One possible operation that gives the phase-independent prediction for our empirical evidence is autocorrelation. The autocorrelation function is defined as the Fourier transform of the power spectrum, $P(\omega)$;

$$\Phi(\tau) = \int_{-\infty}^{\infty} P(\omega) e^{j\omega t} d\omega, \qquad (3.1)$$

Since only the power spectrum of the waveform appears in this definition, it is obvious that $\Phi(\tau)$ is not phase-dependent. Actually, the autocorrelation function of the real stimulus waveforms had identical profiles for both phase conditions used in the experiment (Figure 3.10, third row). Our result is consistent with the fact that the autocorrelation function has particular peaks corresponding to the F_0. It is possible to suppose a mechanism to extract a periodicity at F_0 from the peaks in the autocorrelation function of the stimulus.

In the experiment, the observers' response at F_0 was slightly affected (about 10 %) by phase (Figure 3.9), and some response was seen at multiples of F_0 with random-phase stimuli. Such a phase effect could be explained by the autocorrelation with nonlinearly distorted stimulus waveforms. The autocorrelation is not identical for

the in-phase and random-phase stimuli with the nonlinearity of the system. For the random-phase stimuli, the autocorrelation has several peaks corresponding to the multiples of F_0. These peaks might have the role of interrupting the visibility of F_0 for the random-phase stimuli and of producing exceptions to the rule that "the perceived flicker rate is the missing fundamental frequency". However, both waveforms have dominant peaks with the period of $1/F_0$ (Figure 3.10, bottom). These peaks correspond to the rates that the observers perceived most strongly.

3.3.5 Conclusions

1. The most frequently perceived flicker rate for the compound waveforms corresponds to the missing fundamental component, which is not included in the stimulus.
2. Even though the periodicity of the waveform is unclear, the fundamental frequency is perceived.
3. The phenomenon can be explained by assuming a process that detects periodical peaks in the autocorrelation functions for the nonlinearly distorted waveforms.

3.4 Proposed model of temporal vision

Summary

A qualitative model is proposed here to explain the perception of the missing fundamental frequency for complex flicker. The model consists of the envelope extraction mechanism from the complex signal and periodicity detection mechanism from the signal envelope. To explore the limitation of the model, thresholds of the modulation detection was measured for amplitude modulation (AM) flicker. It was found that the visual sensitivity for AM flicker is affected by the carrier and modulation frequencies. It was proofed that the subjective flicker rate of complex flicker is perceived only when the flicker is well above the threshold of envelope detection.

3.4.1 Introduction

As I stated in the previous section, it was found that the subjective flicker rate of complex waveform is the "missing fundamental" frequency, which is not included in the signal (Fujii et al, 2000). When the periodicity of the waveform is unclear due to the relative phase of the components, the fundamental frequency is also perceived. Furthermore, this signal caused some additional perception at multiple of the

fundamental. To explain this perceptual phenomenon, a process was assumed that detects periodical peaks in the autocorrelation functions for the nonlinearly distorted waveforms. Although the model seemed suitable for the data presented, the detail of the model has not been examined, and therefore it still holds several problems.

Proposed model in the previous section was quite rough. We introduced the mechanisms of visual nonlinearity and the autocorrelation, but did not examine the behaviors of these mechanisms in detail. More detailed model is proposed here and its limitation is tested. In the present model, the input signal is firstly processed by the nonlinearity, which works to extract the waveform envelope. Then, the periodicity of the extracted envelope is detected by the autocorrelation function. About nonlinear mechanism, how much effect would the visual nonlinearity have on flicker perception will be examined. In the previous experiment, it was not clear how far above flicker threshold was the stimuli used. We used complex flicker with nearly 100 % modulation. Because the visual system behaves linearly at lower contrast (e.g. modulation depth) and become nonlinearly at high contrast, we need to know this threshold to discuss the effect of nonlinearity on perception of complex flicker.

3.4.2 A qualitative model for flicker rate perception

To account for the experimental results about missing fundamental flicker, a model is proposed which consists of an initial linear filter followed by nonlinearity and autocorrelation mechanisms. Schematic illustration of this model is shown in Figure 3.11. In the experiment, subjects saw the complex flicker consisting five components and matched its flicker rate to that of sinusoidal flicker. Results showed that subjects could see the fundamental component, which was not included in the complex flicker. The initial linear filter determines the amplitude and phase of the response to the flicker component, but there is no response at the frequency of the fundamental component at this stage. After the complex flicker passes through the linear filter, some kind of nonlinearity works to the flicker. This nonlinearity creates a distortion product at the fundamental frequency and its multiples, according to the relative phase of components. This distortion product corresponds to the envelope of the flicker waveforms.

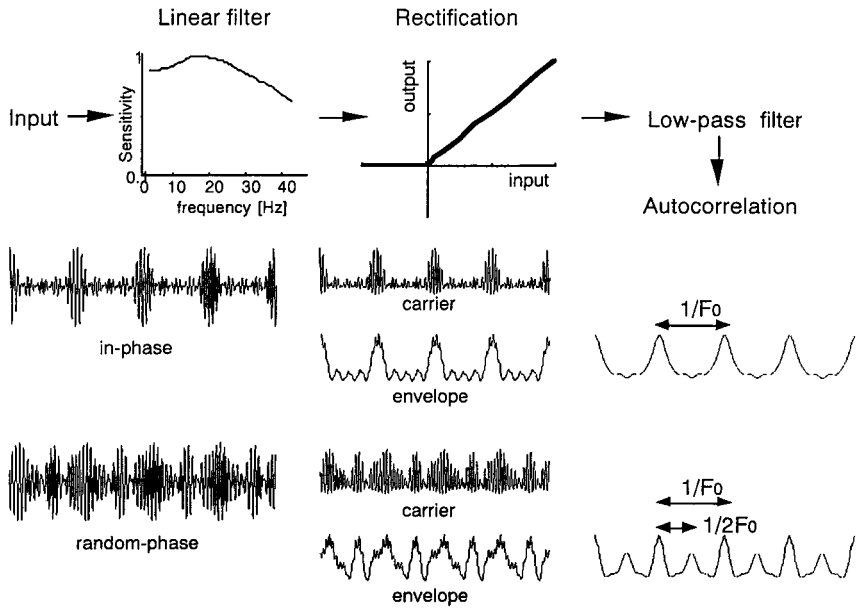

Figure 3.11 Illustration of a model for flicker rate perception in complex waveforms.

It is assumed that a matching of flicker rates is made when the period of the extracted envelope is equal to that of the sinusoidal flicker. To illustrate this, I show in Figure 3.11 that the input waveform is separated into its high frequency component (carrier) and low frequency component (envelope) by nonlinearity and low pass filtering. The carrier component has a distorted waveform due to nonlinear operation, but no assumption is made as to the exact shape of this component. Several researchers proposed the kind of nonlinearity, such as asymmetric nonlinearity (Kelly, 1978), compressive nonlinearity (Eisner et al, 1998), and expansive nonlinearity (Wu et al, 1996), but it is difficult to determine the shape of nonlinearity and its contribution to the flicker rate matching at this point. Rather, we assumed that the envelope component extracted by nonlinear mechanism has much effect on the perceived rate of complex flicker in our previous experiment.

As an operation for extracting periodicity included in the envelope component, we introduce the autocorrelation mechanism after early processing of the linear filter and some kind of nonlinear mechanisms. The envelope of the complex flicker contains periods of fundamental frequency and its multiples. The relative strength of these low frequency components is much affected by the relative phase of components in the complex flicker. Thus it is easy to find that fundamental frequency is perceived for "in-phase" stimulus, but difficult to suppose which periodicity is perceived for "random-

phase" stimulus (see Figure 3.11, middle). By calculating the autocorrelation function, we can see that both of in-phase and random-phase stimuli have dominant peaks at fundamental frequency that all the subjects perceived in the experiment. The decrease in the subjects' response at fundamental frequency for random-phase stimuli is also explained such that the autocorrelation for random-phase stimuli has minor peaks other than fundamental frequency, which corresponds to the multiple of it.

3.4.3 Test of the model: Envelope extraction by nonlinearity

A proposed model mainly consists of the following two processes: the envelope extraction by the nonlinearity, and the periodicity detection by the autocorrelation. Testable predictions of the model are that the flicker rate is perceived only when the envelope of the flicker is extracted, and that the perceived rate of the complex flicker is corresponded to the dominant peaks in the autocorrelation functions. Test of the former prediction is described as following based on the experimental results about threshold of modulation detection for the amplitude-modulation flicker. The purpose of this experiment is to examine the threshold of modulation detection for the complex flicker. In the previous experiment, it was not clear how far above flicker threshold was the stimuli used. Because the visual system behaves linearly at lower contrast (e.g. modulation depth) and become nonlinearly at high contrast, we need to know this threshold to discuss the effect of nonlinearity for the perceived flicker rate of complex waveforms.

Method

Apparatus and stimulus

Same apparatus was used as previous experiment except that the electric circuit was added to compensate the linearity of the LED emission at high frequencies. Linearity of the apparatus was confirmed by the measurement of the luminance waveforms of the stimulus with a luminance meter (TOPCON BM-8) with a response time of 1 ms.

Amplitude-modulation (AM) flicker was used as stimulus. AM waveform L(t) is defined as

$$L(t) = L_0[1 + K(1 + m\cos 2\pi f_m t)\sin 2\pi f_c t], \qquad (3.2)$$

where f_c is carrier frequency, f_m is modulation frequency, L_0 is mean luminance, K is carrier contrast, m is modulation depth, and t is time. Equation 3.2 can be expressed as

$$L(t) = L_0\{K[\frac{m}{2}\sin 2\pi(f_c - f_m)t + \sin 2\pi f_c t + \frac{m}{2}\sin 2\pi(f_c + f_m)t] + 1\}. \quad (3.3)$$

From equation 3.3, it is clear that the AM flicker consists of only three sinusoidal components of f_c (carrier), f_c-f_m, and f_c+f_m (sidebands), and modulation frequency f_m is missing. If the visual system behaves linearly, we could not see the modulation frequency. Contrary, if the modulation frequency is perceived, we can say that some kind of nonlinearity does exist in our visual system. The modulation frequency is contained in the signal envelope, which could be extracted by the nonlinear operation. To examine the effect of the carrier frequency and the modulation depth for visibility of the modulation frequency, modulation depth thresholds were measured for several carrier frequencies ranged from 5 to 50 Hz. Modulation depths were varied between 0.1 and 0.9 during the session. Tested modulation frequency was 2 Hz, and mean luminance was kept constant as 20 cd/m^2 during the experiment.

Procedure and subjects

Modulation detection thresholds were roughly determined by the method of limits as following. The amplitude-modulation flicker was presented successively separated by blank intervals of 1 sec. Each flicker interval lasted 4 sec with rise and fall time of 0.2 sec. During the session, the modulation depth was increased from 0.1 to 0.9 with the step size of 0.1. The subject's task was to specify the interval, at which the subject just perceived the modulation frequency. The procedure was repeated five times for each stimulus configuration. The author was tested as the subject.

Results and discussion

Determined threshold modulation depths are plotted in Figure 3.12. The abscissa represents the carrier frequency and the ordinate shows threshold modulation depth. Each plot represents the average value of the thresholds for five measurements. As the result shows, the sensitivity for the AM flicker peaked around 10 and 20 Hz, and decreased at high and low carrier frequencies for the modulation frequencies of 1, 2, and 3 Hz. For the 0.5 Hz modulations, the sensitivity was low pass characteristic. For high carrier frequencies, the sensitivity decreased more steeply and modulation frequencies could not be seen even for the 100 % modulation at 50 Hz.

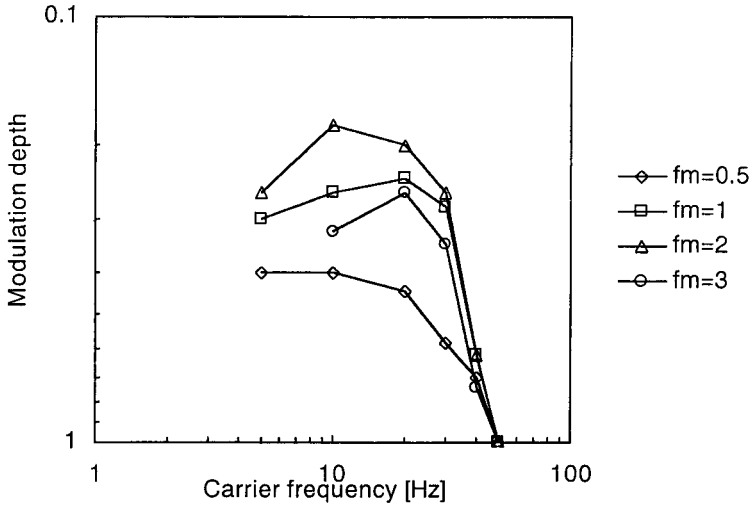

Figure 3.12 Thresholds of amplitude modulation flicker measured for carrier frequencies of 5, 10, 20, 30, 40, and 50 Hz, and modulation frequencies of 0.5, 1, 2, and 3 Hz.

Figure 3.13 re-plots the data of Fig. 3.12 as a function of modulation frequency. Sensitivity peaked at 2 Hz of modulation for the carrier frequencies below 40 Hz. Dashed line in Fig.3.13 is modulation threshold measured by Gorea, Wardak, and Lorenzi (2000) for large (30°) flickering stimuli set at an average luminance of 50 cd/m^2 with amplitude modulated temporal white noise (AM-TWN). This data is shown for comparison with the present one as the similar experimental data available. Even though the experimental conditions are different (e.g. stimulus size, mean luminance), Gorea et al's and present results show similar tendencies. Large difference is only seen at 0.5 Hz modulations. The most likely reason for this discrepancy relates to the different procedures used in the two experiments. While Gorea et al obtained thresholds with 2AFC methods, the present data was collected with criteria of perceived modulation frequencies. For the modulation frequency of 0.5 Hz, it is difficult to distinguish the flicker rate even when the modulation is perceivable.

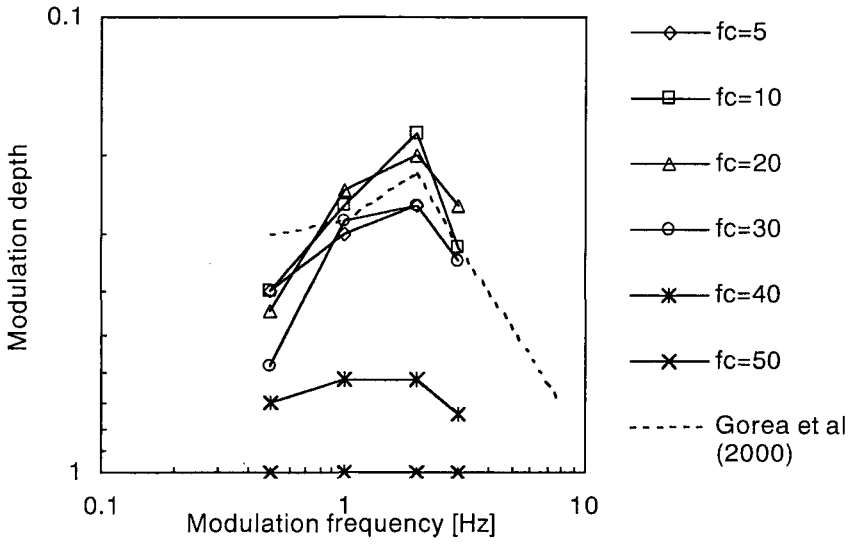

Figure 3.13 Thresholds of amplitude modulation flicker as a function of modulation frequency. For comparison with previous data, measured thresholds for amplitude-modulated temporal white noise (Gorea et al, 2000) are shown as dashed line.

From the observations above, it could be said that the sensitivity for the amplitude modulation flicker is affected by both of the carrier and modulation frequency. As for the effect of the carrier frequency, the present data is similar to the previous data of flicker brightness enhancement (Wu et al, 1996). Brightness enhancement increased with increasing modulation depth at all frequencies, and peaked around 16 Hz at high modulations. Wu et al's results suggest that the visual nonlinearity leading to brightness enhancement is dependent with temporal frequency. The present result is also interpreted as that the nonlinearity is frequency dependent, which works to extract waveform envelope and make it perceivable.

We can say that the previous experiment of complex flicker was conducted in the range well above threshold of modulation detection. To prove this, stimulus condition of the previous experiment was extended and reexamined. The informal observation by the author confirmed that the flicker rate of the fundamental component could not be seen for the stimuli near threshold. At high component frequencies above 40 Hz and at low modulation depth near threshold, the amplitude modulation could be just perceivable but flicker rate could not be distinguishable. Therefore, it can be

concluded that the flicker rate is perceived only when the envelope of the flicker is extracted and exceeded well above threshold.

3.5 Summary

In this chapter, I discussed the underlying mechanism for the temporal perception in vision. To address the problem, I focused on the fundamental properties of the temporal vision mechanism. Psychophysical experiment was performed on subjective flicker rates for complex waveforms. Results showed that human observers perceived a rate at the fundamental frequency, although the energy at this frequency was not included in the signals. It implies the existence of correlation mechanism in temporal vision.

4. Audition

4.1 Introduction

In this chapter, I present a series of studies on physical properties and psychological evaluations of sound. I use a method of autocorrelation in analyzing a sound signal. In addition, I use the interaural cross-correlation function to characterize the spatial properties of sound field. This is based on the autocorrelation function and the interaural cross-correlation function mechanisms, which have been proposed as a model of auditory system. Throughout the numerous studies on subjective responses in the sound field, Ando (1998, 2001) suggested that primary sensations of sound (loudness, pitch, and timbre), and spatial sensations such as localization, diffuseness, and apparent source width (ASW), are processed by the autocorrelation and the cross-correlation mechanisms respectively. Purpose of the present studies is to characterize the acoustical properties of sound field and to understand the mechanism of perception about such properties.

Loudness, pitch, and timbre are considered as primary sensations of sound stimulus. The loudness of a sound is the subjective attribute that is evaluated on a scale from "soft" to "loud". By the method of adjustment, loudness can be defined operationally as the amount of energy in a reference sinusoidal tone or broadband noise that is adjusted to be equally loud as a given stimulus. Loudness is generally correlated with the physical property power. Many sounds are perceived to have a pitch. The pitch of a sound is the attribute that allows it to be positioned on a continuum scale from "low" to "high". The pitch of a sound is defined as the frequency of a sinusoidal tone that is matched to the given sound by the method of adjustment. Perceived pitch of complex sound such as human voices and music instruments is correlated with the frequency of the lowest harmonic component (i.e., fundamental frequency). Also, we can evaluate subjectively the "pitch strength" or "pitchness" of a sound. A very pitchy sound, like a harmonic complex tone, has a clear pitch. For a less-pitchy sound, like that of repetition noise, the perception of pitch is not as immediate and seems weaker in strength. Comparing with loudness and pitch, the timbre of a sound is difficult to be correlated with particular physical property. There is no simple set of physical

properties that corresponds to timbre, and no clear operational definition. According to the American National Standards (ANSI) standard of 1960, timbre is that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar. Using this definition, it is necessary to extract from the mixture of sensations those that might be important.

In natural sound field we have spatial sensations such as direction and distance of sound source, diffuseness, and width of sound source, in addition to sound qualities described above. Information of direction and distance of sound source is important cue for us to navigate in the field. Spatial properties of diffuseness and source width are also important to define the quality of a sound field such as auditorium. To assess the psychological evaluation of sound in the real environment, it is necessary to investigate such spatial sensations as well as primary sensations.

4.2 Related work

Here, I will give a review of studies related to peripheral and central models of auditory system and subjective attribute of sound.

4.2.1 Model of auditory periphery

A starting point for modeling the auditory system is to look carefully at the biological origin to see what may be learned. Our understanding of the auditory periphery to the level of the hair-cell synapses and beyond has been increasing over the past four decades. Here, each of stages will be described briefly in accordance with literatures (Békésy, 1960; Pickles, 1982).

Outer and middle ear

Sound entering the outer ear is subject to a pressure gain at the tympanic membrane relative to the entrance to the ear canal; this pressure gain is maximal in the region between 2 and 5 kHz. The middle ear, which couples sound energy from the external auditory meatus to the cochlea, also has a band-pass pressure transfer function. But, in this case, the peak is near 1 kHz and has a much steeper slope at the low frequencies. These two functions could be combined to obtain an effect of outer and middle ear's frequency characteristic effect.

Cochlea filtering

The cochlea is probably the single most critical component in the auditory pathway. After coupling to the acoustic free-field via the outer and middle ears, the one-

dimensional sound pressure fluctuation is transmitted to the oval window, which starts a traveling wave down the spiraled transmission line of the cochlea. Because of the variation of the mechanical structure of the basilar membrane – the central division of the cochlea – and cochlea fluid, a continuous array of band-pass filters are formed. Fourier components in the pressure variation will travel some distance down the cochlea (further for lower frequencies) before reaching a point where the membrane is in resonance, causing a maximum in basilar membrane motion and the dissipation of the traveling wave. It mean that the cochlea performs a spectral analysis, converting the incident sound-pressure variation into motion at different places of the basilar membrane, with the place of motion encoding spectral location and amplitude of the motion indicating intensity.

Mechanical-neural transduction at hair-cell

The mechanical motion of the basilar membrane is converted to nerve firings via synapses at the base of the inner hair cell. This transduction forms a crucial component of any model of the auditory periphery, since it is the firing patterns on the approximately 30,000 fibers of the auditory nerve that comprise the description of sound used by higher levels of the brain. The inner hair cells have tiny embedded hairs (cilia) that bend when the basilar membrane moves relative to the cochlea fluid, and the cells emit electrical spikes with a probability that depends on the degree of deflection.

There are two properties of the inner hair cells that have particularly important effects on the signals transmitted to higher levels. First, the cells respond to cilia deflection asymmetrically (Hudspeth and Corey, 1977). This introduces a kind of nonlinearity to the input signal. Several complex models of inner hair cell function have been developed that are faithful to the nonlinear properties of mammalian inner hair cells (Hewitt and Meddis, 1991), but the simple half-wave rectification model is chosen by many researchers. Second, at low frequencies, the hair cells tend to fire at a particular phase of the signal – a process called *phase locking*. As the frequency of the input signal increases, phase locking begins to run out at about 1.5 kHz and disappears by 5 kHz (reference). This is because the capacitance of inner hair cells prevents them from sufficiently rapid changing in voltage. In absence of locking to the fine structure of the waveform, the hair cells lock to the signal's amplitude envelope (reference). This produces a reasonable envelope function at high frequencies.

4.2.2 Central mechanism of auditory system

Place model and temporal model of pitch

Models of central processing in the auditory system after auditory nerve have been controversial. Especially for pitch extracting mechanism, there are two main types of models; place models and temporal models. Place models of pitch hold that perceived pitch originates in spatial aspect of the initial pattern of excitation on the basilar membrane. Because primary auditory cells are distributed along the length of the membrane, and this tonotopic organization is maintained throughout the auditory pathway (Langner, 1997), there is good anatomical justification for the place model. A good example of the place model is Goldstein's *optimum processor model* (Goldstein, 1973). In this model, the instantaneous spectral peaks of a signal are extracted. Statistical analysis like a maximum-likelihood method is used to decide the fundamental frequency of the signal. Similar mechanisms have been proposed by Terhardt (1973), Wightman (1973a, b) and Hermes (1988), among others. There are several disadvantages of such models. First, they require more spectral resolution than is known to exhibit in cochlea. Signals with closely spaced spectral peaks can still give rise to a pitch, even though they are not resolved by the cochlea (Moore and Rosen, 1979). Second, place models could not explain certain pitch phenomena, such as iterated noise signals, that are spectrally flat (reference).

In a temporal model, pitch is extracted as the result of temporal processing and periodicity detection on each cochlear channel. Cochlea acts as a spectrum analyzer, but perceived pitch is not based on the spectral peaks from this representation. Rather, the band-passed signals that are output from cochlea are analyzed in time domain. Pitched signals have periodic fluctuations in the envelopes of the band-passed signals. These periodicities are viewed as origin of pitch. A variety of methods have been proposed for measuring the periodicity of sound. Licklider's duplex theory is the first temporal model of pitch (Licklider, 1959). He proposed a method based on a network of delay lines and coincidence detectors oriented in a two-dimensional representation (Figure 4.1). The first dimension corresponded to the distribution of frequency analyzed by cochlea, and the second to the delay time of autocorrelation over the range of periods that evoke a pitch sensation. This network is compiled as that, it calculates a running autocorrelation function in each frequency channel; the peak of the function within a channel indicated the primary pitch to which that channel responded. The information

from multiple channels would be integrated to give rise to a single sensation of pitch.

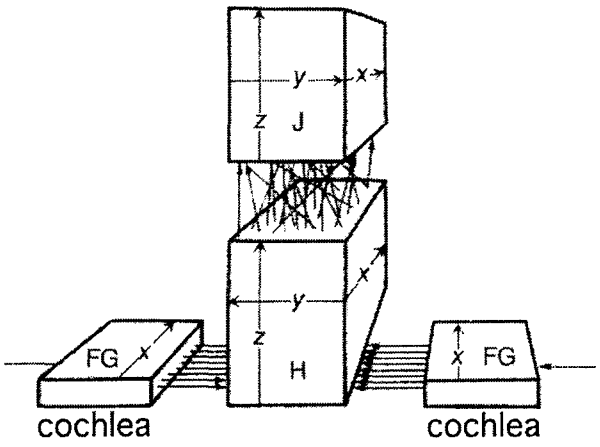

Figure 4.1 Duplex theory for pitch perception (Licklider, 1959).

Since Licklider's formulation, this method has been reintroduced several times. First, van Noorden (1983) proposed calculation for histograms of neural interspike intervals in the cochlear nerve. More recently, the model was reintroduced by Slaney and Lyon (1990), Meddis and Hewitt (1991), and others. Their method is called the autocorrelogram method of pitch analysis and is the preferred model today. Meddis and Hewitt (1991) specifically proposed that the integration of calculated autocorrelation functions (ACFs) in each frequency band produces a summary ACF (Figure 4.2).

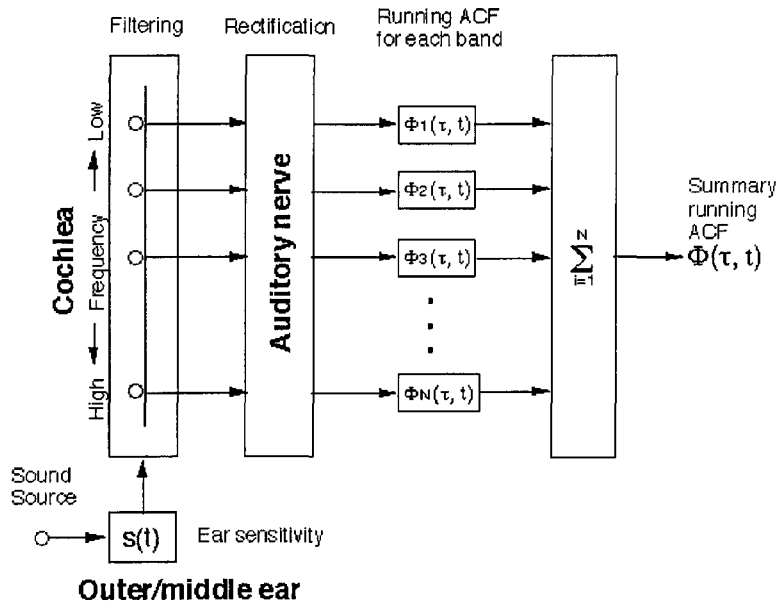

Figure 4.2 The summary ACF model of pitch detection, following Meddis and Hewitt (1991). After passed through the outer and middle ear, the input acoustic signal is separated into frequency subbands by a cochlear filterbank. A rectifying nonlinearity is induced in mechanical-neural transduction at auditory nerve in hair cell. Running ACF is calculated for the rectified waveforms in each frequency band and summed across frequencies to give a summary ACF.

They presented a number of analytic and experimental results showing that this model can quantitatively explain a wide range of pitch perception. An interesting study that can be connected to this pitch model is about neural mechanisms for auditory processing. Recent reports suggest evidence that the information needed for pitch perception is present in the temporal discharge pattern of auditory nerve. Cariani and Delgutte (1996a,b) measured the discharge pattern of auditory nerve fibers in cat in response to a variety of periodic complex sounds. They found that the inter-spike interval distributions for the auditory nerve resemble the summary ACF, and that the properties of these distributions corresponded to the large amount of psychophysical data on pitch perception.

Model of Binaural hearing

In binaural hearing (hearing with two ears), the auditory system is provided with differences in the input signals to the two ears that would not be available in monaural hearing. Consequently, binaural hearing offers a number of advantages over monaural

hearing (Blauert, 1983). The most obvious ones are following:

• Spatial hearing. The formation of an auditory space is improved with regard to the positions (azimuth, elevation, distance) of sound source and the spatial extents of the auditory events.

• Separation of sound signals from concurrent sound source. Simultaneous sound events are spatially separated. This improves the listeners' ability to concentrate on one and disregard the others (i.e. cocktail party effect).

• Suppression of the perceptual effects of reflected sounds. Coloration and reverberance as induced by reflected sounds are reduced. The dominant role of the first wavefront effect (precedence effect) is supported.

A starting point for modeling in binaural hearing is the model of Jeffress (1948) as depicted in Figure 4.3. The main body of this model is the estimation of interaural cross-correlation function (IACF). In this model, it is assumed that the IACF is computed by a delay-coincidence network. Each coincidence detector records coincidences of neural impulses from the two ears after a series of internal time delays. Cross-correlation is an adequate means of analyzing interaural differences in arrival time. Following cues are extracted from IACF: determination of the azimuth of sound source, detection and identification of echoes, and estimation of the amount of spatial impression. Jeffres model was extended by various researchers to cover a broad range of binaural psychophysical phenomena and to include physiological mechanisms (e.g., Lindemann, 1986a,b; Blauert, 1983).

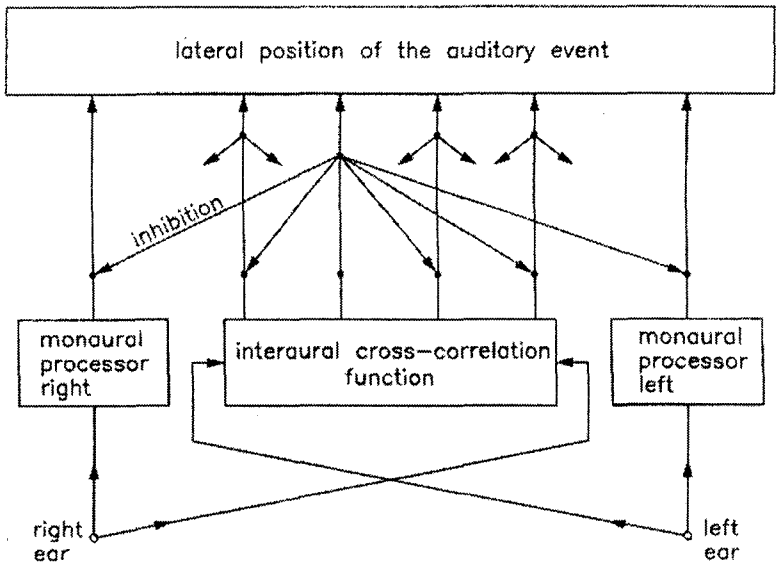

Figure 4.3 IACF model proposed by Jeffres (1948).

4.2.3 Subjective attributes of sound

Loudness

Loudness is generally correlated with sound pressure level (SPL) of sound, but there are other factors affecting perceived loudness. For example, the variation in frequency distribution for broadband noise, bandwidth and center frequency for narrowband noise, and duration of sound causes different loudness for sound with the same intensity. As for these effects on loudness, there is an extensive work by Zwicker and others. Zwicker et al. (1957) firstly reported that when the spacing between groups of pure tones is increased, perceived loudness remains constant until it reaches a critical point, after which the loudness increases. The same effects occurred when the bandwidth of a noise sound was increased. The bandwidth at which loudness summation begins was approximately the same as the critical band determined by methods of masking and threshold estimation. After that, their numerous data and established modeling led to the formulation of the "critical band theory", which become the basis of today's standard loudness meter.

However, there are still a lot of controversies about this theory regarding the basic theory involved. They concerned with the effect of tonal components within the spectrum. Hellman (1982, 1984) investigated the perceived loudness of tone-noise complex in relation to the frequency and amplitude of tone. Their results suggest that

the extent of the increments and decrements in perceived loudness depends on the overall SPL and the interaction between a specific tone frequency and noise spectrum. Tonal components might be contributed more to overall loudness than predicted by a method of loudness summation based on the critical band theory. Merthayasa and Ando (1996) found that loudness increases for band pass noise with its bandwidth narrower than the critical band. Their result contradicts to Zwicker's theory as shown in Figure 4.4. Filtering by a sharp filter as they used, ratio of the tonal components increases when bandwidth decreases. Their result also indicates the effect of the tonal components on loudness estimation. Although a tone corrections for loudness is warranted for certain tone-noise configurations, none of the proposed calculation procedures consider all the valuables relevant to perceived loudness.

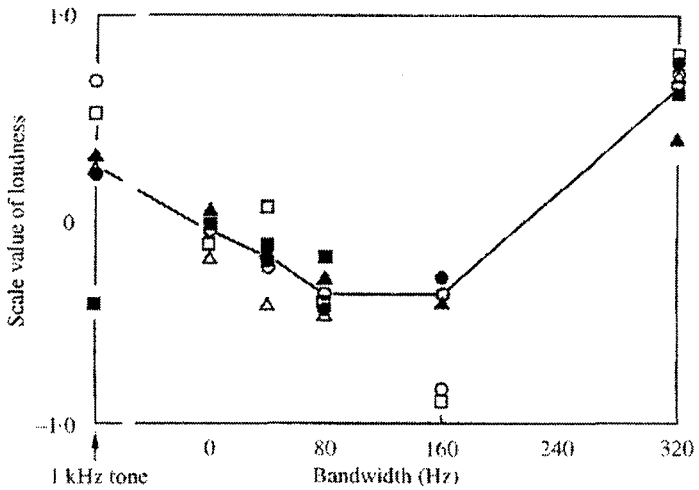

Figure 4.4 Loudness of bandpass noise measured by Merthayasa and Ando (1996).

Pitch and pitch strength

Complex tones, such as vowels or musical tones, consist of a series of sinusoidal components (harmonics) whose frequencies are integer multiples of the repetition rate of the complex, also called as the fundamental frequency (F_0). Complex tones are usually heard as having a single pitch corresponding to F_0. When there is little or no energy present at F_0 (i.e. missing fundamental), we also hear the pitch corresponding to that F_0. The pitch of such a signal is called "periodicity pitch," "residue pitch," "low pitch," or "virtual pitch" (reference). These pitches associated with multi-component spectral patterns can be contrasted with "place," "spectral," or "pure tone" pitches that

are associated with individual frequency components.

In addition to complex tones, random noise can also be manipulated to elicit pitch sensations. For example, "regular interval noise" (RIN) is a sound derived from random noise, which has temporal regularity in the waveform. The most common RIN is "ripple noise," which is produced by delaying a portion of random noise and adding it back to the un-delayed version. The normalized ACF of ripple noise has a single peak at the correlation lag corresponding to the delay. By controlling gain and number of iteration for adding process, we can manipulate the regularity of the signal and consequently the height of this ACF peak. Yost (1996a,b) found that perceived pitch strength of his iterated ripple noise corresponded to the height of the ACF peak. It means that we can detect the degree of regularity in sound and then perceive the strength of pitch. As another example, Ando and his colleagues examined pitch strength of "complex noise" which consists of harmonically related narrow-band noises without fundamental component (Ando et al., 1999). As shown in Figure 4.5, the height of the ACF peak was controlled by the bandwidth of each component. Their result also suggested that the pitch strength is corresponded to the height of the ACF peak (Figure 4.6).

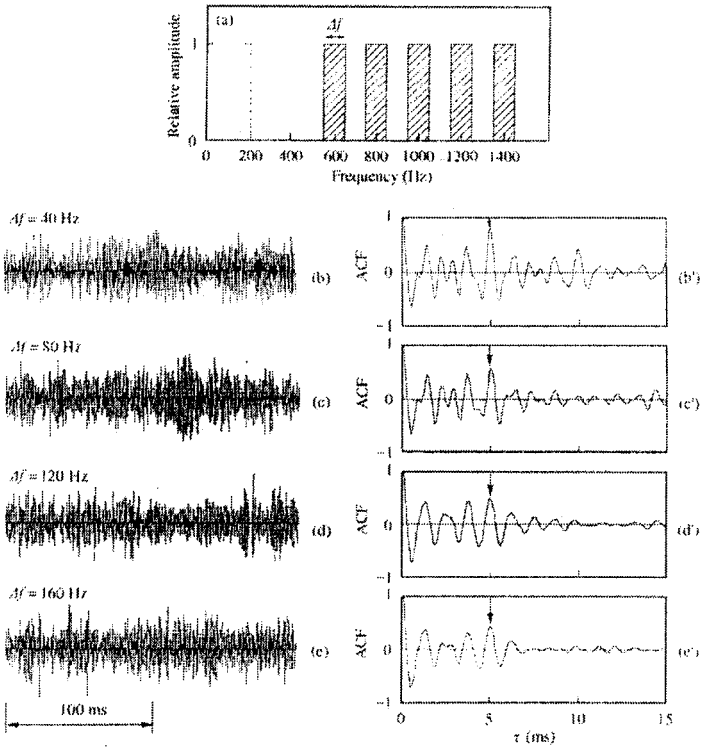

Figure 4.5 Complex noise" containing the center frequencies of 600, 800, 1000, 1200, and 1400 Hz. (a) its fundamental frequency is around 200 Hz. The Δf represents the bandwidth. (b) - (e) Waveforms of the four complex noises , and (b') - (e') their ACFs. From Ando et al. (1999).

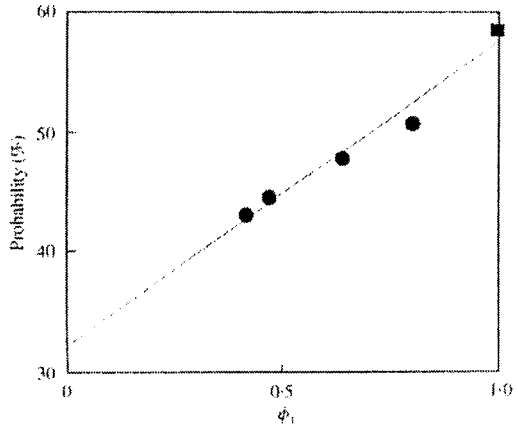

Figure 4.6 Relationship between the ACF peak ϕ_1 and probability of the perceived pitch within 200 ±16 Hz ($r = 0.98$, $p < 0.01$). From Ando et al. (1999).

Timbre

Researchers have long been interested in identifying perceptual dimensions of timbre of sound. Multidimensional scaling (MDS) is a general method for finding underlying dimensions of timbre perception. A typical MDS study of timbre begins with a collection of sounds (for simplicity, complex tones are often used) with differences of pitch, loudness, and duration minimized. Subjects are asked to rate either the similarity or dissimilarity of each pair of sound stimuli. From these judgments, a low-dimensional arrangement of the stimuli is found as that best fit the subjects' ratings. If the set of stimuli has an underlying dimensional structure, the dimensions of the arrangement can be interpreted in terms of perceptual/physical attributes.

Miller and Carterette (1975) made similarity judgment for complex tones simulating instrumental tones. They found that the fundamental frequency was the most salient factor for similarity judgment. Other factors were found as time envelope of stimuli and number of components. Grey (1977) performed similar experiments by using computer-synthesized instrument tones. Salient dimensions were interpreted in terms of the spectral energy distribution and the temporal structures in the transients. Considering these two results and others, timbre might be dependent on both the spectral structure and temporal structure. Especially for the spectral structure, the average spectral centroid, which correlates strongly with *sharpness* or *brightness* of sound, and the consonance, which correlated with *clarity* of sound is consistently found to be principal (Bismarck, 1974; Ohgushi, 1980).

Spatial sensations (localization and spatial impression)

The most tempting field of application for binaural auditory models concerns the ability of binaural hearing to process signals from different sources selectively, and to enhance one of them with regard to the others. Lateral displacement of sound source is perceived when an interaural time difference (ITD) or interaural level difference (ILD) is present. Subjective lateral displacement is assumed to correspond to the location of the maximum peak of the IACF.

Other spatial sensations, such as subjective diffuseness and apparent source width (ASW) have been extensively studied in the field of room acoustics. Keet (1968) reported that the absolute ASW for fixed listening level is highly correlated to the cross-correlation function measured by two microphones with angle of 90° between them. Barron (1971) described the importance of early lateral reflections for spatial

impression, and claimed that the degree of spatial impression is related to the ratio of lateral to non-lateral sound arriving within 80 ms of the direct sound. Damaske and Ando (1972) defined IACC as the maximum value of the interaural cross-correlation function lying within the possible maximum interaural time delay (between -1 ms and + 1 ms). They reported that IACC corresponds to the subjective diffuseness of sound field. Later, it was determined that IACC is a significant factor for the subjective preference of sound fields. Schroeder et al. (1974) conducted the subjective preference tests. They found two significant factors, namely, reverberation time, and IACC. From the systematic investigations in simulated sound fields, Ando (1985) found the four orthogonal factors (listening level: LL, initial time delay gap: Δt_1, subsequent reverberation time: T_{sub}, and IACC). Beranek (1996) also proposed that IACC is one of his proposed six significant factors. As another spatial factor, Barron and Marshall (1981) proposed Lateral Energy Fraction. But the reflection from 90° in the horizontal plane did not always have an advantage for increasing the subjective diffuseness.

4.3 Acoustical properties of aircraft noise measured by temporal and spatial factors

Summary

Acoustical properties of aircraft noise were investigated by means of temporal and spatial factors in sound fields based on the model of auditory-brain system. The model consists of the autocorrelation and crosscorrelation mechanisms for sound signals arriving at two ears and the specialization of human cerebral hemisphere. There are four temporal factors extracted from the autocorrelation function (ACF); 1) sound energy $\Phi(0)$, 2) effective duration of ACF, τ_e, 3) delay time of the first peak, τ_1, and 4) its amplitude ϕ_1. From the interaural cross correlation function (IACF), three spatial factors are extracted as, 1) magnitude of the interaural crosscorrelation, IACC, 2) interaural delay time at IACC, τ_{IACC}, and 3) width of the maximum peak of the IACF, W_{IACC}. It is found that the acoustical properties are well represented by the factors extracted from the ACF and the IACF.

4.3.1 Introduction

This paper describes the acoustical properties of aircraft noise in terms of its temporal and spatial factors. Aircraft noise disturbs peoples' daily lives and sometimes causes serious problems such as hearing loss or has an adverse impact on the growth of unborn

babies, infants, and children (Ando, 1977; Ando and Hattori, 1970, 1973, 1974, 1977a, b). A lot of effort has been spent on noise research and noise reduction technologies (Stephens and Cazier, 1996). Significant progress has been made on reducing noise level but a big problem remains. Noise has only been evaluated by statistic sound pressure level (SPL), but perceived acoustical properties have not been considered sufficiently (Research Committee of Road Traffic Noise in Acoustical Society of Japan, 1999). In particular the relationship between physical properties and psychological affects is not clear. For example, a sound may exist that has a SPL below standards such as EPNL or WECPNL, but that is perceived to be noisy in a given situation. Such an annoyance may be related to primary auditory sensations (pitch, loudness, and timbre) based on the mechanisms in the human auditory-brain system (Ando et al., 1999).

The most plausible mechanism in the auditory system consists of autocorrelators and a crosscorrelator for analyzing sound signals arriving at both ears (Ando, 1998). Perceived pitch and its strength of complex tones or complex noises are expressed by the first peak in the autocorrelation function (ACF) of the signal (Sumioka and Ando, 1996). Loudness is also related to a factor of the ACF, τ_e (Merthayasa and Ando, 1996), not only to the SPL. In addition, spatial properties are important for noise evaluation. Noise sources are usually not fixed, but move spatially. We hear a different sound quality when the noise source is coming or going away. Information on location or direction of the sound source, subjective diffuseness, and apparent source width (ASW) can be expressed by the factors extracted from the interaural crosscorrelation function (IACF) (Ando, 1998; Sato and Ando, 1998). To specify such spatial characteristics, binaural measurements were conducted.

4.3.2 Method

Measurement procedure

Measurements were taken outdoors near the flight course of Kansai International Airport on January 12, 2000, and near Osaka International Airport on December 13, 1999. Measurement locations are illustrated in Figure 4.7.

For measurement of noise along the flight course of Kansai Airport, a dummy head was set near coast. This location is 20 km southwest from 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 the flight data from the

airport. It was cloudy and windless on the ground level during the measurement. The average temperature for the day was 12 °C. Ambient noise level in this area was 43 ± 2 dB.

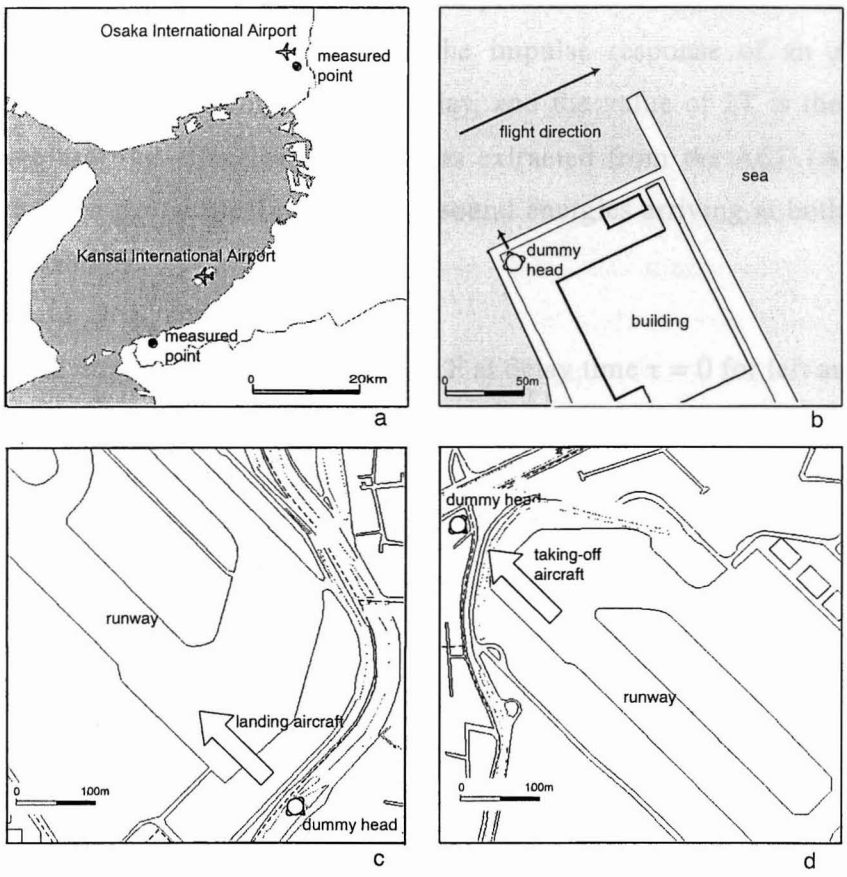

Figure 4.7 (a) Location of two airports and the measurement points. Configuration of each measurement point for (b) level flying, (c) landing, and (d) takeoff.

At the Osaka Airport, two locations were chosen close to a runway to measure the noise from aircraft landing and taking off. The distances between the runway and each measuring point was about 100 m. Ambient noise level in this area was higher because of road traffic (60 ± 2 dB). It was cloudy and windless on the ground level. Temperature was about 10 °C during the measurement. Noise signals were received by two half-inch condenser microphones set at both ear positions of a sphere representing a human head. This dummy head is made of 20-mm-thick styrofoam with a diameter of 200 mm. Microphones were set at 1.5 m above the ground.

Analysis of acoustical factors

An autocorrelation function (ACF) is defined by

$$\Phi_p(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} p'(t)p'(t+\tau)dt, \tag{4.1}$$

where p'(t) = p(t)*s(t), in which p(t) is sound pressure and s(t) is ear sensitivity. For practical reasons, s(t) may be chosen as the impulse response of an A-weighting network. The value τ represents the time delay, and the value of 2T is the integration interval. There are four significant parameters extracted from the ACF (Ando, 1998). The first factor is a geometrical mean of the sound energies arriving at both ears, $\Phi(0)$, which is expressed by,

$$\Phi(0) = \left[\Phi_{ll}(0)\Phi_{rr}(0)\right]^{1/2}, \tag{4.2}$$

where $\Phi_{ll}(0)$ and $\Phi_{rr}(0)$ are the normalized ACF at delay time $\tau = 0$ for left and right ears. Sound pressure level is obtained as SPL = 10 $\log_{10}\Phi(0)$. The second factor is the effective duration of the normalized ACF, τ_e, which is defined by ten-percentile delay of the normalized ACF, representing repetitive features or reverberation contained within the signal itself. The third and fourth factors are the delay time and the amplitude of the first peak of the normalized ACF, τ_1 and ϕ_1. These two factors are closely related to the pitch sensation (Sumioka and Ando, 1996).

For specifying the spatial characteristics of sound signals, three factors were extracted from the interaural crosscorrelation function (IACF). The crosscorrelation function between the sound signals at both ears $f_l(t)$ and $f_r(t)$ is given by,

$$\Phi_{lr}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} f_l'(t)f_r'(t+\tau)dt, \tag{4.3}$$

where $f_l'(t)$ and $f_r'(t)$ are approximately obtained by signals $f_{l,r}(t)$ after passing through the A-weighting network, as in equation (1). Normalized IACF is defined by

$$\phi_{lr}(\tau) = \frac{\Phi_{lr}(\tau)}{\sqrt{\Phi_{ll}(0)\Phi_{rr}(0)}}, \tag{4.4}$$

where the values of $\Phi_{ll}(0)$ and $\Phi_{rr}(0)$ represent the sound energies arriving at left and right ears. The denominator means the geometrical mean of the sound energies arriving at both ears. The magnitude of IACF is defined by,

$$IACC = \left|\phi_{lr}(\tau)\right|_{\max}, \quad |\tau| \le 1 \text{ ms}. \tag{4.5}$$

The value of IACC represents the degree of similarity of sound waves arriving at each ear. This is a significant factor in determining the degree of subjective diffuseness in the

sound field. As IACC decreases the subjective diffuseness increases.

The interaural time delay is defined as τ_{IACC} at which the IACC is decided. It represents the horizontal sound location or direction, and the balance of the sound field. When τ_{IACC} is zero, the front-sound-source image and a well-balanced sound field are perceived. The width of the maximum peak of IACF, W_{IACC}, is defined by the delay time interval 10% below IACC. It is worth noticing that the apparent source width (ASW) could be evaluated by IACC and W_{IACC} (Sato and Ando, 1998).

Conditions for calculating acoustical factors

Aircraft noise lasts for certain duration. Its duration depends on the distance between the receiver and the planes or speed of the planes. Noise was measured 10 s for aircraft landing and 20 s for taking-off. During level flying at high altitude, noise lasted about 60 s. Although the sound pressure level fluctuated throughout the flight, the mean level was constant. Therefore, the measurement time for one session was set to 10 s for level flying aircraft with center of maximum SPL.

As sound signals vary continuously, the acoustical factors described above should be calculated in every certain duration with short interval. In the case of music sources, the integration interval (2T in equation 1) is between 2 to 5 s. This length is based on the theory of "psychological present", which states that humans perceive successive events as one thing (Fraisse, 1982). But in calculating ACF to describe a single syllable for Japanese speech, a much shorter integration interval (30 ms) is used because the speech signal varies in very short time (Shoda and Ando, 1998).

To capture the correct properties of aircraft noise, integration interval for ACF and IACF has been examined. Figures 4.8 (a) and (b) show examples of measured SPL for two types of signals with different τ_e (Figure 4.8, top) integrated for three different intervals. It is clear that an interval of 1 s is too long to capture the fluctuation of sound properties. Such a variation could be caught by 0.25 s or 0.5 s integration. The properties throughout the measuring time are the same for both intervals, but a finer variation could be measured in the case of 0.25 s. Listening to the noise with long τ_e (minimum value: 20 ms), these fine variations could not be heard. For the sound with short τ_e (minimum value: 10 ms), on the other hand, 0.25 s integration matches the actual sound fluctuation. Mouri et al. (2001) reported that the integration interval should be set as $2T \approx 30\ (\tau_e)_{min}$. In this case, recommended 2T is 0.6 s and 0.3 s for the signal with $(\tau_e)_{min}$ of 20 ms and 10 ms, respectively. In the present study, integration interval

was chosen as 0.5 s for signals with $(\tau_e)_{min}$ = 20 ms, and 0.25 s for signals with $(\tau_e)_{min}$ = 10 ms.

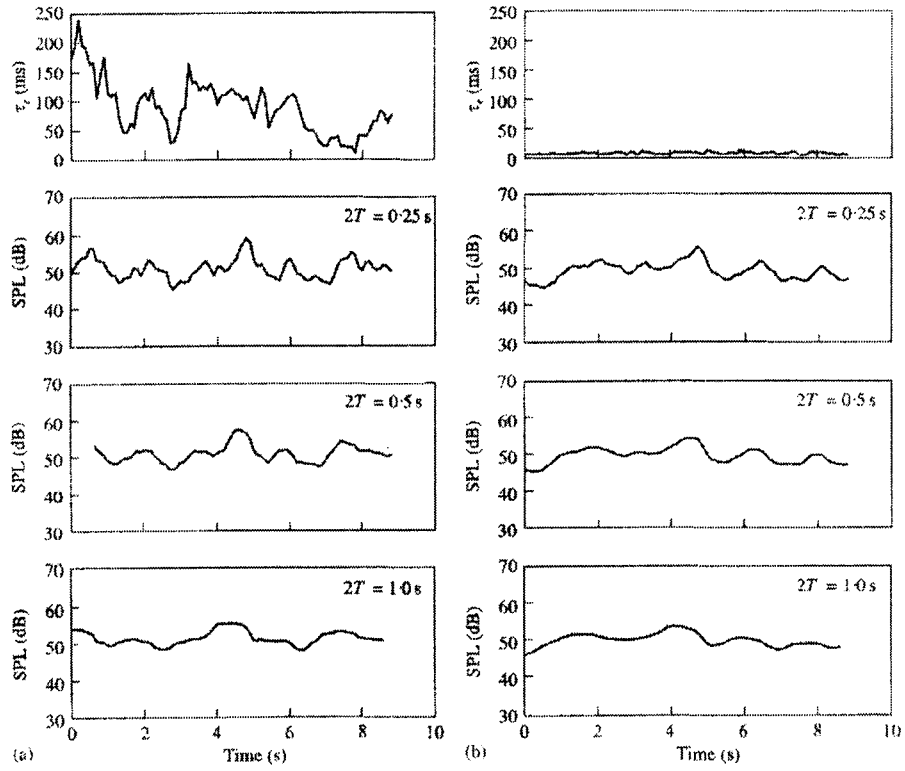

Figure 4.8 Examples of measured SPL for two different types of noise signals with a) $(\tau_e)_{min}$ = 20 ms and b) $(\tau_e)_{min}$ = 10 ms with three different integration intervals, from the second row, 0.25, 0.5, and1.0 s.

4.3.3 Results and discussion

Temporal factors extracted from ACF

An aircraft flying overhead produces a noise on the ground, which rises above the ambient level, reaches a maximum when the aircraft is approximately overhead, and then decreases again below the ambient level. The properties of the aircraft noise vary throughout the flight. The typical case is one in which the noise is predominantly high frequency while the aircraft is approaching and is predominantly low frequency after the aircraft has passed over and is receding. Such characteristics are clearly represented by the factors from the ACF as shown in Figures 4.9 (a) - (c) for landing condition.

Measured SPL is shown as a function of time; at t = 5.0 s the aircraft was directly overhead. The duration above the ambient level was about 10 s. The delay time

and the amplitude of the first peak in ACF, τ_1 and ϕ_1, represent the perceived pitch and its strength. The reciprocal of τ_1 corresponds to the perceived pitch. Results indicate that the perceived pitch varied throughout the flight. As the aircraft approached, τ_1 was about 1 ms with the value of ϕ_1 increasing. The strongest pitch of 3300 Hz was perceived when the aircraft passed overhead, at which the value of τ_1 was 0.3 ms. Such a strong tonal component is emitted from fan exhaust. After the aircraft passed over, τ_1 value increased and ϕ_1 value decreased simultaneously, indicating that the noise was dominated by the lower frequency components produced by jet exhaust.

Power spectra and the ACF measured at t = 1.0, 5.0, and 7.0 s are illustrated in Figures 4.9 (b) and (c). They show that τ_1 and ϕ_1 represent the properties of aircraft noise clearly; at t = 1.0 s there is a small peak around 1000 Hz, which is perceived as a noise with a weak pitch; at t = 5.0 s there is a high frequency component at 3300 Hz perceived as a tonal sound; and at t = 7.0 s the strong peak disappears and the lower frequency components increases below 500 Hz, which is perceived like white noise.

For the same type of aircraft during taking off, the duration above ambient level was approximately 20 s, longer than that of landing condition. The acoustical properties of taking-off aircraft were somewhat different from those of landing aircraft. The ϕ_1 value was always below 0.2, which means the high frequency tonal components were lessened and low frequency components were pronounced. This is possibly because the aircraft is at a higher altitude and the engine power is higher. Thus, high frequency components attenuate and more jet noise is produced.

Figures 4.10 (a) show the measured factors for the aircraft during level flying overhead at an altitude of about 1 km, and measured power spectra and normalized ACF are shown in Figures 4.10 (b) and (c). The noise for level flying aircraft was classified into two typical cases. The SPL throughout the flight fluctuated in the same manner, but the values of τ_1 and ϕ_1 were extremely different for each case. The value of τ_1 for the two cases was almost the same (mean value: 3.06 and 2.45 ms), but the ϕ_1 value for one case varied dramatically throughout the flight. At times with high ϕ_1, a tonal sound was heard and its pitch strength fluctuated in relation to SPL. In other words, when the noise contained a strong tonal component, the total SPL increased. This phenomenon may be related to diffusion, air absorption, or the scattering reflection of sound caused by air conditions such as wind or clouds.

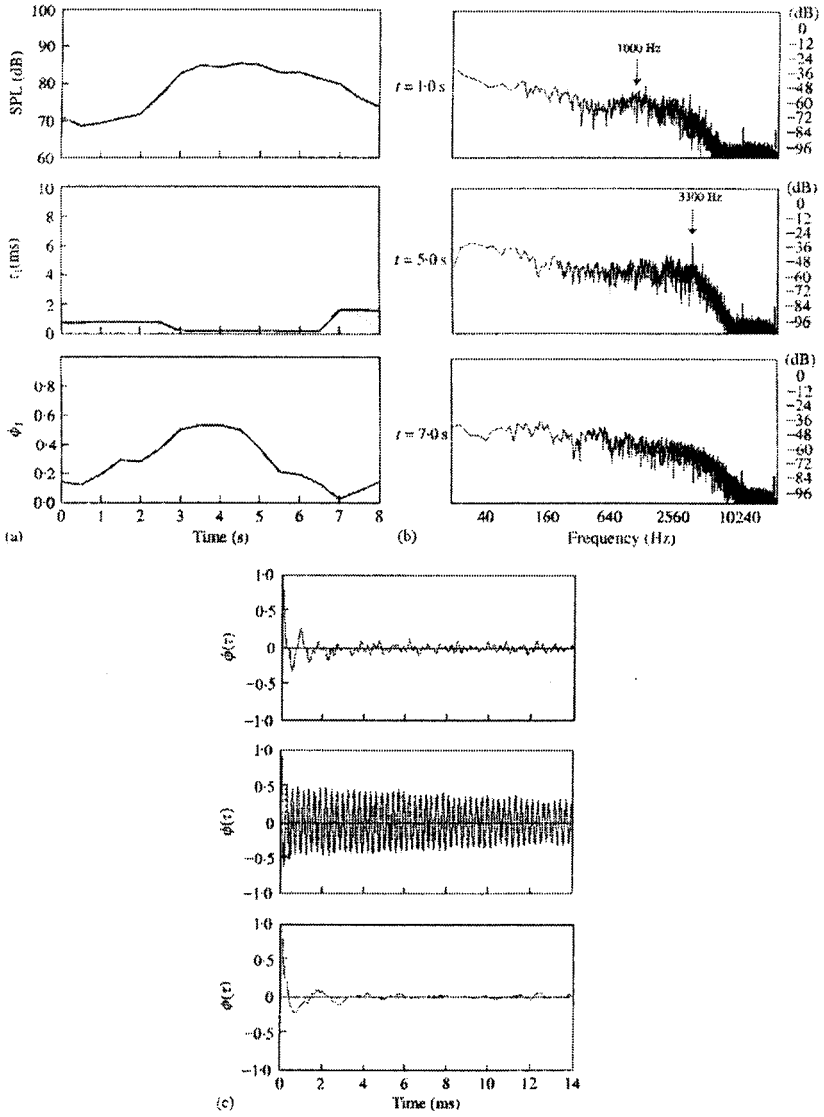

Figure 4.9 Results for a) Measured SPL (top), τ_1 (middle), and ϕ_1 (bottom) for landing aircraft as a function of time, b) power spectra and c) normalized autocorrelation functions at t = 1.0, 5.0, and 7.0 s.

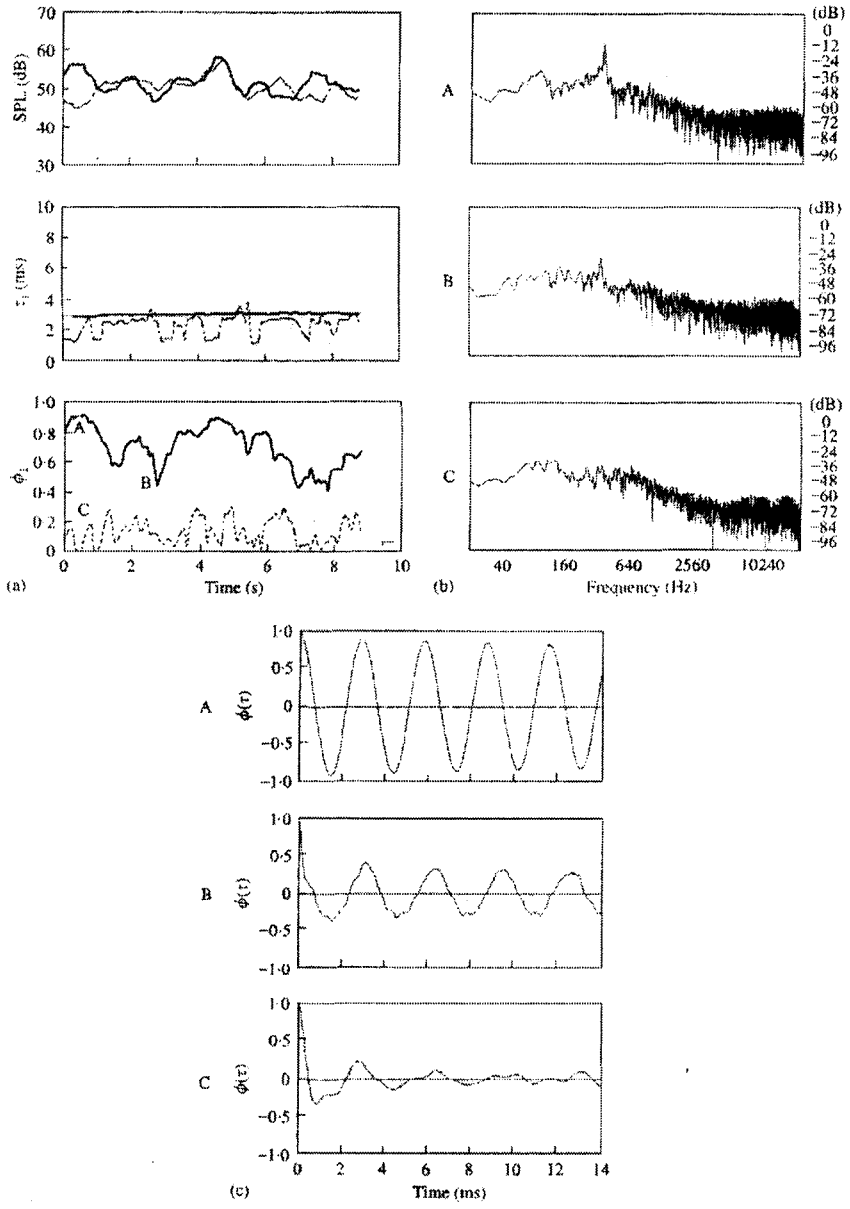

Figure 4.10 (a) Measured SPL, τ_1, and ϕ_1 for level flying aircraft. (thick line), tonal noise, and (dotted line), un-tonal noise. (b) examples of spectrum and (c) normalized ACF, measured at A, B, and C.

Spatial factors extracted from IACF

The normalized interaural crosscorrelation function (IACF) is shown in Figure 4.11 (a) for landing, taking off, and level flying conditions. The values of IACC, τ_{IACC}, and W_{IACC} were found from them. Measured IACC is shown in Figure 4.11 (b) as a function of time. For the landing and taking off conditions, the IACF had a strong peak at $\tau \approx 0$,

meaning that the direction of the noise source is perceived clearly. The value of IACC decreased when the aircraft passed overhead for landing, possibly because the noise was dominated by the high frequency component produced by fan exhaust. The value of W_{IACC} is dependent on the dominant frequency of the sound. For the taking off condition, W_{IACC} was large because of the low frequency components.

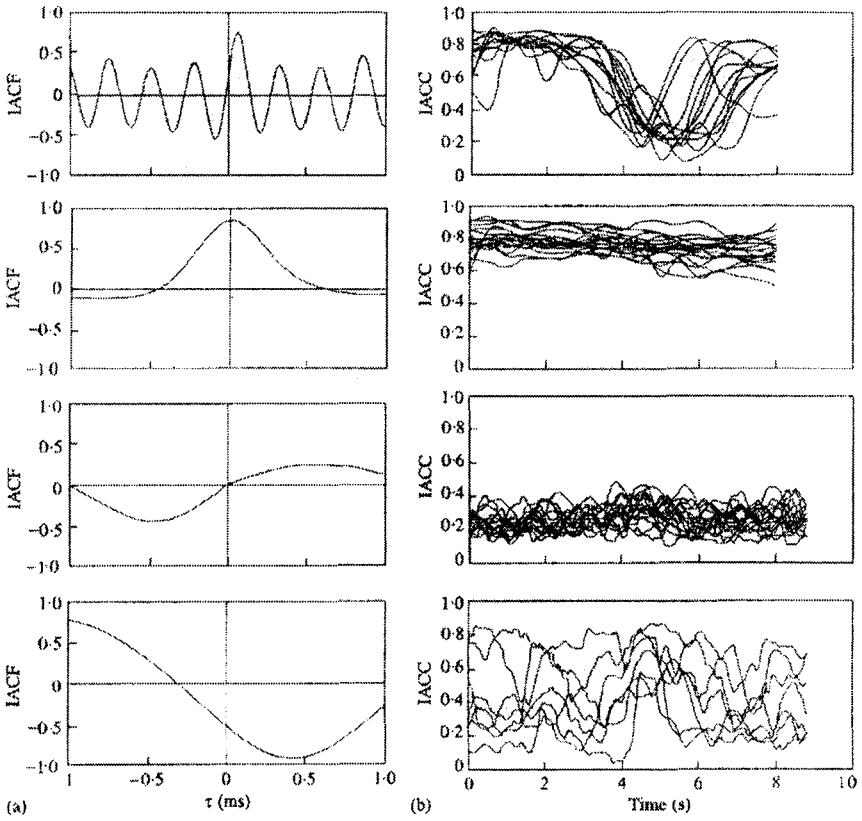

Figure 4.11 (a) Normalized interaural crosscorrelation function and (b) measured IACC as a function of time for the conditions of landing (top), takeoff (second row), level flying 1 (third row), and level flying 2 (bottom).

The value of IACC was generally small for the level flying condition. In this case subjective diffuseness became high or no spatial impression was perceived. For flying aircraft at high altitude, the sound signals may come from various directions because of diffusion or the scattering reflections by clouds. It was also found that the value of IACC increased when the noise was dominated by tonal components. It is possible that such a tonal component reached the ground from the aircraft directly.

The value of τ_{IACC} for landing and taking off aircraft was always close to zero,

which means that the sound source is perceived for a frontal direction. On the contrary, for level flying condition the value of τ_{IACC} could not be calculated in many cases because the peak of the IACF shifted over 1 ms. The value of W_{IACC} was also larger for level flying than for landing or takeoff conditions. As a result, information about sound source direction may be lost and apparent source width (ASW) may become wider.

4.3.4 General discussion

It was found that the measured temporal and spatial factors represent the acoustical properties of aircraft noise well. Although the results have already been reported such that the dominant frequency component varies throughout the flight for landing aircraft (Raney and Cawthon, 1979), the present ACF analysis represents these properties simply and clearly.

For the level flying condition, aircraft noise was classified as either tonal noise or un-tonal noise with low frequency components. For the tonal noise, the value of τ_e becomes longer because of the repetitive component of the signal. It has been reported that loudness increases in proportion to the value of τ_e (Merthayasa and Ando, 1996). It is possible that the aircraft noise including tonal component is perceived louder than un-tonal noise. Psychological experiments should be performed to examine the relationship between loudness and the value of τ_e for aircraft noise.

The spatial properties of aircraft noise are also interesting. It was found that the value of IACC decreases and W_{IACC} increases for the level flying condition. This phenomenon may be related to the scattering reflection by clouds in the sky. The aircraft noise for level flying condition may cause higher subjective diffuseness and wider ASW. Psychological tests on spatial impressions also need to be performed to examine the correspondence between the measured physical properties and psychological perceptions or evaluations for the aircraft noise.

4.4 Physical properties and annoyance of traffic noise

Summary

Temporal and spatial factors of traffic noise were analyzed based on a model of an auditory-brain system. This model consists of the autocorrelation and cross-correlation mechanisms for sound signals arriving at two ears, and takes account of the specialization of left and right cerebral hemispheres. There are three temporal factors

extracted from the autocorrelation function (ACF): (1) the effective duration of ACF τ_e, (2) the delay time of the first peak τ_1, and (3) its amplitude ϕ_1. The four spatial factors extracted from the interaural cross-correlation function (IACF) are (1) listening level, (2) magnitude of the interaural cross-correlation IACC, (3) interaural delay time τ_{IACC}, and (4) width of the maximum peak of the IACF W_{IACC}. Annoyance of the traffic noise reproduced in an anechoic chamber was evaluated by nine subjects in a paired-comparison test. Even though the listening level was constant for all the stimuli, the scale value of annoyance was much different for each vehicle noise. The scale value of annoyance could be well described by the linear combination of these factors.

4.4.1 Introduction

In this section I describe an analysis of traffic noise by a newly developed measurement system (Sakurai et al, 2001) and a laboratory experiment designed to explore the relationship between the physical properties of noise sound and its annoyance (Atagi et al., 2001).

Numerous studies have tried to use the listening level of a noise sound to predict its perceived annoyance. Annoyance depends directory on listening level when sounds are roughly equivalent in other attributes, such as timbre and duration. In the case of aircraft noise, an increase in SPL (sound pressure level) of 10 dB results in a doubling of the subjective annoyance (ISO R/507, 1970) For sound sources having widely different acoustical properties or durations, however, this relationship may no longer hold. We must consider other factors influencing the annoyance.

Cermak and Cornillon (1976) asked their subjects to compare a wide range of road-traffic sounds to find significant factors which affect annoyance. Despite using a variety of sounds having different temporal variance and spectrum shapes, they did not find significant factors contributed to the annoyance other than SPL. Versfeld and Vos (1997) measured the annoyance caused by sounds of military tracked vehicles and civil passenger cars. They found that the annoyance depended on the vehicle type and suggested that the difference between the high-frequency part and the low-frequency part of the spectrum might play a role in the annoyance. For evaluating the unpleasantness or annoyance of noise, Zwicker and Fastl (1999) proposed parameters such as sharpness, roughness, tonality, and fluctuation strength, in addition to loudness.

Recently, Ando (2001) said that perceived annoyance is influenced by primary

and spatial sensations of sound sources and sound fields. As primary sensations, loudness, pitch, and timbre should be considered, and as spatial sensations in a sound field, localization, subjective diffuseness, and apparent source width (ASW) should be considered. Such fundamental subjective attributes for sound fields have been described in relation to a model of an auditory-brain system (Ando, 1998, 2001) This model includes autocorrelation function (ACF) and inter-aural crosscorrelation function (IACF) mechanisms. It has been suggested that primary and spatial sensations of noise fields could be described by the temporal and spatial factors extracted from the ACF and the IACF respectively.

Based on Ando's theory, we have analyzed aircraft noise in terms of temporal and spatial factors extracted from the ACF and IACF. For example, perceived pitch and its strength of complex tones and complex noises are expressed by the first peak in the ACF of the signal. Loudness is related to the ACF factor, τ_e, not only to the listening level. These ACF factors could be possible measures of noise annoyance. In addition, spatial properties are important for noise evaluation. Noise sources are usually not fixed. We hear a different sound quality when the noise source is approaching or departing. Information on the location or direction of the sound source is expressed by the location of the maximum peak of the IACF. Subjective diffuseness and apparent source width can be evaluated by the height and width of the maximum peak of the IACF. These temporal and spatial factors may contribute to annoyance in a complex manner. To simplify the problem, only the ACF factors were investigated in the present laboratory experiment. A paired comparison test for ten subjects revealed that the ACF factor, τ_e, ϕ_1, and the variance of SPL are most highly correlated with the subjective annoyance.

4.4.2 Physical properties of traffic noise

Measurement procedure

Sound recordings were made of civil road traffic, such as a passenger car, a bus, a truck, and a motorbike. The measurement point was 5 m from the center of a road, along a line perpendicular to road. Sounds were received by two 1/2-inch condenser microphones set at the ear positions of a sphere representing a human head. This dummy head was made of 20-mm-thick Styrofoam having a 200-mm diameter. The ear positions were set at 1.5 m above the ground. Received sounds were recorded on a DAT recorder at a sampling rate of 48 kHz, and simultaneously stored on a hard disk of an analyzing computer at a sampling rate of 44.1 kHz for the following analysis.

Analysis of acoustical factors

To evaluate a temporally varying noise sound, we used the running short-time ACF and IACF. Running short-time temporal and spatial factors were used to describe the primary and spatial sensations of a sound field.

Results and discussion

Based on the running short time SPL [dBA] measured by the system, eight standard measures were calculated: (1) mean SPL, (2) variance σ^2 of the SPL, (3) maximum SPL, (4) minimum SPL, (5) – (7) the SPL values exceeded 10 % of the time (L_{10}), 50 % of the time (L_{50}), and 90 % of the time (L_{90}), and (8) equivalent sound level L_{eq}. Most of these standard measures were highly inter-correlated (see Table 4.1). Clearly, all of these factors contain information about the overall sound level and its variability. Therefore, only the median (L_{50}) and variance (σ^2) of the SPL were considered in the subsequent analysis.

Table 4.1 Correlations among eight standard noise measures for nine traffic noises.

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

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

The measured SPL and three ACF factors are represented in Figure 4.12 as a time function. Thick lines and thin lines show the two extremes of measured sounds: one has a clear pitch sensation, and the other has a weak pitch. We call these two sounds tonal noise and un-tonal noise as in a previous study (Fujii et al., 2001). The SPL throughout the measurement varied in the same manner, but the ACF factors were extremely different for each case. The value of τ_1 varied between 1 ms and 10 ms, meaning that perceived pitch varied between 1000 Hz and 100 Hz for both noises. The strength of perceived pitch increases in proportion to the value of ϕ_1. For a tonal noise,

the ϕ_1 value reaches maximum around 0.6 and 0.7. At this time, a strong tonal sound is heard having a pitch of τ_1. When the τ_1 value varies with a high ϕ_1 value, we can perceive the variation of the pitch. However, the ϕ_1 value for an un-tonal noise remained constant around 0.2, despite the variation of τ_1. The perceived pitch for an un-tonal noise is therefore very weak, and it is hard to discriminate pitch fluctuation.

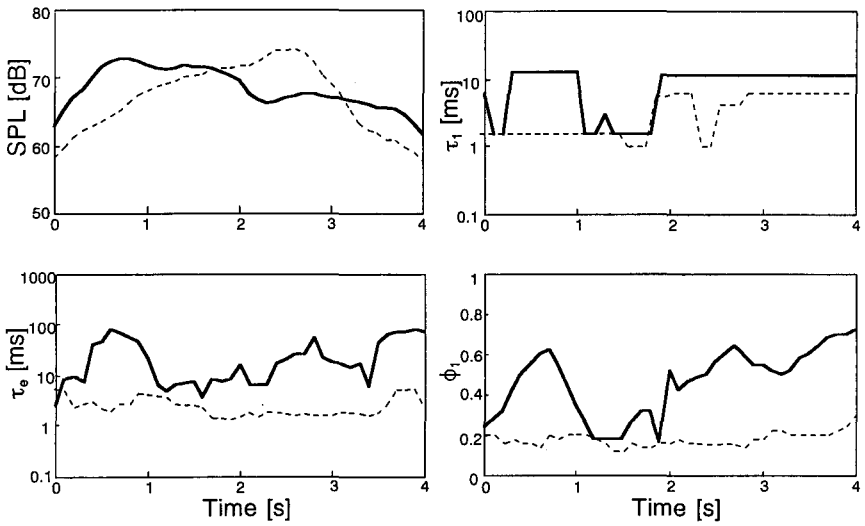

Figure 4.12 Two extreme examples of measured SPL and three ACF factors as time functions. Thick lines show the factors of tonal noise (motor bike), and thin line shows the factors of un-tonal noise (passenger car).

The calculated power spectrum and ACF for tonal and un-tonal noises is shown in Figure 4.13. For the tonal noise, there are several peaks in the spectrum. Generally, the spectrum consists of harmonic components (discrete part) and noise component (continuous part), but it is difficult to identify which peak is a fundamental frequency in the spectrum for a complex sound. When the same sound is analyzed by the ACF, its harmonic structure is easily extracted. Strong periodical peaks in the ACF show that a periodicity corresponding to the pitch is present in the sound. Minor peaks within a period of the ACF give information about the higher-frequency components or timbre of the sound (Meddis and Hewitt, 1991; Cariani and Delgutte, 1996). This information can be used by the measurement system to identify the sound source. For the un-tonal noise, there is no particular peak in the spectrum. This means that the sound has no particular periodicity perceived as pitch. In this case, the ACF decreases to zero without strong periodical peaks. Because the envelope of the ACF is related to the value

of ϕ_1, the value of τ_e is a good measure of the periodicity of the sound signal.

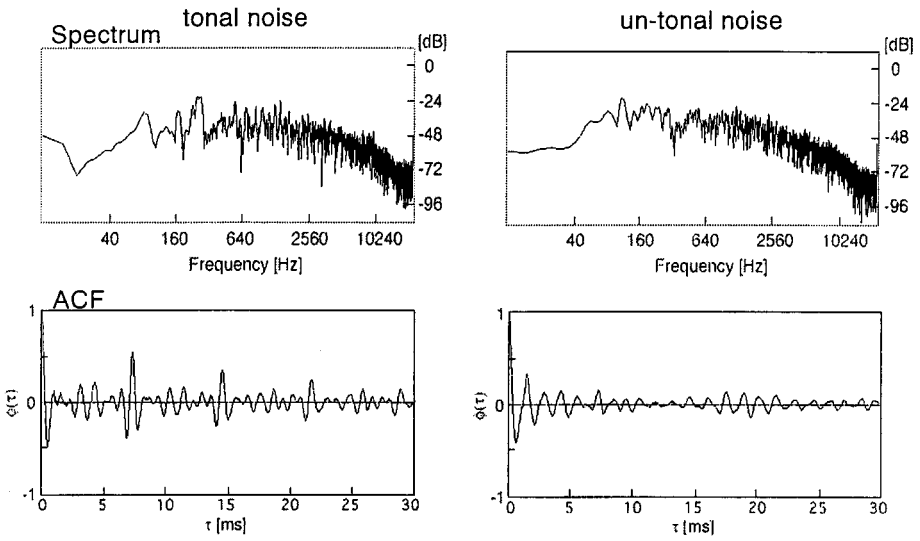

Figure 4.13 Calculated power spectrum and the ACF of tonal noise and un-tonal noise.

The perceived direction of a sound source is represented by the maximum peak of the IACF, because it corresponds to an interaural time difference. As a sound source moves from left to right, the value of τ_{IACC} varies from a minus to plus value, as illustrated in Figure 4.14. The measured values of τ_{IACC} show that the vehicles passed through the receiver from left (right) to right (left).

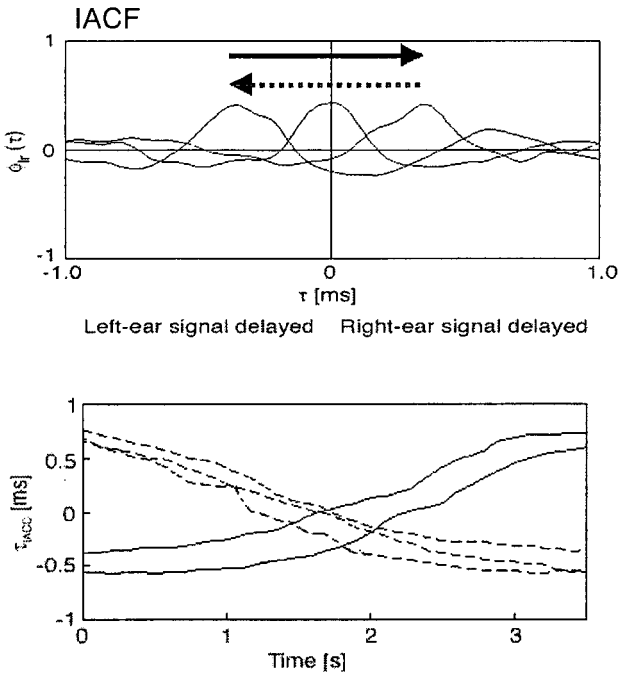

Figure 4.14 Examples of the IACF and the values of τ_{IACC}.

4.4.3 Psychological experiment

Method

Nine recordings of noise sounds were used in the experiment. Each stimulus was edited on computer software to have a 4-sec duration, and contains one vehicle's passage. The maximum level was adjusted to be equal (73 ± 2 dB) and to occur near the middle of the sound. To make the envelope of sounds equal, a 0.5-sec rise and fall time was added to all stimuli.

The cumulative frequency of measured SPL and three ACF factors are shown in Figure 4.15. To characterize the acoustical properties of a stimulus, we used the median and variance of each factor. Our assumption of perceived annoyance is as follows. (1) It has been reported that the median of SPL (L_{50}) is a good measure of annoyance. Therefore, a higher-level sound is considered to be more annoying than lower level sound. (2) Perceived pitch may also be related to annoyance. The calculated τ_1 value of stimuli ranged between 1 ms and 10 ms, corresponding to the perceived pitch of 1000 Hz and 100 Hz. Generally, a sound having a low pitch is not as annoying as a sound having a high pitch. In our experiment, stimuli with a small value of τ_1 might be more annoying. (3) It has been found that the perceived loudness of band pass noise increases in proportion to the value of τ_e (Merthayasa and Ando, 1996; Sato et al., 2001).

Annoyance may also be related to loudness, and consequently to the value of τ_e. (4) Values of ϕ_1 represent perceived pitch strength. A sound having a strong pitch might be more annoying than one having a weak pitch. (5) In general, a noise whose sound level fluctuates is more annoying than the same average noise having a constant sound level. Similarly, a noise whose pitch and timbre fluctuates is more annoying than one having a constant quality (Molino, 1979). To estimate the effects of such fluctuations of sound level and sound quality, we added the variance of SPL and ACF factors to the variables for the annoyance calculation.

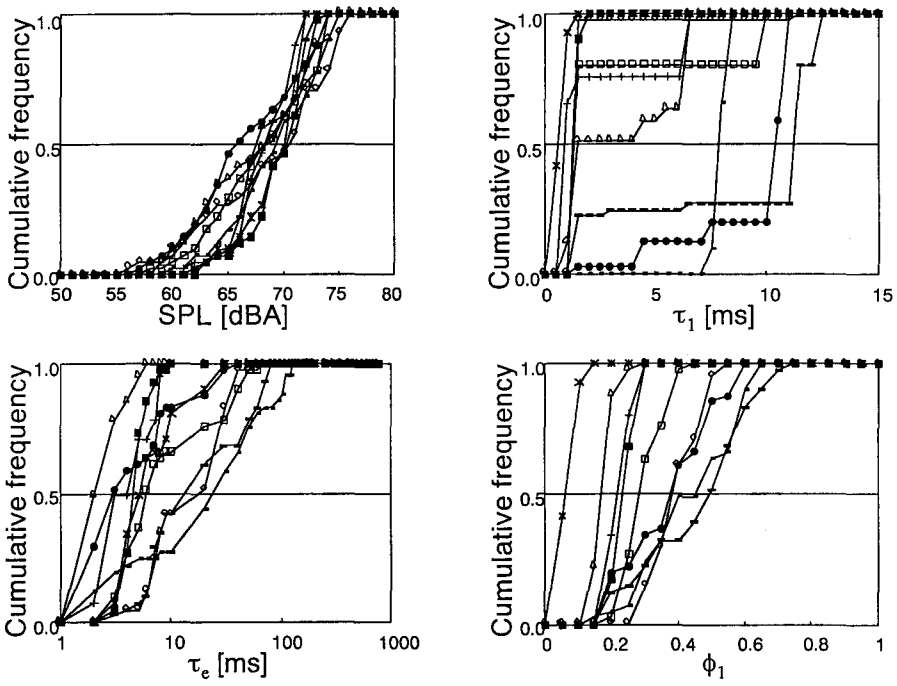

Figure 4.15 Cumulative frequencies of the SPL and the ACF factors measured for the stimuli used in the experiment.

The traffic sounds were reproduced in an anechoic room through a laptop computer, a D/A converter, a power amplifier, and a loudspeaker. As described before, only the effects of the temporal factors were examined in the experiment. A 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. Except for two of the authors (FK and AJ), the rest of the subjects were unaware of the purpose

of the study. Subjective annoyance was measured by a paired comparison method. All possible pairs from the nine sounds (36 pairs) were presented to the subjects in a random order in one session. After the presentation of paired stimuli, the subjects were asked to judge which of the two sounds were more annoying. All subjects had four series of sessions, giving a total of 144 comparisons.

Results and discussion

Collected data were processed by applying "the law of comparative judgment" (case V; Thurstone, 1927). This law is used to produce one-dimensional scale values (SV) for each stimulus from the total matrix of superiorities collected from the paired comparisons. The results were reconfirmed by the goodness of fit (Mosteller, 1951) and the agreement of all subjects' judgments was tested by the chi-square test ($p < 0.05$). With these analyses it was ascertained that the subjects' judgments were reliable and that there were certain underlying criteria upon which they agreed.

The scale values (SV) of annoyance for all the subjects were averaged, and the correlation coefficients, r, were calculated between the annoyance and the median and variance of ACF factors. The correlation matrix between the ACF factors and SV is shown in Table 4.2.

Table 4.2 Correlations between the median and variance of the ACF factors and annoyance.

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

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

Contrary to the predictions, perceived annoyance was not correlated to the SPL. It is considered that the range of the SPL among the stimuli was too small (5 dBA) to affect annoyance. Instead, the variance of the SPL had much effect on annoyance. Although

Cermak and Cornillon (1976) did not find a significant contribution of measures other than L_{eq}, our results suggest that other acoustical factors are more dominant than the L_{50} or L_{eq} when the difference of overall SPL is small.

The values of τ_e and ϕ_1 were significantly correlated to annoyance ($r = 0.56$ and 0.57, $p < 0.05$). This result shows that a sound having a strong tonal component was perceived to be more annoying than the un-tonal noise. The subjects' comment also indicated that they judged a sound having a clear pitch to be more annoying. In the evaluation of the perceived noise level for a tonal sound as used in the experiment, a number of tone corrections are used. Generally, a value is added to the "Perceived Noise Level" (PNL) to give the "Tone Corrected Perceived Noise Level" (PNLT). However, the calculation for this correction is lengthy, and their accuracy is not well established (May, 1978). Instead, by using the value of τ_e and ϕ_1, the effect of the tonal component on perceived annoyance is clearly explained.

Considering the results above, it is considered that the ACF factors and the variance of the SPL independently affect perceived annoyance. To calculate perceived annoyance more precisely, we examined multi-regression analysis by using a linear combination of eight variables, which consist of the median and variance of the ACF factors and the SPL. To obtain an optimal model, all possible combinations were examined. The correlation coefficients and significance levels were used to determine the goodness of fit. The best combination of variables was found as the variance of SPL, the median of τ_e, and the variance of τ_1. Partial regression coefficients of each variable were 0.64, 0.50, and 0.36.

$$SV_{annoyance} \approx 0.64 Var\_SPL + 0.5\tau_e + 0.36 Var\_\tau_1 \qquad (7)$$

Using these tentative values for equation 7, the total correlation coefficient 0.91 was obtained with the significance level $p < 0.05$, as shown in Figure 4.16. 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 sound and pitch fluctuation, which are the important factor for annoyance, could be calculated by the proposed ACF analysis. As for the other factors, which we did not concern in this study, roughness and sharpness of sound could be considered in the future work.

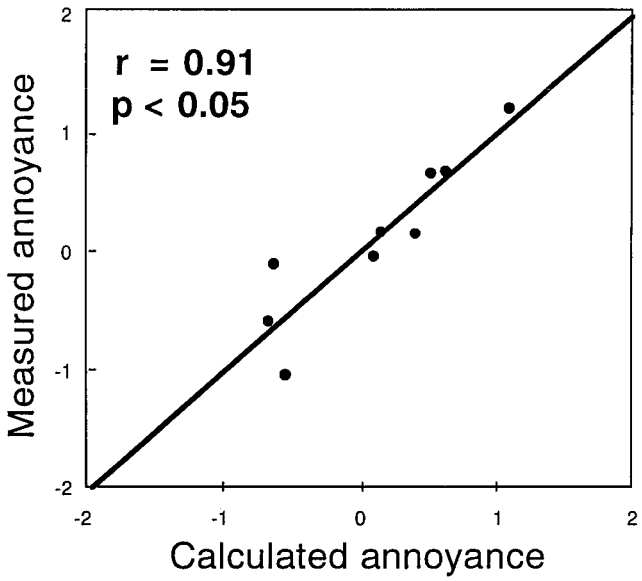

Figure 4.16 Relationship between the measured annoyance and the calculated annoyance by using the linear combination of the ACF factors.

4.4.4 Conclusion

The purpose of this study was to describe the acoustical properties of traffic noise and to explore the relationship between the described properties and perceived annoyance. From the results we concluded that: (1) The ACF analysis is effective in characterizing sounds, such as the perceived pitch and timbre. The IACF analysis is effective in describing the spatial information of the noise source. (2) Perceived annoyance is greatly affected by the variation of the SPL and other primary sensations, when the difference of the overall SPL is small. (3) The combination of the ACF factors can calculate perceived annoyance sufficiently.

4.5 Summary

In this chapter, I presented a series of studies on physical properties and psychological evaluations of sound. I used the ACF in analyzing a sound signal and the interaural cross-correlation function (IACF) to characterize the spatial properties of sound field. It was found that the acoustical properties are well represented by the factors extracted from the ACF and the IACF, and that the subjective evaluations for sound signal are explained by the combinations of the ACF factors.

5. Application

5.1 Introduction

In this section, I will present two engineering applications based on the correlation model in human visual and auditory system. One is a method of analyzing textural features for image retrieval and pattern recognition. Another is about feature extraction for speech recognition technology. These techniques are using our ability for detecting periodical structure in visual and auditory signals. Also, they intend to evaluate our subjective attributes in visual and auditory sensations.

5.2 Analysis of textural features for image retrieval

Summary

I present a new method for measuring textural properties especially corresponding to human visual perception. Proposed method is a basis of the image retrieval system, pattern recognition, and machine vision. Physical parameters on perceived contrast, coarseness, and regularity are extracted from the autocorrelation function on the gray scale image.

5.2.1 Background and previous work

Recently, a huge archive of image, film, and photographs is digitized in various fields. Consequently a demand of automated image retrieval system is increasing. Tools assisting image search within a large database have broad applications, such as medical image query, video editing, and materials for architectural design, product design, and so on. A retrieval system serves the purpose of saving human users' time and effort of browsing the entire database. Therefore, it is expected that the retrieved images resemble the visual properties of the prototype pattern provided by the human user. To construct such a system, it is important that the features used for pattern comparisons are faithful to those used by humans (Liu and Picard, 1996). Considering an application for pattern recognition, we have to choose a set of features for measuring human perceptual similarity.

To describe visual properties of image, texture might have an important role,

because a textured region exhibit certain degree of homogeneity and can be regarded as containing a same or near same information over neighborhood. However, there is a difficulty in evaluation of subjective judgment of similarity and preference of texture, which are inevitable for image comparison and classification. This difficulty comes from multidimensional nature of texture perception. We need to know how many features are related to texture perception. As human texture perception, Rao and Lohse (1996) have indicated that the most salient perceptual dimensions in texture dissimilarity judgment can be described as "repetitiveness", "contrast", and "coarseness". Hence, it is effective for a retrieval system to use in modeling textural features, which relate to these perceptual dimensions. I propose here a set of image features based on the correlation model to capture the properties of human texture perception.

5.2.2 Method

As I presented in Chapter 2, the autocorrelation function (ACF) analysis provides useful measures for representing three salient perceptual properties of texture, namely, contrast, coarseness, and regularity. Especially, regularity is the most important property for texture discrimination. Our method is based on the nature of the ACF such that the periodical component in the image is extracted. The ACF gives the correlation between pixel pairs for every direction and distance in the image. If the image contains periodicity, its ACF also has periodical structure. Period of the ACF (δ_1) corresponds to the period of the image, and amplitude of the maximum peak in the ACF (ϕ_1) is related to the degree of periodicity. Such properties of the ACF are well corresponded to the perceptual properties of textures, coarseness and regularity. Also, perceived contrast of the image is related to the root mean square (RMS) of the gray scale distribution. The value of RMS is effectively calculated in the ACF analysis because when the correlation lag is zero ($\Phi(0)$), the autocorrelation is the RMS itself (see 2.3.2 Autocorrelation analysis of natural textures).

5.2.3 Results

To examine the efficiency of the ACF analysis, we analyzed 16 textures from the "Brodatz texture database" (Brodatz, 1966), that were used by Tamura et al. (1978). This texture set includes a variety of natural and man-made objects (Figure 2.9). The

analyzed data consists of 256 × 256 images with 8-bit (256) gray levels. The ACF was calculated and three textural properties were extracted from each image. Results are shown in Figure 5.1, 5.2, and 5.3 for contrast, coarseness, and regularity. In each Figure, textures are replaced from left to right, top to bottom, in ranking order by computation for each property.

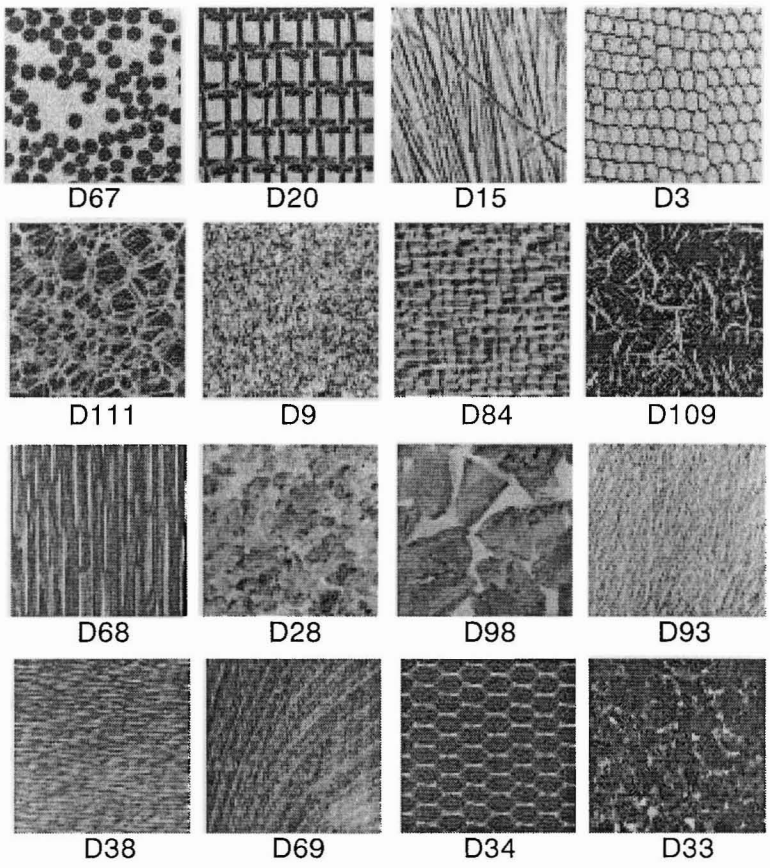

Figure 5.1 Test samples displayed in ranking order by computation for contrast. From left to right, top to bottom, the images are from the highest contrast to the lowest contrast.

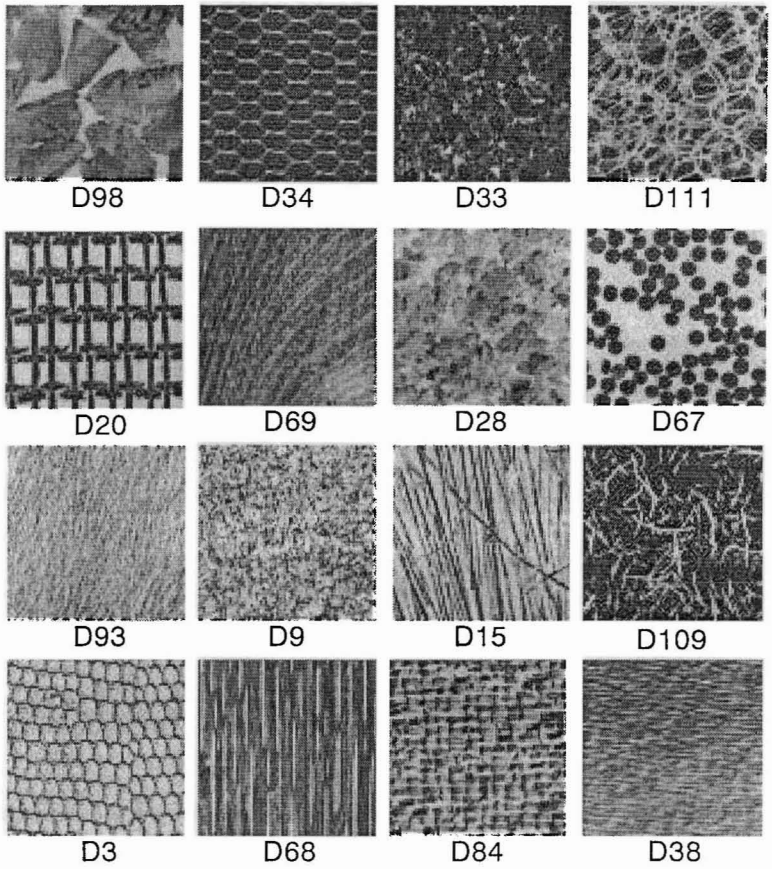

Figure 5.2 Test samples displayed in ranking order by computation for coarseness. From left to right, top to bottom, the images are from coarse to fine.

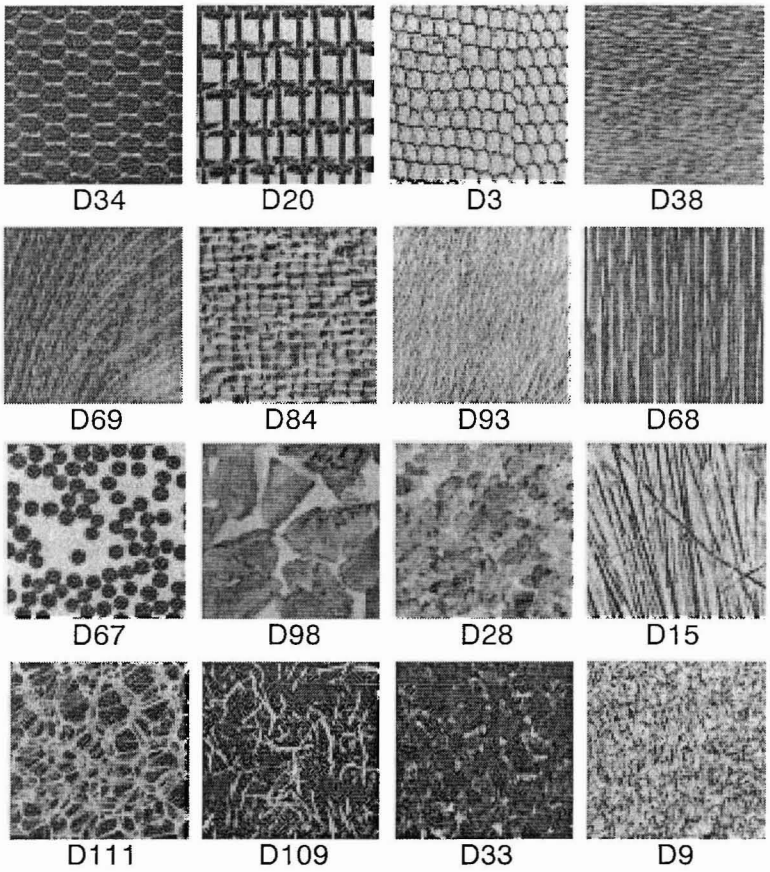

Figure 5.3 Test samples displayed in ranking order by computation for regularity. From left to right, top to bottom, the images are from regular to random.

5.3 Feature extraction for speech recognition

Summary

A method of feature extraction for speech recognition system in the real environment is proposed. By using a minimum set of parameters corresponding to the human auditory attributes, we can characterize speech signals. Our method is based on the ACF and IACF model of human auditory system. The ACF analysis is effective to extract subjective attributes of sound, such as pitch and timbre. Also, the IACF analysis could be used to characterize spatial attribute of sound field. Identification is based on the extracted parameters' minimum distance between input signal and syllable set in the database.

5.3.1 Background and previous work

In the technology of speech recognition, there is a widely accepted method such that

analyzing a given speech signal within a series of short duration, constructing the time sequence vectors of extracted features, then comparing them with the vector of stored template signals. The main problem in the speech analysis is to find out the sufficient feature set by which we could discriminate a given signal with others. Among a various method proposed, frequency spectrum estimation, Cepstrum, and linear predictive coding (LPC) is generally used. Because the characteristics of speech signals are represented in the spectral structure, estimation of the spectrum shape could be effective for speech identification.

However, there remain some serious problems in the present method. Firstly, we need complex parameter to approximate a spectrum shape, because the speech signals contain a large amount of frequency information. It may cause a prediction error. These parameters also include ones, which does not reflect our auditory sensation. To adapt the individual voice differences, the recognition system must identify the difference in spectral shapes. Secondly, these methods do not work well with noise. Spectral shapes of the speech signal are much affected by the presence of background noise and reverberation in the real environment. Thirdly, when considering speech recognition in the real environment, it needs to separate the plural sources arriving from different directions and localize the target source. Therefore we need to consider the spatial attributes in the sound field. These problems, namely, "speaker adaptation", "noise robustness", and "source separation", are the very topic that speech recognition technology is currently facing.

Here, I propose a method for implementation of speech recognition system in the real environment by using a minimum set of parameters corresponding to the human auditory characteristics. Our method is based on the ACF and IACF model of human auditory system. As described in Chapter 4, the ACF analysis is effective to extract subjective attributes of sound, pitch and timbre. Also, the IACF analysis could be used to characterize spatial attribute of sound field.

There is effective duration time τ_e ACF, as the most important factor (feature quantity) in the analysis of ACF. Effective duration τ_e is defined as a 10% delay time. Effects of the repetition and reverberation components included in the signal are described by τ_e. In addition, the fine structure of ACF including peak and dip contains much information on the periodicity of the signal. The most effective factor in the analysis of speech signal is information on a pitch. Delay time and amplitude of the

principal peak of ACF are the factors, which corresponds to a pitch of the voice and its strength. The principal peak is a largest peak of ACF in most cases, and successive peaks appear in its period. The local peaks within a period between principal peaks show the time structure of the high frequency range. Information on the tone color is included. Especially in case of the voice signal, features of the resonant frequency of vocal tract called a formant are shown. The above ACF factors seem to contain all features necessary for the voice recognition.

The value of τ_{IACC}, in the range between -1ms and +1ms in the IACF is an important factor on the horizontal direction of the sound source. Clear sense of direction is perceived, when IACC has large value. When IACC has the small value, the subjective diffuseness increases, and sense of direction becomes unclear. It is possible to obtain the apparent width of sound source by IACC and W_{IACC}.

5.3.2 Method

The specific equipment for this method is consisted of binaural microphones, low-pass filter, A/D converter, and computer for calculating ACF and IACF as it is shown in Figure 5.4. Figure 5.4 also shows a flowchart of the method for identifying the syllable. Collected speech signal is converted into digital signal through the A/D converter and low pass filter and stored in the memory. Stored signal is read out, and the factors are deduced by the calculation of ACF and IACF. Finally, it is compared with the database storing the template of the syllables.

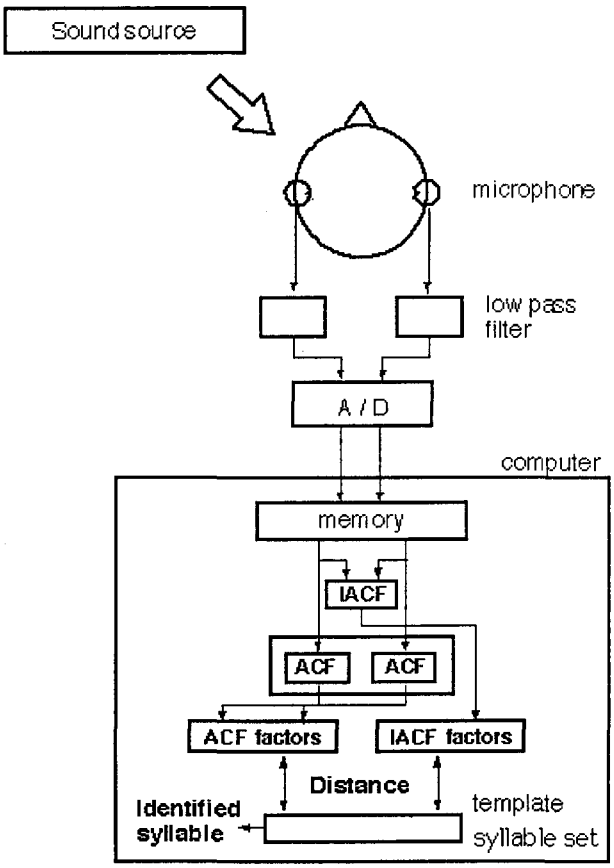

Figure 5.4 Specific equipment for the proposed syllable identification method. It is consists of binaural microphones, low-pass filter, A/D converter, and computer for calculating ACF and IACF.

As following, the concrete calculation method of ACF and IACF is described. For the target speech signal, running ACF and running IACF are calculated for short time segment (it is called the frame) F_k (t), as shown in Figure 5.5. Usually the integral interval 2T (length of the frame) is set as between 20 and 30 ms, overlapped with neighboring frames. Short time running ACF will be calculated as following equation.

$$\Phi_p(\tau,t,T) = \frac{1}{2T}\int_{t-T}^{t+T} p'(t)p'(t+\tau)dt \tag{5.1}$$

Normalized ACF will have the 1 maximum value at $\tau = 0$. Binaural sound pressure level is given by the average of two monaural levels. Effective duration time τ_e is defined by delay time t as the 10-percentile attenuation of the ACF. Since that initial ACF linearly attenuates is generally observed, it is possible to easily require τ_e by the linear regression.

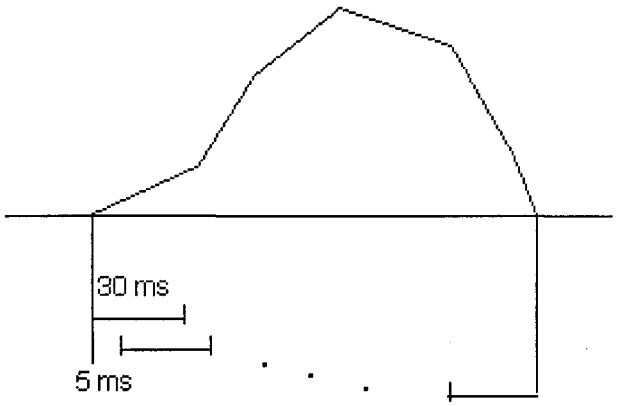

Figure 5.5 Short time segment of speech signal for calculation of running ACF.

Concretely, τ_e is decided using least mean square method (LMS) for the peak of ACF appeared in some fixed time Δt. In Figure 5.6, the example of normalized ACF is shown. The delay time and amplitude of the normalized ACF is defined as τ_k, ϕ_k, k=1, 2, ..., N. The interval in search of the peaks is up to the maximum peak of ACF appears from delay time $\tau = 0$. It is correspondent to the one period of the ACF. The largest peak of ACF is correspondent to a pitch of the sound source, and the local peaks within the largest peak are correspondent to the formant.

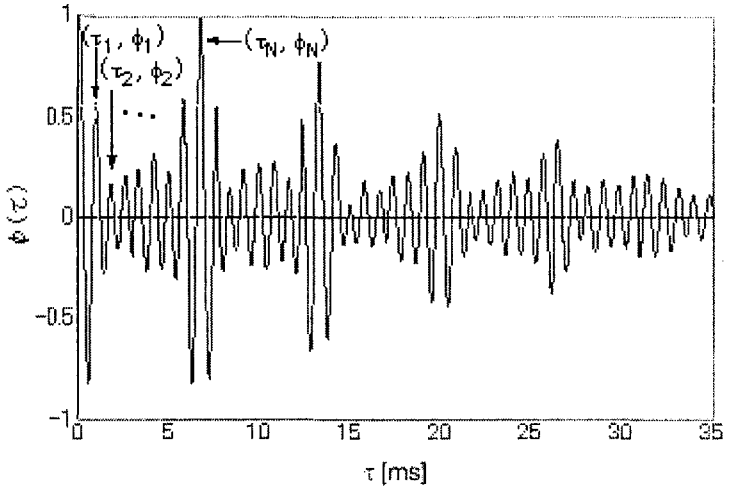

Figure 5.6 The example of normalized ACF calculated for speech signal.

Finally, the method for identifying the syllable is described. Identification is based on the extracted parameters' distance between input signal and syllables of the template. The template is a set of ACF factors on the whole syllable set calculated

beforehand. Distance D(x) of the template (shown by symbol b) and input data (shown by symbol a) is calculated like the following equation (x: $\Phi(0)$, τ_e, τ_k, ϕ_k, k=1,2, ..., N).

$$D(\Phi(0)) = \{\sum_{k=1}^{N} \left|\log(\Phi(0)_k)^a - \log(\Phi(0)_k)^b\right|\} / N \quad (5.2)$$

Equation 11 calculates the distance on $\Phi(0)$. N shows the number of the analysis frame. The distance is also obtained on other independent factors in the similar equation. Total D of the distance is shown by the following equation.

$$D = \sum_{x=1}^{M} W^x D(x) \quad (5.3)$$

M in equation 12 is a number of the factor, and W is a weighting factor. It is judged that the template, in which calculated distance D is the smallest, is a syllable of the input signal. In the actual sound field, the recognition at the high accuracy becomes possible by adding the IACF factors.

5.3.3 Results

As an example of the implementation of proposed method, Figure 5.7 shows the prediction of speech intelligibility in the real sound field. Experiment was carried out as follows. Target syllable and interference sound (white noise or another syllable) was presented from loudspeaker placed in front or lateral direction of the subject. Subjects were asked to answer what they hear as target syllable. Intelligibility is represented by a percentage of the correct answer on the abscissa. To predict intelligibility, the distance of the ACF and IACF factors were calculated between target-only signal and mixed signal. Percentage of which the least distance was found for target syllable is represented on the ordinate as predicted intelligibility. As shown in Figure 5.7, the result gave high correlation ($r = 0.86$, $p < 0.05$) between measured and calculated intelligibility. We can say that that the recognition by the proposed method reflects human sensibility.

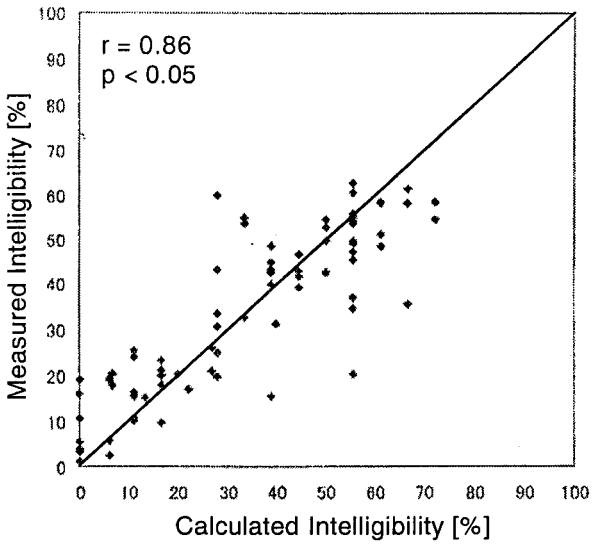

Figure 5.7 Measured and calculated speech intelligibility for single syllable presented simultaneously with interference sound.

5.4 Summary

In this chapter, I proposed two engineering applications based on the correlation model in human visual and auditory system. One is an analyzing method of textural features for image retrieval and pattern recognition. Another is sound feature extraction for speech recognition technology. These techniques use our ability for detecting periodical structure in visual and auditory signals.

6. Conclusion

6.1 Summary of results

As I outlined in Chapter 1, this dissertation covers the research area of perception and psychological evaluation in spatial vision, temporal vision, and audition. The fundamental result that I have presented is a demonstration that it is possible to apply the correlation mechanism to the process in vision and auditory system. The results gained in each study are summarized as follows.

In Chapter 2, **Spatial vision**, I presented a series of studies on physical properties and psychological evaluations of two-dimensional spatial pattern. I showed that the autocorrelation function (ACF) analysis provides useful measures for representing three salient perceptual properties of texture, namely, contrast, coarseness, and regularity. Another experiment showed that the degree of regularity is a salient cue for texture preference judgment. Described ACF model offered the advantage of extracting perceptual properties and evaluating subjective reaction in texture perception.

In Chapter 3, **Temporal vision**, I discussed the underlying mechanism for the temporal perception in vision. To address the problem, I focused on the fundamental properties of the temporal vision mechanism. Psychophysical experiment was performed on subjective flicker rates for complex waveforms. Results showed that human observers perceived a rate at the fundamental frequency, although the energy at this frequency was not included in the signals. It implies the existence of correlation mechanism in temporal vision.

In Chapter 4, **Audition**, I presented a series of studies on physical properties and psychological evaluations of sound. I used the ACF in analyzing a sound signal and the interaural cross-correlation function (IACF) to characterize the spatial properties of sound field. It was found that the acoustical properties are well represented by the factors extracted from the ACF and the IACF, and that the subjective evaluations for sound signal are explained by the combinations of the ACF factors.

In Chapter 5, **Application**, I proposed two engineering applications based on the correlation model in human visual and auditory system. One is an analyzing method of textural features for image retrieval and pattern recognition. Another is sound feature

extraction for speech recognition technology. These techniques use our ability for detecting periodical structure in visual and auditory signals.

6.2 Future directions

One of the most important future research is to get into more sophisticated model for visual and auditory perception. In this dissertation, I presented workable correlation models for evaluating perception and subjective response in spatial vision, in temporal vision, and in audition. But we are still far from understanding of the mechanisms in vision and audition. Especially for the high level recognition of the real environment, there remains amount of room for future research. I believe that the approach I take in this dissertation contributes to the understanding of the mechanisms in perception and recognition of complex phenomena.

Understanding of the neural mechanism is also important. To discuss the analogous subjective attributes in vision and audition, we need to assume the general neural strategy in the brain. As one of such mechanism, "neural timing net" is proposed (Cariani, 2001). Timing nets constitute a new and general neural network strategy for performing temporal computations on neural spike trains: extraction of common periodicities, detection of recurring temporal patterns, and formation and separation of invariant spike patterns. Although the main concern in the paper is with the importance of the timing information in auditory perception, the analogous strategies in visual perception are also suggested. The possibilities of extension of this model to visual perception could be examined in the future.

In this dissertation, I discussed the correlation mechanism of spatial and temporal vision separately. As a next step, they could be expanded to the general model of visual system including spatial and temporal properties. Adelson and Bergen (1991) proposed the generalized concept of "plenoptic functions", which can describe the structure of the information in the light rays coming on an observer. Plenoptic functions characterize local change along one or more dimensions of a single function. Their basic notion about visual perception is based on the edge detection mechanism in early vision. We can also consider the periodicity detection mechanism in somewhat late visual pathway for describing visual perception. As for a model of spatio-temporal vision, Reichardt's motion detectors and motion energy models have been proposed (Reichardt, 1961; van Santen and Sperling, 1984, 1985; Adelson and Bergen, 1985). These models

referred to spatial and temporal correlation detection mechanisms. Therefore they could be applied to spatial-only pattern and temporal-only pattern of visual information.

Another interesting and important research direction related to this dissertation is the engineering application using human abilities and characteristics in perception. Firstly, we can consider the implementation of human mechanism in the environment recognition system, such as image recognition and speech recognition. Secondly, the man-machine interface could be accomplished which reflects human perceptual properties. An example in this direction is the image and sound retrieval system using salient cues in visual and auditory perception.

References

Adelson, E. H. and Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. Journal of the Optical Society of America, A, **2**, 284-299.

Adelson, E. H. and Bergen, J. R. (1991). The plenoptic function and the elements of early vision. In Landy, M., and Movshon, J. A (eds.). Computational models of visual processing., pp. 3-20. Cambridge, MA: MIT Press.

Amandasun, M. and King, R. (1989). Textural features corresponding to textural properties. IEEE Transactions on Systems, Man and Cybernetics, **19**, 1264-1274.

Ando, Y. (1977). Effects of noise on duration experience. Journal of Sound and Vibration, **55**, 600-603.

Ando, Y. (1985). Concert Hall Acoustics. Springer-Verlag, Heidelberg.

Ando, Y. (1998). Architectural acoustics-Blending sound sources, sound fields, and listeners. New York: AIP/Springer-Verlag.

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

Ando, Y. and Hattori, H. (1970). Effect of noise during fetal life upon postnatal adaptability (statistical study of the reaction of babies to air-craft noise). Journal of the Acoustical Society of America, **47**, 1128-1130.

Ando, Y. and Hattori, H. (1973). Statistical studies on the effect of intense noise during human fetal life. Journal of Sound and Vibration, **27**, 101-110.

Ando, Y. and Hattori, H. (1974). Reaction of infants to aircraft noise and effect on sleep of babies (in Japanese). Zibi Rinsyo, **67**, 129-136.

Ando, Y. and Hattori, H. (1977a). Effects of noise on human placental lactogen (HPL) levels in maternal plasma. British Journal of Obstet and Gynaeco, **84**, 115-118.

Ando, Y. and Hattori, H. (1977). Effects of noise on sleep of babies. Journal of the Acoustical Society of America, **62**, 199-204.

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., pp. 63-99. Southampton: WIT Press.

Atagi, J., Fujii, K., and Ando, Y. (2001). Temporal and spatial factors of traffic noise and its annoyance. Proceedings of 17th International Congress on Acoustics.

Badcock, D. R. and Derrington, A. M. (1989). Detecting the displacement of spatial beats: no role for distortion products. Vision Research, **29**, 731-739.

Barron, M. and Marshall, A. H. (1981). Spatial impression due to early lateral reflections in concert halls, the derivation of a physical measure. Journal of Sound and

Vibration, **66**, 1-14.

Bartley, S. H., Paczewitz, G., and Valsi, E. (1957). Brightness enhancement and the stimulus cycle. Journal of Psychology, **43**, 187-192.

Beck, J. (1966). Perceptual grouping produced y changes in orientation and shape. Science, **154**, 538-540.

Ben-Av, M. B. and Sagi, D. (1995). Perceptual grouping y similarity and proximity: experimental results can be predicted by intensity autocorrelations. Vision Research, **35**, 853-866.

Beranek, L. L. (1996). Concert and Opera Halls: How They Sound. New York: Acoustical Society of America.

Blauert, J. (1983). Spatial Hearing: The psychophysics of human sound localization. Cambridge, MA: MIT Press.

Bovik, A. C., Clark, M., and Geisler, W. S. (1990). Multichannel texture analysis using localized spatial filters. IEEE Transactions on Pattern Analysis and Machine Intelligence, **12**, 55-73.

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.

Brodatz, P. (1966). Textures. New York: Dover.

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. (2001). Neural timing nets. Neural Networks, **14**, 737-753.

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, rate pitch, and the dominance region for pitch. Journal of Neurophysiology, **76**, 1717-1734.

Cermak, G. W. and Cornillon, P. C. (1976). Multidimensional analysis of judgments about traffic noise. Journal of the Acoustical Society of America, **59**, 1412-1420.

Cho, R. Y., Yang, V., and Hallett, P. E. (2000). Reliability and dimensionality of judgments of visually textured materials. Perception &Psychophysics, **62**, 735-752.

Cross, G. R. and Jain, A. K. (1983). Markov random field texture models. IEEE Transactions on Pattern Analysis and Machine Intelligence, **5**, 25-39.

Damaske, P. and Ando, Y. (1972). Interaural crosscorrelation for multichannel loudspeaker reproduction. Acustica, **27**, 232-238.

de Lange, H. (1952). Experiments on flicker and some calculations on an electrical analogue of the foveal systems. Physica, **8**, 935-950.

de Lange, H. (1958). Research into the dynamic nature of the human fovea-cortex systems with intermittent and modulated light. Journal of the Optical Society of America, **48**, 777-789.

DeValois, K. K., DeValois, R. L., and Yund, E. W. (1979). Responses of striate cortex cells to gratings and checkerboard patterns. Journal of Physiology (London), **291**, 483-505.

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.

Eisner, A., Shapiro, A. G., and Middleton, J. A. (1998). Equivalence between temporal frequency and modulation depth for flicker response suppression: analysis of three-process model of visual adaptation. Journal of the Optical Society of America, A, **15**, 1987-2002.

Fraisse, P. (1982). Rhythm and tempo. In Deutsch, D (ed.) Psychology of music. Orlando, Fl: Academic Press.

Fraisse, P. (1984). Perception and estimation of time. Annual Review of Psychology, **35**, 1-36.

Francos, J. M., Zvi Meiri, A., and Porat, B. (1993). A 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.

Gibson, J. J. (1950). The perception of the Visual World. Boston, Houghton Mifflin.

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.

Gorea, A., Wardak, C., and Lorenzi, C. (2000). Visual sensitivity to temporal modulation of temporal noise. Vision Research, **40**, 3817-3822.

Granit, R. and Harper, P. (1930). Comparative studies on the peripheral and central retina: II. synaptic reactions in the eye. American Journal of Physiology, **95**, 211-228.

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. Journal of the Acoustical Society of America, **61**, 1270-1277.

Hammett, S. T. and Smith, A. T. (1994). Temporal eats in the human visual system. Vision Research, **34**, 2833-2840.

Handel, S. and Yoder, J. (1975). The effects of intensity and interval rhythms on the perception of auditory and visual temporal patterns. Quarterly Journal of Experimental

Psychology, **27**, 111-122.

Haralick, R. M. (1979). Statistical and structural approaches to textures. Proceedings of IEEE, **67**, 786-804.

Haralick, R. M., Shanmugam, K., and Dinstein, I. (1973). Textural features for image classification. IEEE Transactions on System, Man, and Cybernetics, **SMC-3**, 610-621.

Hecht, S. and Verrijp, C .D. (1933). Intermittent stimulation by light. III the relation between intensity and critical fusion frequency for different retinal locations. Journal of General Physiology, **17**, 251-264.

Hellman, R. P. (1982). Loudness, annoyance, and noisiness produced y 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.

Hermes, D. J. (1988). Measurement of pitch by subharmonic summation. Journal of the Acoustical Society of America, **83**, 257-264.

Hess, R. F. and Snowden, R. J. (1992). Temporal properties of human visual filters: Number, shapes and spatial covariation. Vision Research, **32**, 47-59.

Hewitt, M. J. and Meddis, R. (1991). An evaluation of eight computer models of mammalian inner hair-cell function. Journal of the Acoustical Society of America, **90**, 904-917.

Hubel, D. H. and Wiesel, T. N. (1959). Receptive fields of single neurons in the cat's striate cortex. Journal of Physiology (London), **148**, 574-591.

Hudspeth, A. J. and Corey, D. P. (1977). Sensitivity, polarity, and conductance change in the response of vertebrate hair cells to controlled mechanical stimuli. Proceedings of National Academy of Science, USA, **74**, 2407-2411.

International Standardization Organization, Geneva (1970). Procedure for Describing Aircraft Noise Around an Airport, ISO R/507.

Jain, A. K. and Farrokhnia, F. (1991). Unsupervised texture segmentation using gabor filters. Pattern Recognition, **24**, 1167-1185.

Jeffress, L. A. (1948). A place theory of sound localization. Journal of Comparative Physiology and Psychology, **61**, 468-486.

Johnston, A. and Clifford, C. W. C. (1995). A unified account of three apparent motion illusions. Vision Research, **35**, 1109-1123.

Julesz, B. (1980). Spatial nonlinearities in the instantaneous perception of textures with identical power spectra. Philosophical Transactions Royal Society of London B, **290**,

83-94.

Julesz, B., Gilbert, E. N., and Victor, J. D. (1973). Inabitability of humans to discriminate between visual textures that agree in second-order statistics - revisited. Perception, **2**, 391-405.

Keesey, U. T. (1972). Flicker and pattern detection: A comparison of thresholds. Journal of the Optical Society of America, **62**, 446-448.

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

Kelly, D. H. (1961). Visual responses to time-dependent stimuli, I: Amplitude sensitivity measurements. Journal of the Optical Society of America, **59**, 422-429.

Kelly, D. H. (1969). Diffusion model of linear flicker responses. Journal of the Optical Society of America, **59**, 1665-1670.

Kelly, D. H. (1971). Theory of flicker and transient responses. I. uniform fields. Journal of the Optical Society of America, **61**, 537-546.

Kelly, D. H. and Savoiek, R. E. (1978). Theory of flicker and transient responses. III. an essential nonlinearity. Journal of Optical Society of America, **68**, 1481-1490.

Koffka, K. (1935). Principles of Gestalt psychology. New York: Harcourt Brace Janovich.

Krauskopf, J. (1980). Discrimination and detection of changes in luminance. Vision Research, **20**, 671-677.

Kremers, J., Lee, B. B., Pokorny, J., and Smith, V. C. (1993). Responses of macaque ganglion cells and human observers to compound periodic waveforms. Vision Research, **33**, 1997-2011.

Langner, G. (1997). Neural processing and representation of periodicity pitch. Acta Otolaryngology, **532**, 68-76.

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. New York: McGraw-Hill.

Lindemann, W. (1986a). Extension of a binaural cross-correlation model by contralateral inhibition. I. simulation of lateralization for stationary signals. Journal of the Acoustical Society of America, **80**, 1608-1622.

Lindemann, W. (1986b). Extension of a binaural cross-correlation model by contralateral inhibition. II. the law of the first wave front. Journal of the Acoustical Society of America, **80**, 1623-1630.

Liu, F. and Picard, R. W. (1996). Periodicity, directionality, and randomness: Wold features for image modeling and retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence, **18**, 722-733.

Macleod, I. A., Williams, D. R., and Makous, W. (1992). A visual nonlinearity fed by single cones. Vision Research, **32**, 347-363.

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. (1984). Temporal frequency discrimination above threshold. Vision Research, **24**, 1873-1880.

Mandler, M. B. and Bowker, D. O. (1980). Shifts in apparent flicker rate following flicker adaptation. Investigative Ophthalmology &Visual Science, **19**, 45.

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.

Matsuyama, T., Miura, S., and Nagao, M. (1983). Structural analysis of natural textures by Fourier transformation. Computer Vision, Graphics, and Image Processing, **24**, 347-362.

May, D. N. (1978). Basic subjective responses to noise. In May, D. N (ed.) Handbook of noise assessment., pp. 3-38. New York: Van Nostrand Reinhold Co.

Meddis, R. (1986). Simulation of mechanical to neural transduction in the auditory receptor. Journal of the Acoustical Society of America, **79**, 702-711.

Meddis, R. (1988). Simulation of auditory-neural transduction: Further studies. Journal of the Acoustical Society of America, **83**, 1056-1063.

Meddis, R. and Hewitt, M. J. (1991a). Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification. 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 Hewitt, M. J. (1992). Modeling the identification of concurrent vowels with different fundamental frequencies. Journal of the Acoustical Society of America, **91**, 233-245.

Merthayasa, I. G. N. and Ando, Y. (1996). Variation in the autocorrelation function of narrow and noises; their effect on loudness judgment. Japan and Sweden Symposium on Medical Effects of Noise.

Miller, J. R. and Carterette, E. C. (1975). Perceptual space for musical structures. Journal of the Acoustical Society of America, **58**, 711-720.

Mishima, J. (1951). Fundamental research on the constancy of mental tempo. Japanese Journal of Psychology, **22**, 12-28.

Mishima, J. (1956). On the factors of the mental tempo. Japanese Journal of Psychological Research, **4**, 27-37.

Molino, J. A. (1979). Annoyance and noise. In Harris, C. M (ed.) Handbook of noise control., pp.16 (1-9). New York: McGraw-Hill.

Moore, B. C. and Rosen, S. M. (1979). Tune recognition with reduced pitch and interval information. Quarterly Journal of Experimental Psychology, **31**, 229-40.

Mosteller, F. (1951). Remarks on paired comparisons: III a test of significance for paired comparisons when equal standard deviations and equal correlations are assumed. Psychometrika, **16**, 207-218.

Mouri, K., Akiyama, K., and Ando, Y. (2000). 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**, 87-96.

Nachmias, J. and Rogowitz, B. E. (1983). Masking by spatially modulated gratings. Vision Research, **23**, 1621-1629.

Navon, D. (1977). Forest before trees, The precedence of global features in visual perception. Cognitive psychology, **9**, 353-383.

Ohgushi, K. (1980). Physical and psychological factors governing timbre of complex tones. Journal of the acoustical Society of Japan, **36**, 253-259. (in Japanese).

Pickles, J. O. (1982). An introduction to the physiology of hearing. Academic Press.

Raney, J. P. and Cawthon, J. M. (1979). Aircraft noise. In Harris, C. M (ed.) Handbook of Noise Control, pp. 34 (1-18). New York: McGraw-Hill.

Rao, A. R. and Lohse, G. L. (1996). Towards a texture naming system: Identifying relevant dimensions of texture. Vision Research, **36**, 1649-1669.

Reichardt, W. (1961). Autocorrelation, a principle for the evaluation of sensory information. In Rosenblith, W. A (ed.) Sensory communication., pp.303-317. Cambridge, MA: MIT Press.

Research Committee of Road Traffic Noise in Acoustical Society of Japan. (1999). ASJ prediction model 1998 for road traffic noise report from research committee of road traffic noise in acoustical society of Japan (in Japanese). Journal of the Acoustical Society of Japan, **55**, 281-324.

Roehrig, W. C. (1967). The influence of the portion of the retina stimulated on the critical flicker-fusion threshold. Journal of Psychology, **48**, 57-63.

Sakai, K. and Finkel, L. H. (1995). Characterization of the spatial-frequency spectrum in the perception of shape from texture. Journal of the Optical Society of America, A, **12**, 1208-1224.

Sakai, K. and Finkel, L. H. (1997). Spatial-frequency analysis in the perception of perspective depth. Network: Computation in Neural Systems, **8**, 335-352.

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. (1998). On the apparent source width (ASW) for bandpass noises related to the IACC and the width of the interaural cross-correlation function (WIACC). Journal of the Acoustical Society of America, **105**, 1234.

Sato, S., Kitamura, T., Sakai, H., and Ando, Y. (2001). The loudness of "complex noise" in relation to the factors extracted from the auto-correlation function. Journal of Sound and Vibration, **241**, 97-103.

Schiller, P., Sandell, J., and Maunsell, J. (1986). Functions of the on and off channels of the visual system. Nature, **322**, 824-825.

Schroeder, M. R., Gottlob, D., and Siebrasse, K. F. (1974). Comparative study of curopean concert halls, correlation of subjective preference with geometric and acoustic parameters. Journal of the Acoustical Society of America, **56**, 1195-1201.

Shoda, T. and Ando, Y. (1998). Calculation of speech intelligibility using four orthogonal factors extracted from the autocorrelation function of sound source and sound field signals. Journal of the Acoustical Society of America, **103**, 2999.

Slaney, M. and Lyon, R. F. (1990). A perceptual pitch detector. Proceedings of the 1990 International Conference on Acoustics, Speech, and Signal Processing., pp. 357-360.

American National Standards. (1960). Acoustical Terminology S1. Acoustical Society of America.

Stephens, D. G. and Cazier, F. W. J. (1996). Nasa noise reduction program for advanced subsonic transports. Noise Control Engineering, **44**, 135-140.

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**, 2720.

Tamura, H., Mori, S., and Yamawaki, T. (1978). Textural features corresponding to visual perception. IEEE Transactions on Systems, Man and Cybernetics, 8, 460-472.

Terhardt, E. (1973). Pitch, consonance, and harmony. Journal of the Acoustical Society of America, **55**, 1061-1069.

Tomita, F., Shirai, Y., and Tsuji, S. (1982). Description of textures by a structural analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, **4**, 183-191.

Turner, M. R. (1986). Texture discrimination by gabor functions. Biological Cybernetics, **55**, 71-82.

Uttal, W. R. (1975). An autocorrelation theory of form detection. Hillsdale, NJ: Erlbaum.

van Noorden, L. (1983). Two channel pitch perception. In Clynes, M (ed.) Music, Mind, and Brain: The Neuropsychology of Music, pp. 251-269. New York: Plenum Press.

van Santen, J. P. H. and Sperling, G. (1984). Temporal covariance model of human motion perception. Journal of the Optical Society of America, A, **1**, 451-473.

van Santen, J. P. H. and Sperling, G. (1985). Elaborated reichardt detectors. Journal of the Optical Society of America, A, **2**, 300-321.

Versfeld, N. J. and Vos, J. (1997). Annoyance caused by sounds of wheeled and tracked vehicles. Journal of the Acoustical Society of America, **101**, 2677-2685.

von Bekesy. (1960). Experiments in Hearing. New York: Wiley.

von Bismarck, G. (1974). Sharpness as an attribute of the timbre of steady sounds. Acustica, **30**, 159-172.

Wechsler, H. (1980). Texture analysis a survey. Signal Processing, **2**, 271-282.
Wightman, F. L. (1973a). The pattern-transformation model of pitch. Journal of the Acoustical Society of America, **54**, 407-416.

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

Wu, S., Burns, S. A., Reeves, A., and Elsner, A. E. (1996). Flicker brightness enhancement and visual nonlinearity. Vision Research, **36**, 1573-1583.

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

Yost, W. A. (1996b). A time domain description for the pitch strength of iterated rippled noise. Journal of the Acoustical Society of America, **100**, 1066-1078.

Zwicker, E. and Fastl, H. (1999). Psychoacoustics: Facts and Models. Springer-Verlag, Berlin.

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.

List of Publications

Full papers

Fujii, K., Kita, S., Matsushima, T., and Ando, Y. (2000). The missing fundamental phenomenon in temporal vision. Psychological Research, 64 (2), 149-154.

Fujii, K., Soeta, Y., and Ando, Y. Acoustical properties of aircraft noise measured by temporal and spatial factors. Journal of Sound and Vibration, 241 (1), 69-78.

Fujii, K., Sugi, S., and Ando, Y. Textural properties corresponding to visual perception based on the correlation mechanism in the visual system. Psychological Research (in review).

Fujii, K., and Ando, Y. Relationship between the visual property of texture and subjective preference. Journal of Architectural Planning and Environmental Engineering AIJ (in review).

International conferences

Fujii, K., Kita, S., Matsushima, T. and Ando, Y. (2000). Missing fundamental phenomenon in temporal vision: subjective flicker rates for complex waveforms. 27 th International Congress of Psychology, Stockholm, Sweden.

Atagi, J., Fujii, K., and Ando, Y. (2001). Temporal and Spatial Factors of Traffic Noise and Its Annoyance. 17 th International Congress on Acoustics, Rome, Italy.