A Pilot Study of Autocorrelation Analysis and Subjective Preferences of Images of Camphor Leaves Moving in the Wind

Yoshiharu Soeta and Yoichi Ando
Graduate School of Science and Technology, Kobe University, Rokkodai, Nada, Kobe, 657-8501, Japan

(Received 20 June 2000; accepted 10 April 2001)

To clarify the temporal features of camphor leaves moving in the wind, sequential visual images of their movements were recorded. They were analyzed by using an autocorrelation function (ACF). To discuss the preferred temporal properties of the sequential images, paired-comparison tests of the analyzed images were conducted. The scale values of the preferences can not be attributed solely to wind speed. Remarkably, the most preferred condition is an ACF effective duration of around 0.3–0.4 s.

Keywords: autocorrelation function (ACF), subjective preference, effective duration of the ACF

1. INTRODUCTION

The living environment is composed of spatial and temporal elements. To design pleasant living environments, both spatial and temporal factors in our physical environment must be understood as the physical factors influencing our subjective evaluations of environments. Spatial elements have frequently been considered in architectural design, however, temporal elements have rarely been taken into account. Temporal factors are deeply related to responses in the human brain caused by changes in the physical environment [1-3]. Temporal factors must be considered as well.

Field [4] presented the idea that images from the natural environment are not simply random patterns; they actually possess a number of consistent statistical properties. He also suggested that understanding these properties could lead to a better understanding of how the mammalian visual system codes information. He suggested ways to relate the statistical properties of the natural environment to cortical-cell behavior. Runderman and Bialek [5] studied the statistics of an ensemble of images taken in the woods and showed that distributions of local quantities such as contrast are scale invariant and have nearly exponential tails. Schaaf and Hateren [6] obtained a large set of power spectra for natural images and confirmed previous reports on their variability. Parraga et al. [7] explored the question of whether observed differences in contrast sensitivity reflect actual differences in the chrominance and luminance components of natural scenes. A great deal of research has been done on spatial elements. Temporal elements, by contrast, have been somewhat neglected. The natural environment is constantly changing, so it is important to analyze natural images in the temporal domain. Dong and Atick [8] asserted that natural time-varying images possess spatiotemporal correlations and measured these power spectra. They showed relationships between the temporal aspects of natural images and neural coding. Unfortunately, temporal aspects, with the exception of frequency components, were not measured, and the relationships between natural images and subjective attributes (e.g., subjective preferences) remain unexamined.

The features of the natural environment must be understood before evaluating psychological reactions to it. Ando [9] suggested that the same approach used to analyze sound fields in a concert hall might be used to analyze subjective preferences in the visual environment. He postulated that such preferences may be affected by four factors (1) the lighting level, (2) movement-function properties of reflective surfaces, (3) properties of reflective surfaces, including their color and form, and (4) spatial perception, including distance. The second factor is temporal, the others are spatial. Although a number of studies have dealt with the human response to flickers, they were mostly concerned with the sensitivity to flickering stimuli (e.g. de Lange [10], Kelly [11], Kermers et al. [12]). By contrast, this paper is concerned with preferences in natural time-varying images. As an example, the physical features and subjective preferences of images of camphor leaves moving in the wind were investigated. Leaves are very common in the visual environment, and camphor leaves are often found in the streets and gardens of Japan. Subjective preference was initially chosen as a primitive and convenient measure, it is also a fairly simple and direct dependent variable.

First, the images were analyzed by using an autocorrelation function (ACF). We focused on dynamic changes in the lighting levels of the leaves' reflective surfaces, and investigated the orthogonal properties of the physical factors extracted from the ACF at different wind speeds. Secondly, to evaluate the preferred temporal properties, paired-comparison tests were conducted. Our objective was to clarify the temporal features and determine whether the scale values of the subjective preferences were related to the physical parameters extracted from the ACF. A full characterization of the physical features of natural scenes is virtually impossible. Time-variant images possess a multitude of structures and regularities at many level of complexity. Our ultimate goal is to establish a method for evaluating the temporal factors in the visual environment that affect human preferences. Such a method could be used to design richer environments in which the passage of time can be enjoyed.

2. ANALYSIS OF PHYSICAL FEATURES

2.1 Measurement of gray levels

Temporal images of windblown camphor leaves were recorded between 12:30 and 1:00 p.m. on a sunny winter day using a video camera. The images were taken on a hillside in an open space (25 x 70 m) on the Kobe University Campus. The camera was placed in front of the reflective surfaces of leaves at a distance of 1.1 m and a height of 1.0 m, the elevation angle was 30 degs. The wind speed was concurrently recorded. We analyzed four ten-second images with different degrees of movement. This movement was caused by wind blowing at speeds of 0.71±0.03, 1.15±0.14, 2.40±0.08, and 3.12±0.29 (SD) m/s. The images were digitized using a video capture board at a resolution of 320 x 240 pixels and a gray level of 8 bits/pixel (256 density levels). The frame rate was 30 per second. To find the area, in which the dynamic changes in the lighting level would be analyzed, the gray levels in the central area of each image were changed to 1) 1 x 1, 2) 2 x 2, 3) 4 x 4, 4) 8 x 8, 5) 16 x 16, 6) 32 x 32, 7) 66 x 66, 8) 98 x 98, 9) 130 x 130, and 10) 196 x 196 pixels. This is shown in Fig. 1. They were analyzed by using the ACF. The actual visual angles of these areas were 1) 0.01, 2) 0.02, 3) 0.05, 4) 0.1, 5) 0.2, 6) 0.4, 7) 0.8, 8) 1.2, 9) 1.6, and 10) 2.4 degs. The results show that the ACF factors leveled off at visual angles below 0.05 deg. Consequently, one pixel, the minimum unit, was used

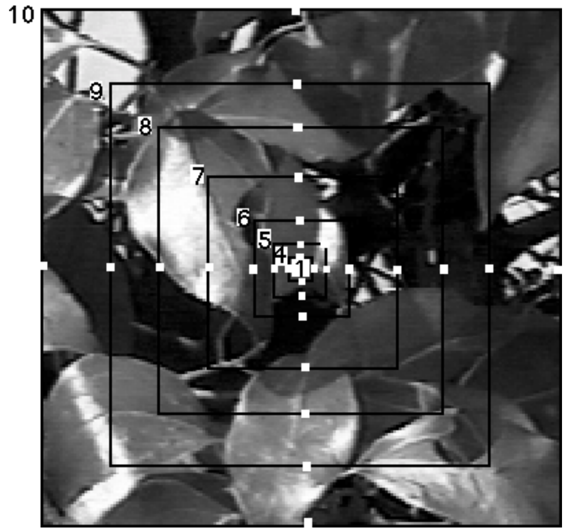

Fig. 1. Areas integrated corresponding to visual angles of 1) 0.01, 2) 0.02, 3) 0.05, 4) 0.1, 5) 0.2, 6) 0.4, 7) 0.8, 8) 1.2, 9) 1.6, 10) 2.4 degs. White dots show positions analyzed corresponding to a visual angle of 0.01.

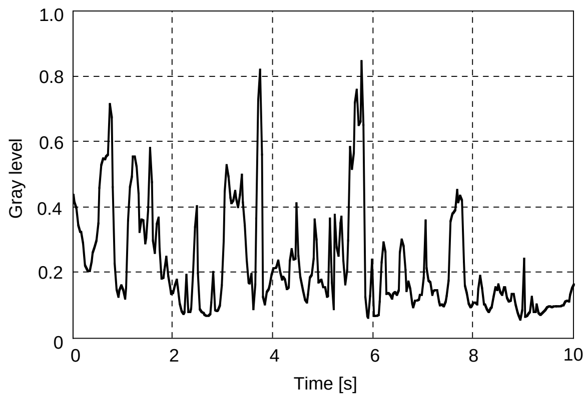

Fig. 2. Sequential gray-level data at a wind speed of 3.1 m/s.

for our analysis. The gray levels at 29 points in the digitized images (the white dots in Fig. 1) were also analyzed by using the ACF. An example of the sequential gray-level data is shown in Fig. 2. We did not explicitly consider any compression of the gray-level data in our analysis because it is believed that this effect is only significant at extremely high or low luminance. Our data did not fall into either category. However, if the luminance in the environment had been high or low, we would have had to take compression into account.

2.2 Analysis of gray levels by using the autocorrelation function (ACF)

The ACF is the fundamental measure of time-sequential data. It has been recently discovered that the delay time of the peak of ACF measurements is closely related to the perceived flicker rates of a lighting source [13]. In addition, important subjective

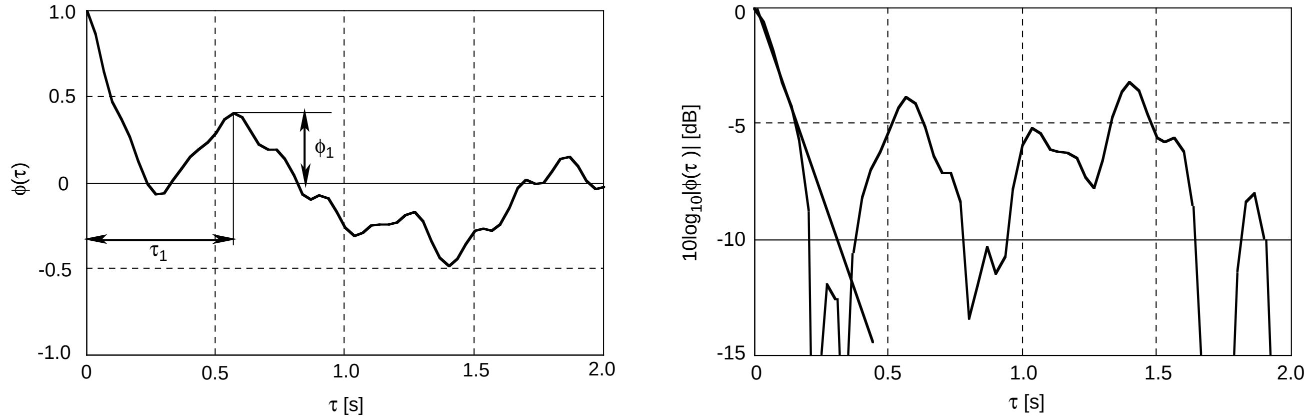

Fig. 3. (a) Example of normalized gray-level ACF analyzed for a wind speed of 3.12 m/s and the definitions of ϕ_1 and τ_1. (b) Example of straight-line regression with initial part of normalized ACF. The effective duration of ACF (τ_e) is defined as the cross point of −10 dB at given delay.

Table 1. The correlation coefficients between factors.

Factor	τ_e	$\log\Phi(0)$	ϕ_1	τ_1
τ_e	1.00	−0.14	0.39	0.31
$\log\Phi(0)$	–	1.00	−0.21	−0.04
ϕ_1	–	–	1.00	0.14
τ_1	–	–	–	1.00

attributes of a sound field, such as the missing fundamental phenomenon and subjective preferences of temporal factors, can be expressed by the ACF of the source signal [9]. The ACF provides the same information as the power-density spectrum of a signal. The subjective attributes are assumed to be described by one of four orthogonal factors that can be extracted from it. An example of a measured ACF is shown in Fig. 3(a). The four factors follow [9,14]. The first is the effective duration of the normalized ACF, τ_e, defined by ten-percentile delay. The second is the energy at the origin of the delay, $\Phi(0)$. The fine structure of a normalized ACF includes peaks and dips and the associated delays. The symbol τ_n represents the delay time, and ϕ_n represents the amplitude of the n-th peak (n = 1, 2, ...). Because there was a certain degree of coherence between parameters τ_n (n = 1, 2, ...), and ϕ_n, these parameters may be represented by the last two factors. These factors are the first peak, ϕ_1 and the delay time, τ_1. A moving ACF can be expressed as a function of time, t, such that [9]:

$$\phi(\tau) = \phi(\tau; t, T) = \frac{\Phi(\tau; t, T)}{\Phi(0; t, T)} \qquad (1)$$

where

$$\Phi(\tau; t, T) = \frac{1}{2T}\int_{t-T}^{t+T} b(s)b(s+\tau)ds; \qquad (2)$$

2T is the integral interval, τ is the time delay, and b(s) is the gray level. The average value of b(s) is zero. To derive the degree of decay in a signal, the effective duration of the ACF, τ_e, must be determined. To do this, a straight-line regression of ACF is done using only the initial declining portion. This regression fits lines to the ACF such that 0 dB > 10 log $|\phi(\tau)|$ > −5 dB. The collect value of τ_e is that point where the line associated with −10 dB intersects the τ axis of the ten-percentile delay. This is shown in Fig. 3(b). The values of τ_e, $\Phi(0)$, ϕ_1, and τ_1 were analyzed as a function of running time for 2T = 2.5 s, with a 100-ms interval. Note that 2T = 2.5 s is the minimum duration of subjective and brain responses [1].

2.3 Results

The cumulative frequencies of the four factors, τ_e, $\Phi(0)$, ϕ_1, and τ_1 at 29 positions are shown in Figs. 4(a) thru (d). As indicated in Table 1, there were no correlations between the four parameters. The median (50%) value of τ_e decreased as the wind speed increased, except at 1.15 m/s. The median value of τ_1 showed the maximum at a wind speed of 0.71, and the minimum at a wind speed of 3.12 m/s. Bifurcation is observed at wind speeds of 0.71 and 1.15 m/s. Linear relationships between wind speeds and four factors were not found. There is a possibility that wind fluctuation and slight change that the anemometer couldn't measure had effects on the results.

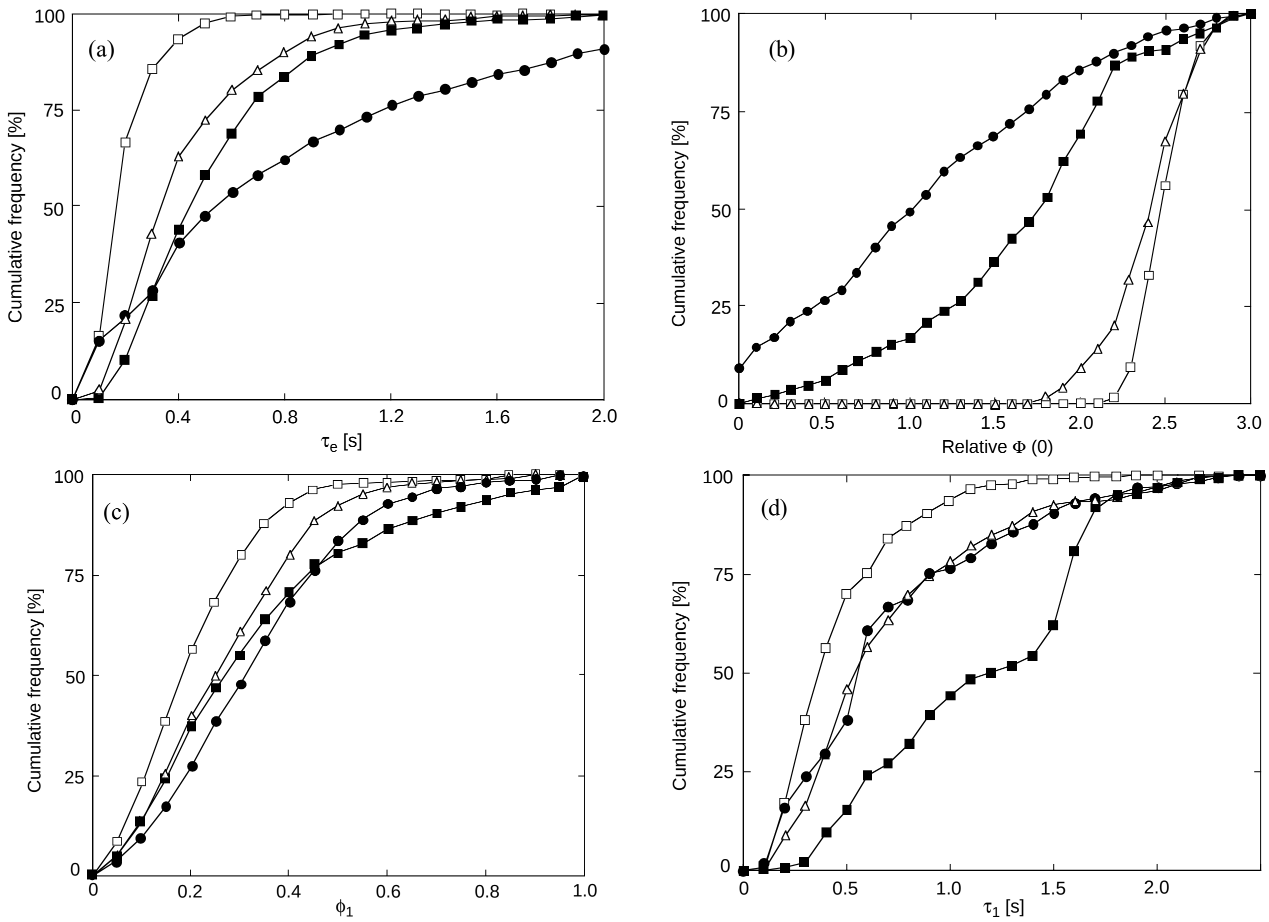

Fig. 4. Cumulative frequencies of (a) τ_e, (b) $\Phi(0)$, (c) ϕ_1, and (d) τ_1 observed at 29 positions. Wind speeds are represented with following symbols: ■ 0.7, ● 1.1, △ 2.4, and □ 3.1 [m/s].

3. SUBJECTIVE PREFERENCES

3.1 Preference tests

The five-second sequential images analyzed here were displayed on a CRT placed in a dark room at a distance of 1.1 m from subjects' eyes. Considering the focal vision, the area of the images presented was within a 1.2-deg visual angle. Ten subjects (22 to 32 years old) with normal or corrected visual acuity participated in paired-comparison testing. Images were presented at 1.0-s intervals, and the interval between comparison pairs was 4.0 s to allow time for the subjects to respond. Each subject judged asked which image they preferred to watch, and they responded by pushing a button. The tests were performed for all combinations of the pairs, i.e., 12 pairs, interchanging the order in each pair per session, a total of 6 sessions were conducted for each subject. Scale values based on the subjective judgment of each subject, were obtained by applying the law of comparative judgment. They are regarded as the linear psychological distance between images [15,16].

3.2 Relationship between scale values of preference and physical parameters

The scale values of the preferences were reconfirmed by a goodness-of-fit based on Thurstone's model (Case V) [17]. By a chi-squared test, agreement between the fitted data and the observed data more than 10% significance level was obtained. It can be said that the results have a certain degree of agreement between them. The global-scale value of the subjective preferences showed a maximum at wind speeds of 2.40 m/s, and a minimum at wind speeds of 1.15 m/s. There was a certain degree of agreement between all the subjects, especially at wind speeds between 1.15 and 2.40 m/s. Relatively large individual differences were found at wind speeds of 0.71 and 3.12 m/s.

Median values of four factors and global-scale value of the subjective preferences are listed in Table 2. Scale values of preference as a function of the median values of τ_e is shown in Fig. 5. By applying analyses of variance to the scale values of preference, the effect of τ_e is found to be significant ($F = 22.7$, $p < 0.01$). The results of the Fisher's LSD test to the

Table 2. Median values of four factors and global-scale value of the subjective preferences.

Wind speed [m/s]	Scale value of subjective preference	Factor			
		τ_e [s]	logΦ(0)	ϕ_1	τ_1 [s]
0.71	0.38	0.44	−2.66	0.27	1.13
1.15	−0.85	0.53	−3.37	0.31	0.53
2.40	0.68	0.32	−1.99	0.25	0.53
3.12	−0.22	0.18	−1.93	0.18	0.33

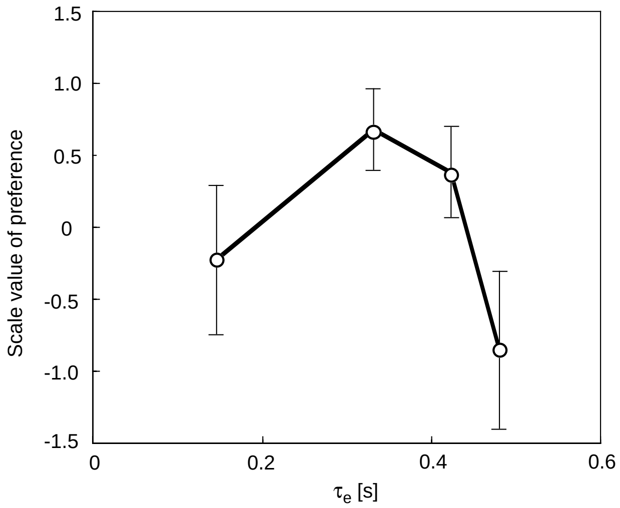

Fig. 5. Scale value of preference for ten subjects obtained by paired-comparison tests. Shown as function of median value of τ_e. Error bars are standard deviation.

scale values of preference didn't show a significant difference at τ_e between 0.32 and 0.44 s only. The most preferred τ_e and ϕ_1 values for each subject were estimated by fitting the suitable polynomial curve to a graph in with plotted scale values. The most preferred τ_e values for the subjects were 0.39, 0.24, 0.38, 0.41, 0.39, 0.30, 0.39, 0.27, 0.28, and 0.25 s. The most preferred ϕ_1 values for the subjects were 0.19–0.29. The scale values of these preferences were not related to the other factors, $\Phi(0)$ and τ_1, extracted from the ACF.

4. DISCUSSION

The statistical properties of static images have been studied for many years [4-7], and interestingly, the power spectra of such images have been found to behave regularly, with values roughly corresponding to $1/f_s^2$, where f_s is the spatial frequency. Knowledge regarding regularities for time-varying images, by contrast, is very limited. Dong and Atick [8] have suggested that the natural time-variant scenes are not-Gaussian. This is consistent with our findings. Furthermore, they have suggested that the spatiotemporal power spectrum is the quantity of data that neurons in the early stages of the visual system can evaluate and utilize in recording visual inputs in the retina and lateral geniculate nucleus. Our analysis characterizes the temporal properties of time-varying images. More systematic studies are needed to understand the spatial properties and apply this knowledge to our understanding of the visual systems.

Vision and auditory senses receive different kinds of signal, light and sound, and these signals are processed in different parts of the brain. Despite this, it is unreasonable to assume that they undergo completely different processes. It is highly possible that the mechanisms are similar. We know that a perceived pitch for a complex tone consists of harmonics which correspond to a fundamental frequency. According to the pattern-transformation model [18], the auditory system detects the pitch not from the temporal waveform per se but by means of peaks in the ACF of the stimulus. The ACF provides information about the pitch, which is independent of the stimulus waveforms. Fujii et. al [13] described a phenomenon demonstrating a possible similarity between the visual and auditory systems. They called it the "missing fundamental". It is also known in auditory pitch sensation. When a signal contains only a number of harmonics without a fundamental frequency, we hear the fundamental frequency as a pitch. They demonstrated the existence of a missing fundamental phenomenon in temporal vision, and explained it by assuming a process in which periodical peaks in the ACF are detected. In our findings the delay time of the first peak of the ACF, τ_1, corresponds to the pitch of the moving leaves. The bifurcation of the τ_1 values represents the hierarchy of the moving leaves, i.e., the movements of the leaves and branches. When the wind is fast, the movements of the leaves have a clear pitch; when it is slow, the pitch is fuzzy.

The magnitude of the normalized ACF at τ_1, ϕ_1, corresponds to the pitch strength of the moving leaves. The fact that the preference evaluation curve as a function of ϕ_1 peaking at 0.19–0.29 indicates that preferences can be obtained at a certain degree of time-variant fluctuation in temporal images. Since there is an apparent little correlation between τ_e and ϕ_1 as show in Table 1, a relationship between scale values of preference and ϕ_1 values is considered to be superficial in this study.

The effective duration of the normalized ACF, τ_e, represents a kind of repetitive feature containing the signal itself. The fact that the preferred values of τ_e are around 0.3–0.4 s indicates that preferences can be obtained at a certain degree of repetition in temporal images. It is notable that in sound fields with reflections, the preferred delay time of this reflection is deeply psychologically related to the τ_e of the source signal [9,14]. In visual fields, it is supposed that center activities may correspond to the direct sound, and a coherent area within focal vision corresponds to the reflections. As a result, it is also meaningful to introduce values of τ_e for subjective preference judgments in the visual environment.

In this paper, an attempt was made to show that subjective preference is closely related to the temporal information contained in natural images. This conclusion is based on a small sample consisting of images of camphor leaves blown by the wind at different speeds. Clearly, such a sample is insufficient to serve as the basis of any form of proof. Further analysis based on a much larger population of images is necessary.

5. CONCLUSIONS

The results of experiments described in this paper demonstrate that:

(1) Orthogonal properties of the physical factors form the ACF, τ_e, $\Phi(0)$, ϕ_1, and τ_1 are extracted.

(2) Remarkably, the preferred values of τ_e are approximately 0.3–0.4s.

ACKNOWLEDGMENTS

This research was supported by a Research Fellowship from the Japan Society for the Promotion of Science for Young Scientists.

REFERENCES

[1] Ando, Y., and Chen, C. (1996). On the analysis of autocorrelation function of α-waves on the left and right cerebral hemispheres and in relation to the delay time of single sound reflection, J. Archi. Plann. Environ. Engng. AIJ., 488, 67-73.

[2] Chen, C., and Ando, Y. (1996). On the analysis of autocorrelation function of α-waves on the left and right cerebral hemispheres and subjective preferences for the reverberation time of music sound field, J. Archi. Plann. Environ. Engng. AIJ., 489, 73-80.

[3] Chen, C., Ryugo, H., and Ando, Y. (1997). Relationship between subjective preference and the autocorrelation function of left and right cortical α-waves responding to the noise-burst tempo, J. Archi. Plann. Environ. Engng. AIJ., 497, 67-74.

[4] Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells, J. Opt. Soc. Am., A 4, 2379-2394.

[5] Runderman, D. L., and Bialek, W. (1994). Statistics of natural images: scaling in the woods, Phys. Rev. Lett. 73, 814-817.

[6] Schaaf, A. van. der., and Hateren, J. H. van. (1996). Modelling of the power spectra of natural images: statistics and information, Vision Res., 36, 2759-2770.

[7] Parraga, C. A., Brelstaff, G., Troscianko, T., and Moorehead, I. R. (1998). Color and luminance information in natural scenes, J. Opt. Soc. Am., A 15, 563-569.

[8] Dong, D. W., and Atick, J. J. (1995). Statistics of natural time-varying images, Network, 5, 517-548.

[9] Ando, Y. (1998). Architectural Acoustics, Blending Sound Sources, Sound Fields, and Listeners, AIP Press Springer-Verlag, New York.

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

[11] Kelly, D. H. (1961). Visual responses to time-dependent stimuli: I. Amplitude sensitivity measurements. J. Opt. Soc. Am. 51, 422-429.

[12] 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 Res., 33, 1997-2011.

[13] Fujii, K., Kita, S., Matushima, T., and Ando, Y. (2000). The missing fundamental phenomenon in temporal vision, Psychol. Res., 64, 149-154.

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

[15] Thurstone, L. L. (1927). A law of comparative judgment, Psychol. Rev., 34, 273-289.

[16] Gullikson, H. (1956). A least square solution for paired comparisons with incomplete data, Psychometrika, 21, 125-134.

[17] Mosteller, F. (1951). Remarks on the method of paired comparisons III, Psychometrika, 16, 207-218.

[18] Wightman, F. L. (1973). The pattern-transformation model of pitch, J. Acoust Soc Am, 54, 407-416.