Matching a Tonal Tempo with Camphor Leaves Moving in the Wind

Yoshiharu Soeta, Yoshinobu Uchida, and Yoichi Ando

Graduate School of Science and Technology, Kobe University, Rokkodai, Nada, Kobe, 657-8501, Japan

(Received 26 December 2000; accepted 5 October 2001)

Matching tests between an image of camphor leaves moving in the wind and periodical sound pulses were performed to find the tonal tempo matching with the image. Subjects viewed a displayed image of camphor leaves moving in different speeds of wind while simultaneously listening to periodical sound pulses having periodicities ranging from 0.08−1.28 s. Results show that the matching period of sound pulses is about half of the delay time at the first peak of the autocorrelation function which was calculated by the gray levels of moving-leaves images.

Keywords: autocorrelation function (ACF), matching a tonal tempo, the delay time at the first peak of the ACF

1. INTRODUCTION

Our visual and auditory systems provide us with the majority of the information we obtain from the environment. Designing optimal physical environments that incorporate the passage of time requires an understanding of how we perceive both visual and auditory information, and of the congruency between them as a function of time [1].

A number of studies have dealt with the relationship between audio and visual perception. For example, Gebhard and Mowbray [2] and Myers et al. [3] investigated matching repeating tone bursts to light pulses. To extend the knowledge regarding the interactions between sound on the impression of video and of video on the impression of sound, some experiments used audiovisual media (e.g., compact disc recordings of commercially-used excerpts of broadcasts and original soundtracks, film excerpts of animals, sequences of cinematic images excerpted from a motion picture and so on) [4-6]. These studies were mostly concerned with the degree of matching and its evaluative dimension.

It has been proposed that the ability to interpret an action film depends on the combination of semantic (i.e., meaning) and formal (e.g., temporal) information flowing across auditory and visual channels [4]. Sugano and Iwamiya [7] concentrated on the temporal information in visual motion and investigated visual and auditory factors to determine the congruency of music and moving images. Subjects had to adjust the speed of a ball, moving in a circular or square pattern, to match changes in music tempo. The pace of the ball tended to reflect the point at which both periods of movement of the ball and periods of musical rhythm were the same or had integer ratios. When the visual motion demonstrated accents, i.e., made a square, this tendency became dominant. Investigators believed accents of visual motion and clearness of musical meter were factors. They also examined the effects of the synchronization between auditory and visual accents and those matching between musical tempo and visual speed on the congruency of motion picture and music [8]. They showed that the effects of the synchronization had a greater influence on judging congruency. The purpose of this study is to clarify the relationship between the temporal factors in images and sounds which contribute to perceptions of congruency between them. We therefore performed a test to match periodical pulses of sound with the time-variant images. Images of camphor leaves moving in the wind were selected as visual stimuli because camphor leaves are often found in the streets and gardens in Japan. Pulses of sound were selected for the sake of simplicity because they have no complicated semantic information or melody. One of our goals is to establish a method for selecting music that matches a lively visual environment. Such a method could be used to design richer auditory and visual environments in which the passage of time can be enjoyed.

2. METHOD

Temporal images of windblown camphor leaves were recorded by using a video camera between 12:30 and 1:00 p.m. on a sunny winter day. The images were taken on a hillside in an open space (about 25 x 70 m) on the Kobe University Campus, as shown in Figure 1. The camera was placed in front of the

Fig. 1. Camphor leaves used in this study; white dots show 29 positions analyzed.

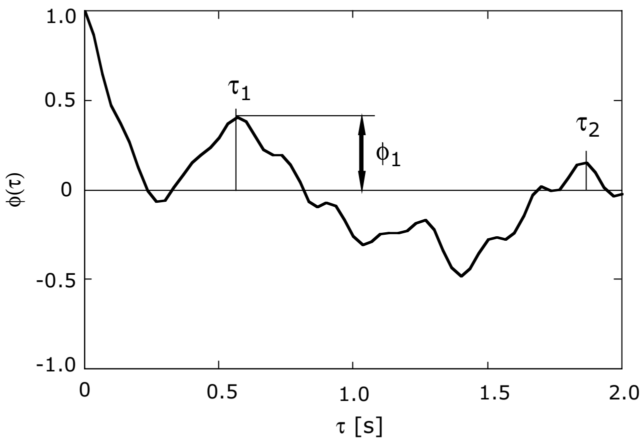

Fig. 2. Definition of temporal factors τ_1, and example of normalized ACF of gray-level of camphor leaves moving in a wind of 3.12 m/s.

reflective surfaces of leaves at a distance of 1.1 m and a height of 1.0 m, and the elevation angle was 30 degrees. The wind speed was concurrently recorded. We selected three five-second images at different wind speeds, so the 0.71±(0.03), 2.40±(0.08), and 3.12±(0.29) m/s values in the parenthesis signify the standard deviation. 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. The gray levels in the digitized images (white dots corresponding to a visual angle of 0.01 in Figure 1) at 29 points were analyzed by using the autocorrelation function (ACF). An example of a measured ACF is shown in Figure 2. The ACF provides the same information as the power spectral density of a signal. There are four factors extracted from the ACF [1,9]. The first is the effective duration of the normalized ACF (NACF), τ_e, defined by ten-percentile delay.

Table 1. Median value of ACF four factors.

	Wind speed [m/s]		
Factor	0.71	2.40	3.12
τ_e [s]	0.44	0.32	0.17
$\log\Phi(0)$	−2.66	−1.99	−1.93
ϕ_1	0.27	0.25	0.18
τ_1 [s]	1.13	0.53	0.33

The second is the energy at zero delay, $\Phi(0)$. The last two factors are the first peak, ϕ_1 and the delay time, τ_1. The fine structure of a NACF 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.

A moving ACF can be expressed as a function of time, t, such that [1]:

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

where

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

$2T$ is the integral interval, τ is the time delay, and $b(s)$ is the gray level as a function of time. The average value of $b(s)$ is zero. The value of τ_1, corresponding to the pitch of the moving leaves, was analyzed as a function of running time for $2T$= 2.5 s, at a 100-ms interval. The other three factors are analyzed and listed in Table 1 [10]. The integral interval of $2T$= 2.5 s was selected in accordance with the psychological present, which is the short duration needed to make subjective preference judgments and for the brain to respond to certain music signals [11,12]. The cumulative frequencies of the τ_1 at 29 positions are shown in Figure 3. The median (50%) value of τ_1 was 0.33 s at a wind speed of 3.12, 0.53 s at a wind speed of 2.40, and 1.12 s at a wind speed of 0.71 m/s.

Five-second sequential pulses, 0.08, 0.16, 0.32, 0.64, and 1.28 s having pulse-widths of 0.063 ms, were used for the matching tests. Seven subjects, from ages 21 to 24 years old and having normal hearing and binocular vision, participated in this study. The subjects were seated in a dark anechoic chamber with comfortable thermal environments where they looked at images and listened to sounds at the same time. The monitor presenting the visual stimuli was placed in the front

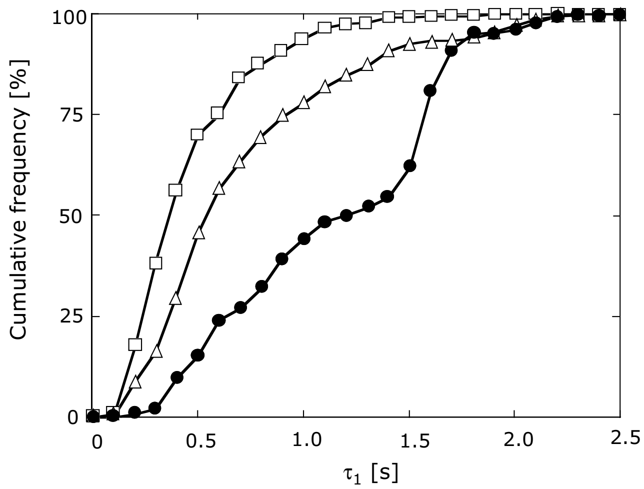

Fig. 3. Cumulative frequencies of values of τ_1 obtained at 29 positions. Average wind speeds represented by: ● 0.7, △ 2.4, and □ 3.1 [m/s].

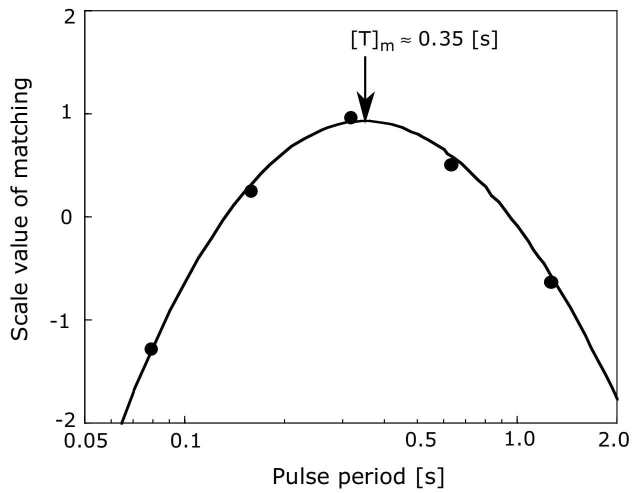

Fig. 4. Example of obtaining most-matched-pulse period $[T]_m$ with movement of camphor leaves (Subject E, wind speed ≈ 2.40 m/s).

of subjects' eye position at a distance of 1.1 m to keep foveal fixation (natural binocular). The loudspeaker was placed on the monitor. The subjects' head and eye positions were unconstrained. The sound pressure level at the center position of the subject's head was kept constant at a peak of 78 dBA. The matching tests were conducted by the paired-comparison method varying the period of pulsed sound. The trial started with the first interval lasting 5 s, followed by a blank duration of 1 s, then a second interval lasting 5 s. During the subsequent 4-s blank duration, the subjects judged which sound pulses more subjectively matched up with the movement of camphor leaves. Ten pairs of combined 5-level periods constituted one series, and 10 series were conducted for all seven subjects for each image of camphor leaves moving at different wind speeds.

Table 2. Most-matched-pulse periods [s] of each subject and global values for images at three different wind speeds.

	Wind speed of image [m/s]		
Subject	0.71	2.40	3.12
A	0.64	0.34	0.15
B	0.58	0.19	0.13
C	0.33	0.34	0.18
D	0.60	0.37	0.24
E	0.65	0.35	0.22
F	0.60	0.56	0.21
G	0.58	0.39	0.22
Averaged (Global)	0.56	0.36	0.21

3. RESULTS

By applying the paired-comparison method, the scale values of the matching judgment of each subject were calculated according to the law of comparative judgment (case V) [13,14] and were reconfirmed by a goodness-of-fit test [15]. The results of each subject indicated the model had a good match between fitted values and the observed values (according to a chi-squared test at 10% significance level).

The most-matched-pulse period $[T]_m$, whose "m" suffix denotes the most matched condition of each subject, was obtained by fitting a suitable polynomial curve to the graph of the scale value. Figure 4 shows an example of the matching evaluation curve for subject E at a wind speed of 2.40 m/s, to obtain $[T]_m \approx 0.35$ s. The $[T]_m$ of each subject and global values for the images at three different wind speeds are listed in Table 2. The global values of the $[T]_m$ were 0.21 s at a wind speed of 3.12 m/s, 0.36 at a wind speed of 2.40 m/s, and 0.56 s at a wind speed of 0.71 m/s. The effect of the image of camphor leaves on the most matched pulse periods was examined using one-way analysis of variance (ANOVA). A significant effect of the image was found ($p < 0.001$).

Figures 5 (a)-(c) show scale values of matching as a function of normalized pulse periods for all subjects and its global values. A matching evaluation curve may be expressed from these figures as:

$$S \approx -\alpha |x|^{\beta} \tag{3}$$

where α and β are the weighting coefficients and $x = \log_{10} T - \log_{10}[T]_m$. To simplify equation (3), the coefficient β may be

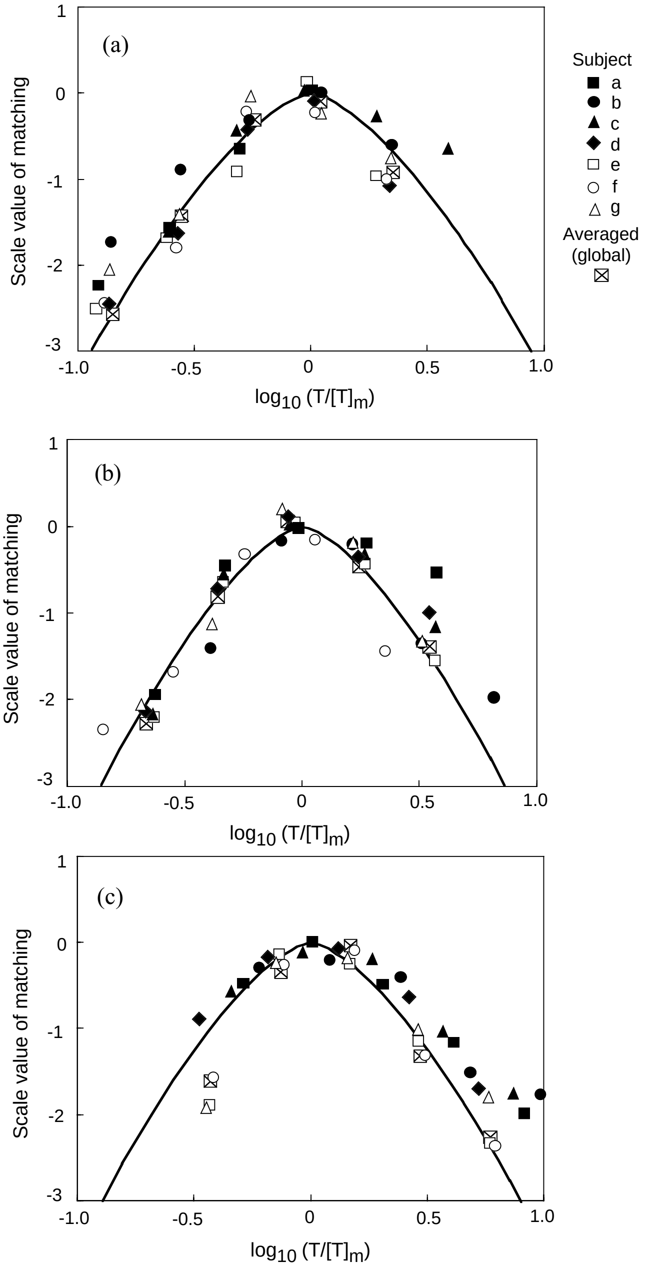

Fig. 5. Scale values of matching as a function of the period of pulse sounds normalized by most-matched-pulse period $[T]_m$ at (a) wind speed ≈ 0.71 m/s, (b) wind speed ≈ 2.40 m/s, and (c) wind speed ≈ 3.12 m/s.

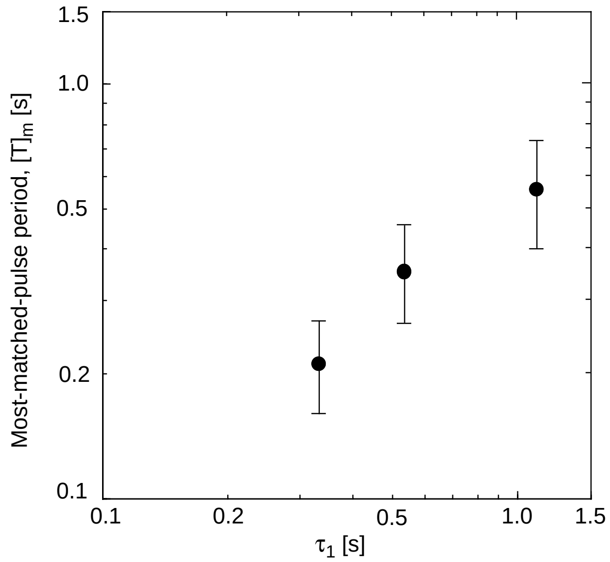

Fig. 6. Relationship between most-matched-pulse period $[T]_m$ and delay time of first peak of ACF, τ_1. Ranges of the $[T]_m$ graphically obtained by all subjects at 0.1 below the maximum matching score.

fixed at a certain value so that the matching evaluation curve can be expressed by the sole coefficient α. After obtaining the $[T]_m$ for each subject, we identified values of α and β. The values of β estimated by using a quasi-Newton numerical method in global values were 1.18 at a wind speed of 3.12 m/s, 1.94 at a wind speed of 2.40 m/s, and 1.43 at a wind speed of 0.71 m/s. The average value of β was 1.52 (≈ 3/2), thus fixed at 3/2 [1,9]. The solid line in Figures 5 indicates the matching curve represented by equation (3) with $\beta = 3/2$. Thus the characteristics of the matching curve can be approximately expressed by only the coefficient α, which is the sole coefficient describing the sharpness of the matching curve, indicating the strength of matching for the $[T]_m$. Table 3 shows the results of the coefficient α of each subject and its global values for each image. The effect of the image of camphor leaves on the α value was examined using the one-way ANOVA, but a significant effect of the image was not apparent.

Figure 6 shows the $[T]_m$ as a function of τ_1. A remarkable finding is that the matched period of sound pulses is approximately half of τ_1, and is nearly equal to τ_e. The other factors, $\Phi(0)$ and ϕ_1, were not related to the most-matched-pulse periods in this experiment.

4. DISCUSSION

In order to describe the phenomenon of the missing fundamental of complex tones, rippled noises, and complex noises, perceived pitch is described in terms of both τ_1 and ϕ_1, which are defined in the fine structures of the NACF of a sound source [16]. The values of τ_1 are defined by the time delay at the first maximum peak of the NACF. The ϕ_1 is the magnitude of NACF at τ_1, which corresponds to the strength of pitch. The period of sequential pulse sound used in this investigation can be determined by τ_1 also. Although metrical accent in music and its clarity were proposed as factors that determine the congruency of music and motion pictures [7], they might be described by the τ_1 and ϕ_1. In the case of matching judgment between motion image and music, therefore, it is recommended to introduce factors of the ACF

Table 3 Coefficient of α of each subject and averaged values calculated by equation (3), when β was fixed at 3/2.

Subject	Wind speed of image [m/s]		
	0.71	2.40	3.12
A	2.27	2.59	2.83
B	1.97	3.00	2.16
C	2.17	3.43	2.31
D	2.56	3.34	3.20
E	3.83	3.88	3.12
F	3.64	3.38	3.23
G	3.27	3.60	2.57
Averaged (global)	3.52	3.74	3.26

of the source signal, τ_1 and ϕ_1.

In this study, the median values of τ_1 of images of camphor leaves moving in the wind were used as representative values. As shown in Figure 3, calculated values of τ_1 converged around the median values at wind speeds of 2.40 and 3.12 m/s. Calculated values of τ_1, however, diverged around 0.6 and 1.5 s at a wind speed of 0.71 m/s. The bifurcation of the τ_1 values represents the hierarchy of the moving leaves, i.e., the movements of the leaves and branches [10]. When the movements of the leaves and branches coincide or when only leaves or branches move, calculated values of τ_1 converge. When they don't coincide, the values of τ_1 can diverge. In the latter case, further experiments will be needed to select representative values, e.g., whether subjects matched pulse sounds and movements of leaves or those of branches.

Sugano and Iwamiya [7] showed that the matched speed of a moving ball tended to be concentrated at the point at which both periods of the movement and those of musical rhythm were the same or had integer ratios. In our experiment, the matched period of sound pulses $[T]_m$ was about half of τ_1 of the image of camphor leaves. The leaves move in one direction and then back again. The period of this movement corresponds to τ_1. It is considered that when the period of sound pulses and the duration of the leaves' movement in one direction matched, the subjects judged that sound pulses subjectively matched up with the movement of the leaves. Consequently, the $[T]_m$ was about half of τ_1 of the image of camphor leaves. At a wind speed of 0.71 m/s, the values of τ_1 diverged around 0.6 and 1.5 s, however, the middle of these diverged values was nearly equal to the median values τ_1. It is conceivable that subjects judged that sound pulses matched up with the tempo of movement between the leaves and branches in this case.

In this study, the matched pulse periods were consistently shorter than the first peak of the ACF, τ_1, of the image. Myers et al. [3] showed that the matched tone burst rates were consistently higher than the flash rates when a tone burst rate was adjusted to match that of an LED. There is a possibility that the auditory system is more sensitive than the visual one in temporal matching judgment.

Matching periods between repeating tone bursts and light pulses have been investigated [2,3]. It has recently been discovered that the delay time of the peak of ACF, τ_1, measurements is closely related to the perceived flicker rates of a lighting source [17]. Therefore, matching periods between tone bursts and flickering lights with compound waveforms may be explained by τ_1 of a lighting source. In this study, τ_1 derived from the images of windblown camphor leaves represents the tempo of moving leaves and τ_1 is closely related to the matching rates of a sound pulse. There is a slight difference of the meaning between τ_1 derived from a lighting source and an image, however, both of τ_1 correspond well to human perception, i.e., the perceived flicker rates or the matching rates of a sound pulse. As a result, in order to describe any subjective judgments for dynamic environments having both auditory and visual components, temporal factors, τ_1, extracted from the ACF should be introduced.

The weighting coefficient β in equation (3), estimated by using a quasi-Newton numerical method, was nearly equal to 3/2. This is consistent with the preference judgment for sound fields and flickering lights [1,9,18]. Equation (3), which represents the matching evaluation curve in the present study, corresponds to the preference evaluation curve of sound fields, which could mean that the theories on subjective preference of sound fields might also be applicable to studies such as the congruency of music and motion sequences.

The effective duration of the normalized ACF, τ_e, represents a kind of repetitive feature containing the signal itself. The relationship between τ_e and subjective preferences of related images of camphor leaves moving in the wind are discussed, and the preferred value of τ_e is around 0.3–0.4 s [10]. The most matched period of sound pulses $[T]_m$ was nearly equal to τ_e in the present study. There is a possibility that a repetitive feature containing an image is related to the matching judgment between music and motion images. Clearly, such a result is insufficient to serve as the basis of any form of proof. Further experiments based on much larger populations of images and sounds are necessary.

5. CONCLUSION

The results show that the matching period of sound pulses is approximately half of the first peak of the autocorrelation function, which was calculated from the gray levels of images.

ACKNOWLEDGMENT

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

REFERENCES

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

[2] Gebhard, J. W., and Mowbray, G. H. (1959). On discriminating the rate of visual flicker and auditory flutter, Am. J. Psychol., 72, 521-529.

[3] Myers, A. K., Cotton, B., and Hilp, H. A. (1981). Matching the rate of concurrent tone bursts and light flashes as a function of surround luminance, Percept. Psychophysics, 30, 33-38.

[4] Bolivar, V. J., Cohen, A. J., and Fentress, J. C. (1994). Semantic and formal congruency in music and motion pictures: Effects on the interpretation of visual action, Psychomusicology, 13, 28-59.

[5] Lipscomb, S. D., and Kendall, R. A. (1994). Perceptual judgement of the relationship between musical and visual components in film, Psychomusicology, 13, 60-98.

[6] Iwamiya, S. (1994). Interaction between auditory and visual processing when listening to music in an audio visual context: 1. Matching 2. Audio quality, Psychomusicology, 13, 133-154.

[7] Sugano, Y., and Iwamiya, S. (1999). Effects of synchronization between musical rhythm and visual motion on the congruency of music and motion picture (in Japanese). Journal of Music Perception and Cognition, 5, 1-10.

[8] Sugano, Y., and Iwamiya, S. (2000). The effects of synchronization between auditory and visual accents and those of matching between musical tempo and visual speed on the emotional impression of combinations of motion picture and music (in Japanese). J. Acoust. Soc. Jpn., 56, 695-704.

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

[10] Soeta, Y., and Ando, Y. (2001). Autocorrelation analysis and subjective preferences of images of camphor leaves moving in the wind. Journal of Temporal Design in Architecture and the Environment, 1, 6-11.

[11] 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.

[12] Fraisse, P. (1984). Perception and estimation of time, Ann. Rev. Psychol., 35, 1-36.

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

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

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

[16] Ando, Y., Sakai, H., and Sato, S. (2000). Formulae describing subjective attributes for sound fields based on a model of the auditory-brain system, J. Sound Vibration, 232, 101-127.

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

[18] Soeta, Y., Uetani, S., and Ando, Y. (2002). The relationship between subjective preference and alpha wave activity in relation to temporal frequency and mean luminance of a flickering light, to be published in J. Opt. Soc. Am. A.