Duration Sensation When Listening to Pure Tone and Complex Tone

Kazi Saifuddin, Takaji Matsushima and Yoichi Ando

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

(Received 12 Nobember 2001; accepted 20 December 2001)

The duration sensation (DS) is a primary sensation of the sound stimuli of complex-tone, because musical notes include information of pitch and duration as well as loudness. Paired-comparison tests while changing the residue pitch or fundamental frequency, which correspond to the delay time of the first peak (τ_1) extracted from the autocorrelation function (ACF) were conducted. In Experiment 1, the white noise was used as a reference, and complex-tones (two pure-tone components) with three fundamental frequencies (250, 500, and 1000 Hz) were used as the test stimuli. Ten subjects judged whether or not the white-noise duration was longer than the complex-tone duration. Throughout investigation, the sound pressure level was fixed at 80 dB(A). And, the rise and fall times, defined as the time at a −3 dB drop from the steady level were kept a constant 1 ms. Results show that the DS of pure-tone and bandpass-noise stimuli is judged to be longer when is longer τ_1. Results of experiment 2 show that the DS of complex-tone stimuli is the same as that of the pure tone, if values of τ_1 are the same.

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

1. INTRODUCTION

A model of the auditory-brain systems related to the autocorrelation function (ACF) and interaural cross-correlation mechanisms (IACCs) of sound stimuli was described in 1998 by Ando [1]. Based on this model, he proposed a primary sensation theory of a sound stimulus, $p(t)$, located in front of a listener [2]. In this theory, four primary sensations are recognized: loudness, pitch, timbre, and duration sensation (DS). The DS in the auditory-brain system is closely connected to the factors extracted from the autocorrelation function (ACF) as a signal process [3-5]. The ACF is equivalent to the power density spectrum of the $p(t)$. Thus, the ACF can be defined by

$$\Phi(\tau) = \frac{1}{2T}\int_{-T}^{+T} p(t)p(t+\tau)dt, \tag{1}$$

where $p(t)$ is the sound signal at the entrances of the ears, τ is the delay time, and $2T$ is the integration interval. When the ACF is normalized, $\phi(\tau)$ can be defined by

$$\phi(0) = \frac{\Phi(\tau)}{\Phi(0)}, \tag{2}$$

where $\Phi(0)$ is the energy at the beginning ($\tau = 0$) of the ACF within each integration interval ($2T$). The perception of pitch is depends on τ_1, the time delay of the first peak in the ACF. It was found throughout these studies that higher values of τ_1 or lower pitches in the ACF of pure-tone and bandpass-noise stimuli (with and without reverberation effects) produce a longer DS [3-5].

We conducted two experiments: (1) DS was examined in relation to the τ_1 of the ACF of complex-tone stimuli for three fundamental frequencies (250, 500, and 1000 Hz) compared with that for white-noise stimuli. (2) DS was compared directly between complex-tone and pure-tone stimuli. The complex tones of the 3000 Hz and 3500 Hz components (fundamental frequency 500 Hz) were compared with the pure-tone stimuli of 500 and 3000 Hz.

2. EXPERIMENTAL METHOD

2.1 Source Stimuli

Experiment 1: Three complex-tone stimuli with three fundamental frequencies f_f = 250 (components of 3000 Hz and 3250 Hz), 500 (components of 3000 Hz and 3500 Hz) and 1000 Hz (components of 3000 Hz and 4000 Hz) were used (Fig. 1). Thus, complex-tone stimuli consisted of two pure-tone components. The starting phases of the two harmonic components were set to zero. Physical duration of three complex tones were 150 ms. Ten white-noise stimuli with durations of 140 to 230 ms in 10-ms steps were compared.

Experiment 2: one complex-tone stimulus with a

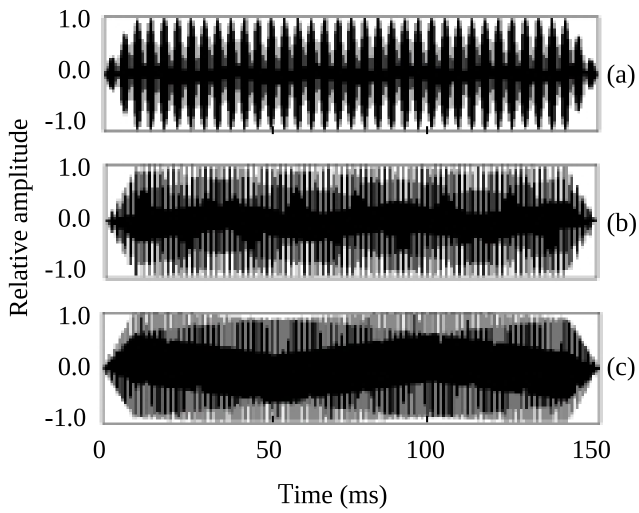

Fig.1. Actual waveforms of three complex tones used. (a) Complex components: 3000 HZ and 3250 Hz with fundamental frequency $f_f = 250$ Hz. (b) 3000 Hz and 3500 Hz with $f_f = 500$ Hz. (c) 3000 Hz and 4000 Hz $f_f = 1000$ Hz. Physical duration of each stimulus was 150 ms including 1-ms rise and fall times.

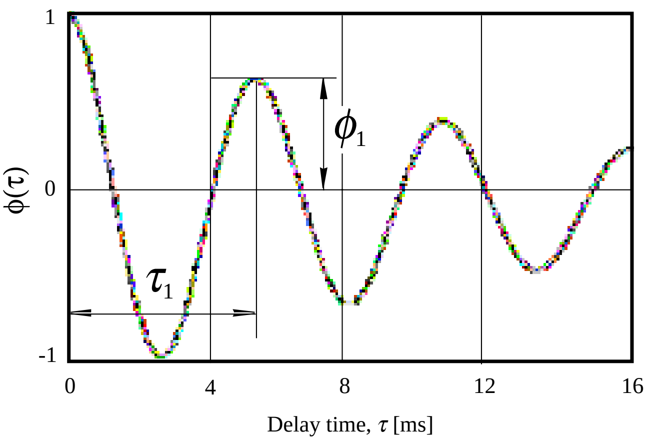

Fig. 2. Factors τ_1 and ϕ_1 defined by the delay time and amplitude, respectively, of the first peak of the autocorrelation function (ACF).

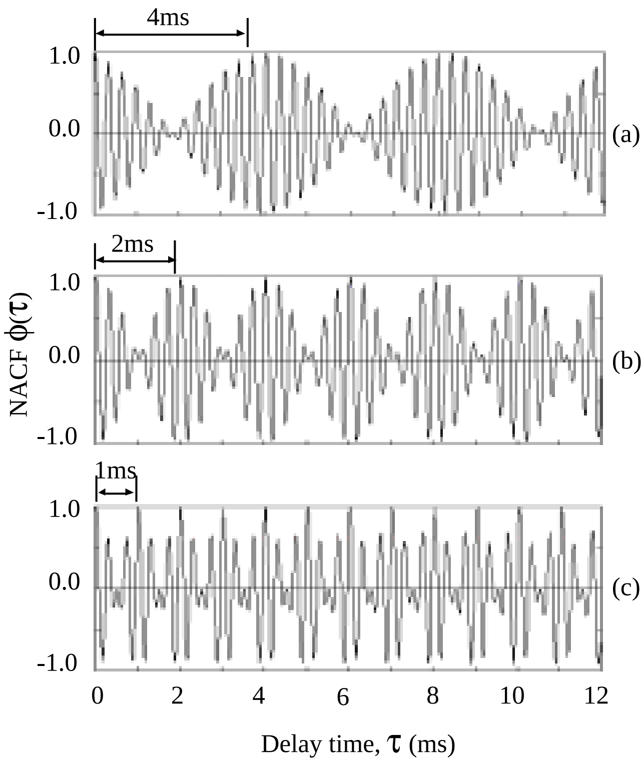

Fig. 3. Normalized autocorrelation function (NACF) of three complex-tone stimuli. (a) Complex components: 3000 Hz and 3250 Hz with fundamental frequency $f_f = 250$ Hz. (b) 3000 Hz and 3500 Hz with $f_f = 500$ Hz. (c) 3000 Hz and 4000 Hz $f_f = 1000$ Hz.

fundamental frequency of 500 Hz and two pure-tone stimuli (500 and 3000 Hz) were used. The stimuli in both experiments included 1-ms rise and fall times, defined as the time at a –3 dB drop from the steady level.

2.2 ACF Analysis

The ACF of the three complex-tone and two pure-tone stimuli reproduced in a soundproofed chamber was analyzed with the integration interval ($2T = 2$ s). As shown in Fig. 2, ACF components, the factors τ_1 and ϕ_1 were defined, respectively, as the delay time and amplitude of the first maximum peak. The amplitudes of the major peaks (ϕ_1, ϕ_2, ϕ_3, ..., ϕ_n) of the complex tones and of all the peaks of the pure tones were found to be almost parallel. The time delay of the first peak (τ_1) of the complex-tone stimuli as shown in Fig. 3 corresponds to the missing fundamentals, if the components are below about 5000 Hz and $f_f < 1200$ Hz [3-5].

2.3 Stimuli Presentation

The source stimuli were presented in a dark soundproof chamber from a single loudspeaker at a horizontal distance of 74 (±1) cm from the center of the head of the seated listener. The sound pressure levels (SPLs) of the stimuli were kept constant at 80 dB(A).

Experiment 1, the complex-tone stimulus was presented first with a constant 150-ms duration. The white-noise stimuli were presented next. The intra- and inter-pair (response time) gaps, respectively, were 1 and 3 s. Each pair of stimuli was presented five times randomly within every session for each subject.

Experiment 2, pairs consisting of two pure tones and one complex tone were presented randomly to obtain scale values. Three durations (140, 150, and 160 ms) for each stimulus were used. The total number of pairs were 9 (all combinations). The parameter was the pitch in relation to the τ_1 of the ACF both of the pure-tone and complex-tone stimuli.

2.4 Subjective Tests

Ten students participated in both experiments as subjects of normal hearing levels (22 and 36 years old). Each subject was seated in the dark soundproof chamber in front of the loudspeaker. Experiment 1, they were asked to judge whether the duration of the first or second of a pair of stimuli was longer or shorter than that of the other. They reported to push a button only when the second stimulus (white noise) judged longer than the first stimulus. Each subject participated in three sessions with three different sets of stimuli. The pause between sessions was 5 minutes for both experiments avoiding fatigue effects. Experiment 2, the subjects reported pushing the button in one of two ways to indicate whether the first or second stimulus judged longer.

3. RESULTS

3.1 Experiment 1

The cumulative frequencies judgments of the subjects for the three different stimuli presented first in the pair are shown in Fig. 4. The reference physical duration was 150 ms, and the 50% line of the psychometric function of the duration judgments was figured out. Values on the 50% line of the cumulative frequencies for all of subjects were 190.0, 185.0, and 180.0 ms (26.7, 23.3, and 20.0% longer than 150 ms) for three complex-tone stimuli of f_f = 250, 500, and 1000 Hz, respectively. In this figure, values on the 50% line of the cumulative frequencies of the duration judgment with the individual differences (standard deviation = 7.37) are plotted in Fig. 5 also. Significant differences between judged durations of f_f = 250 Hz and 500 Hz and of f_f = 250 Hz and 1000 Hz are achieved.

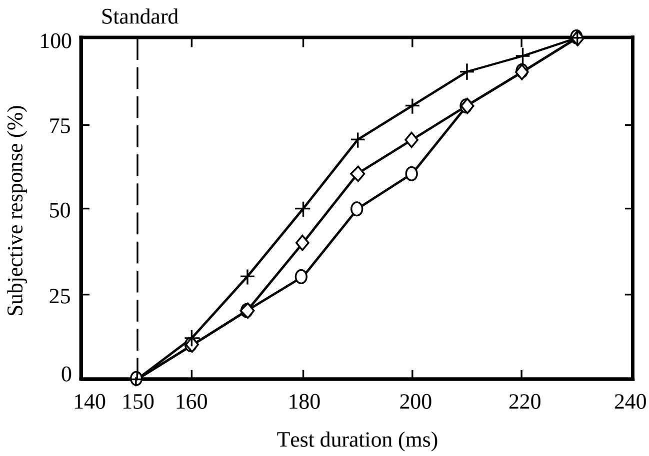

Fig. 4. Cumulative frequencies of judged responses of duration in paired-comparison test for three stimuli. Stimuli were three complex tones with different fundamental frequencies (f_f). Dashed line indicates standard-stimulus duration of 150 ms. × : f_f = 250 Hz, ◇ : f_f = 500 Hz, ○ : f_f = 1000 Hz.

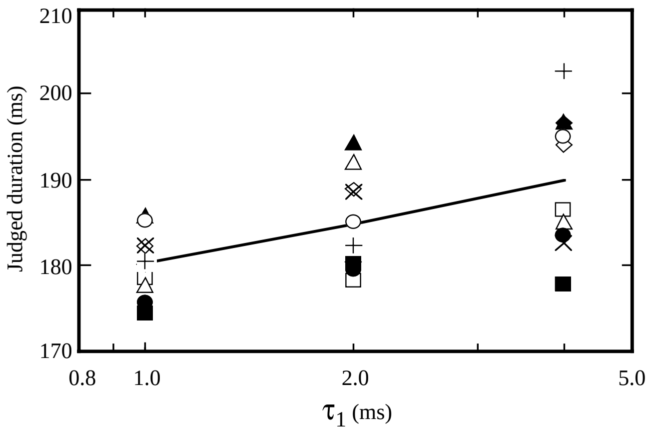

Fig. 5. Judged durations for complex-tone stimuli with three fundamental frequencies. Different symbols indicate different individual scores of DS obtained by ten subjects. ■ : JA, □ : TK, ◆ : KF, ◇ : RS, ▲ : NA, △ : TH, ● : KS, ○ : NK, × : YO, + : NA . Solid line shows average values.

3.2 Experiment 2

The results from experiment 2 provide the scale value (SV) obtained using the "simple method of calculating individual subjective responses of paired-comparison test" [6] between pure-tone and complex-tone stimuli. The average data are shown in Fig. 6. The individual scale values of the judged durations for the ten subjects are shown in Fig. 7. The three curves in Fig. 6 show the results for comparisons between pairs of pure-tone and pure-tone, complex-tone and complex-tone, and pure-tone and complex-tone stimuli.

As shown in Fig. 6, the duration judgments of complex-tone stimuli are quite similar to those of the pure-tone, when the value of τ_1 (τ_1 = 2 ms) is the same or the same pitch. The DS of the pure tones (3000 Hz; τ_1 = 0.33 ms) are found to be significantly shorter ($p < 0.01$) than those of the pure tone (frequency: 500 Hz; τ_1 = 2 ms) and complex tone (fundamental frequency: 500 Hz; τ_1 = 2 ms).

Scale values of individuals are shown in Fig. 7, which are obtained by the method in Appendix. Results of test of goodness of fit of the model are listed in Table 1 for ten subjects. These individual data confirm almost above mentioned results in the range of the standard deviation.

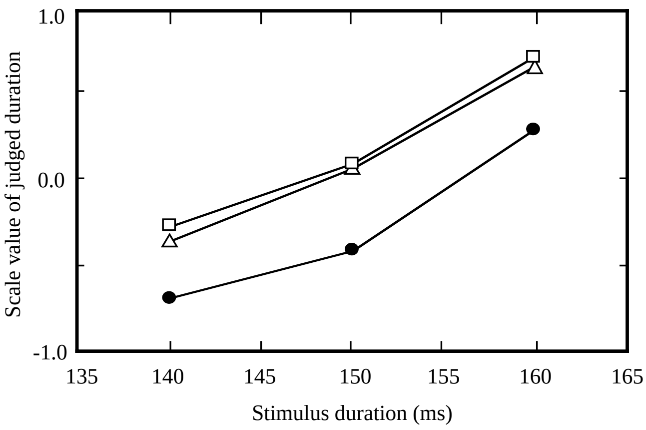

Fig. 6. Scale values obtained by paired-comparison tests. □ : complex tone, f_f = 500 Hz with 3000 Hz and 3500 Hz pure-tone components; △ : 500-Hz pure tone; ● : 3000-Hz pure tone.

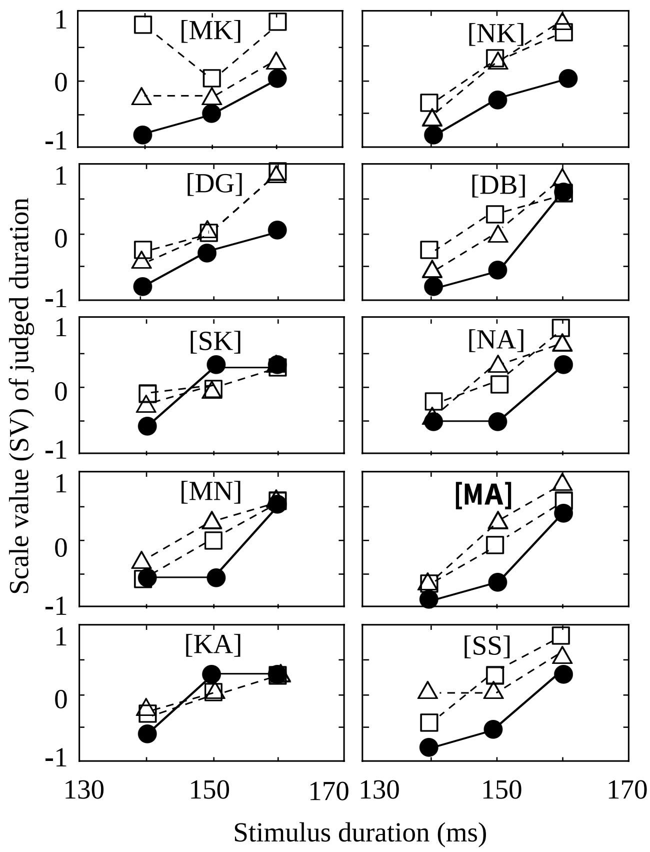

Fig. 7. Individual scale values of obtained by paired-comparison tests [Appendix]. □ : complex tone, f_f = 500 Hz with 3000 Hz and 3500 Hz pure-tone components; △ : 500-Hz pure tone; ● : 3000-Hz pure tone.

4. DISCUSSION

4.1 Experiment 1

The DS increases with decreasing τ_1 of the fundamental frequency of the complex-tone. The judged duration for each subject and the average duration of the three stimuli with a complex tone are significantly longer ($p < 0.01$) between the complex tones with fundamental frequencies of 250 and 1000 Hz.

Figure 4 shows that the DS are different when the physical duration (PD) of the stimuli are different. The duration sensation may be expressed as a function of τ_1 and PD, such that

$$DS \approx f(\tau_1,\ PD). \tag{3}$$

4.2 Experiment 2

Results of experiment 2 indicate that the duration sensations are almost the same for the pure-tone (τ_1 = 2 ms) and complex-tone (τ_1 = 2 ms) stimuli. However, DS of the pure tone (τ_1 = 0.33 ms: 3000 Hz) is significantly ($p < 0.05$) shorter than that of the pure-tone (τ_1 = 2 ms: 500 Hz) as well as the complex-tone (τ_1 = 2 ms: 500 Hz). This implies that the in the ACF of the sound stimulus can determine the duration sensation (DS). As shown in Fig. 6, the scale values of DS between the two pure tones (τ_1 = 2 and 0.33 ms) are almost parallel, so that effects of τ_1 and PD on DS are independent holding Eq. (3). It has been achieved that this tendency is true for the individual DS with a few marked exceptions. For instance, both values of d and λ showing poorness of fit for subjects MK and KA are exceed 19 and 13% (K > 7) as indicated in Table 1, respectively.

On the basis of the results of duration judgments a new concept of DS can be introduced. The DS is considered one of primary psychological phenomena of the stimuli similar to pitch and loudness. Thus, results may be interpreted based on the "auditory-brain model" [1] and the "theory of primary sensation" [2]. The time duration investigated (140–230 ms) corresponds to rather short musical note. For this kind of practical purpose, further investigations are recommended.

5. CONCLUSIONS

Significant results of this study are briefly summarized below.

1. The DS of complex-tone stimuli with lower fundamental frequencies or a lower pitch (a larger τ_1) is judged to be longer than that of stimuli with higher fundamental frequencies or a higher pitch.
2. Effects of the τ_1 extracted from the ACF on the DS are almost the same on the scale value for the pure-tone (τ_1 = 2 ms) and complex-tone (τ_1 = 2 ms) stimuli. And, DS of the pure-tone stimulus (τ_1 = 0.33 ms) with higher pitch is significantly shorter than that of pure-tone and complex-tone stimuli with lower pitch (τ_1 = 2 ms).

Table 1 Results of tests of goodness of fit for ten subjects (see Appendix).

Subject	K	d	λ	σ[a)]
MK	8	22.2	0.21	0.57
DG	6	16.7	0.02	0.54
SK	6	16.7	0.05	0.31
MN	4	11.1	0.09	0.50
KA	7	19.4	0.13	0.52
NK	6	16.7	0.15	0.61
DB	4	11.1	0.06	0.62
NA	5	13.9	0.04	0.52
MA	5	13.9	0.13	0.65
SS	6	16.7	0.03	0.56

3. Thus, DS can be expressed as a function of $\tau_1 = 2$ and PD, but it cannot be expressed by the frequency components for the complex tone.

APPENDIX
METHOD OF OBTAINTING INDIVIDUAL SCALE VALUE [1]

Considering the fact that members of an audience, including children and aged people, are quite diverse, a method for subjective judgments should be as simple as possible. For this purpose, the paired-comparison method is selected. Another method, for example, the method of magnitude estimate is too difficult for most people, except for the skilled subjects in laboratory experiments. The paired-comparison method usually needs a number of judgments for a single pair. However, from a single observation datum for a set of sound fields, an approximate method is described for obtaining the scale value of subjective preference. The method is based on the law of comparative judgment using the linear range of normal distribution between the probability and the scale value.

The probability that a sound field B is preferred to another sound field A is expressed by

$$P(B > A) = \frac{1}{\sigma_d\sqrt{2p}}\int_0^{\infty}\exp\left(-\frac{(X_d - \langle X_d\rangle)^2}{2\sigma_d}\right)dX_d$$
$$P(B > A) = \frac{1}{\sqrt{2p}}\int_{Z_{ab}}^{\infty}\exp\left(-\frac{y^2}{2}\right)dy \qquad \text{(A.1)}$$

where

$$Z_{ab} = -\frac{\langle X_d\rangle}{\sigma_d} \qquad \text{(A.2)}$$

$\langle X_d \rangle$ is the average scale value between sound fields A and B, $X_d = X_b - X_a$, if σ_d is being used as the unit for the scale value: $\sigma_d = 1$.

Thus,

$$P(B > A) = \mathrm{erfc}[Z_{ab}] \qquad \text{(A.3)}$$

$$Z_{ab} = \mathrm{erfc}^{-1}[P(B > A)] \qquad \text{(A.4)}$$

The first order approximation of the Taylor series of Equation (A.4) is given by

$$Z_{ab} = \sqrt{2\pi}(P(B < A) - 1/2) \qquad \text{(A.5)}$$

The linear range can be obtained for

$$0.05 \le P(B > A) \le 0.95.$$

Let us now consider the number of sound fields, given by F (i, j = 1, 2, ..., F), and suppose a single response of each pair, for simplicity. Then the probability P(B > A) in Eq. (A.5) is replaced by [6]

$$P(i > j) = \frac{1}{F}\sum_{i=1}^{F} Y_i \qquad \text{(A.6)}$$

where $Y_i = 1$ responds to a preference of i over j, $Y_i = Y_j = 0.5$ (i = j), while $Y_i = 0$ corresponds a preference of j over i. In order to improve the precision of the probability P(i > j), a certain minimum number of sound fields within the linear range are needed, to keep the accuracy high when using Eq. (A.6). This is performed by a preliminary investigation, avoiding any extreme sound field outside the linear range. In this a manner, the scale value $S_i = Z_{ij}$ (i = 1, 2, ..., F) may be obtained approximately, when Z_{ij} with P (j > i) is obtained by Eq. (A.5).

Next, let us consider an error in a single observation. The poorness of fit for the model is defined by

$$\lambda = \frac{\sum_{(i,j)}\left|S_i - S_j\right|_{poor}}{\sum_{(i,j)}\left|S_i - S_j\right|}, \quad 0 \le \lambda \le 1 \qquad \text{(A.7)}$$

where

$$\begin{aligned}\left|S_i - S_j\right|_{poor} &= S_i - S_j > 0, \quad \text{if} \quad Y_i = 0 \\ &= 0, \qquad\qquad\quad \text{if} \quad Y_i = 1\end{aligned} \qquad \text{(A.8)}$$

Thus, in spite of j is being preferred over i ($Y_i = 0$), it is possible that $S_j - S_i < 0$, and the amount $S_j - S_{i\,poor}$ is added as in Eq. (A.7).

When i is preferred over j ($Y_i = 1$), it is natural that $S_j – S_i > 0$, and the amount is not added to the numerator.

The value of λ corresponds to the average error of the scale value. This should be small enough, say, less than 10%.

Another observation is that, when the poorness number is K, according to the condition expressed by Eq. (A.8), then the percentage of violations d is defined by

$$d = \frac{2K}{N(N-1)} \times 100 \qquad (A.9)$$

ACKNOWLEDGMENTS

We thank those who participated in this study as experimental subjects.

REFERENCES

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

[2] Y. Ando (2001). A theory of primary sensations and spatial sensations measuring environmental noise. J. Sound Vib., 241, 3-18.

[3] K. Saifuddin, H. Sakai and Y. Ando (2001). Duration sensation of simulated environmental noise in relation to its autocorrelation function. J. Sound Vib., 241, 117-127.

[4] K. Saifuddin and Y. Ando (2001). Duration sensation when listening to bandpass noises in a room. Journal of Temporal Design in Architecture and the Environment, 1, 27-32. (http://www.jtdweb.org/journal/).

[5] Y. Ando, K. Saifuddin and S. Sato (2001). Duration sensation when listening to bandpass noises. J. Sound Vib., (to be published).

[6] Y. Ando and P. K. Singh (1996). A simple method of calculating individual subjective responses by paired-comparison tests. Mem. Grad. School Sci. & Technol., Kobe Univ., 14-A, 57-66.