| | |
|---|---|
| Title | Estimation of reaction time for birdsongs and effects of background noise and listener’s age |
| Author(s) | Shimokura, Ryota; Soeta, Yoshiharu |
| Citation | Applied Acoustics. 2022, 194, p. 108785 |
| Version Type | AM |
| URL | https://hdl.handle.net/11094/88312 |
| rights | © 2022. This manuscript version is made available under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. |
| Note | |

# Estimation of reaction time for birdsongs and effects of background noise and listener's age

Ryota Shimokura* *: corresponding author

Graduate School of Engineering Science, Osaka University

Room D436, 1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan

E-mail: rshimo@sys.es.osaka-u.ac.jp, Tel/Fax: +81-6-6850-6376

Yoshiharu Soeta

Medical and Biological Engineering Research Group, Biomedical Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Ikeda, Osaka 563-8577, Japan

E-mail: y.soeta@aist.go.jp, Tel: +81-72-751-9496

## Highlights

- Recordings of birdsong are often used as an auditory signal for visually-challenged people in public spaces in Japan.
- To examine the detectability of these sounds, we measured reaction times (RTs) for six types of birdsong.
- The RT for the Cuckoo were the fastest in the silent condition, and the RTs for six birdsongs could be estimated on the basis of the durations required to reach a specified sound exposure level.
- In the noisy conditions, spectral masking via bandpass noise prolonged the RT, while the bandpass noise with a center frequency higher than that of the birdsong tended to shorten the RT.

## ABSTRACT

Recordings of birdsong are often used as an auditory signal for visually-challenged people in public spaces in Japan. To examine the detectability of these sounds, we measured reaction times (RTs) for six types of birdsong: Jay, Cuckoo, Himalayan cuckoo, Japanese grosbeak, Japanese white-eye, and Japanese bush warbler. We presented birdsong stimuli with irregular timing to 20 younger and 20 older participants in combination with silence, white noise, or bandpass noise with a center frequency of 1, 2, or 4 kHz. Regardless of the output levels and participant age, the RT for the Jay, Cuckoo, and Himalayan cuckoo songs was lower than 0.75 sec in the silent condition, and the averaged RT obtained by the younger participants could be estimated on the basis of the duration required to reach a specified sound exposure level for the birdsong stimuli. In the noisy conditions, spectral masking via bandpass noise prolonged the RT, while the bandpass noise with a center frequency higher than that of the birdsong tended to shorten the RT, especially in older participants.

**Keywords**: reaction time, birdsong, sound exposure level, spectral masking

## 1. Introduction

Birdsong is often used in public spaces in Japan (e.g., entrance gates of stations and crosswalks) to serve as auditory signals for visually-challenged people. The sound sign also informs them positions of stairs or public toilets. These natural or synthesized natural sounds were chosen because they have shorter simple reaction times (RT) than artificial sounds [1], and thus is associated to relatively high detectability although other factors (e.g., intensity, frequency and timbre) may influence it. And the other reason to choose the birdsongs is that they are hearable clearly and localized easily in environmental noises [2]. However, the birdsongs are different according to species, communication types, or even individual bodies, the RT should be estimated by their acoustical characteristics to design the detectable sound sign.

RT is known to decrease monotonically as sound intensity increases [3-5], and it is correlated with the loudness of a sound [6-9]. Although some psychoacoustics studies have used steady-state sounds (i.e., pure tones with a constant frequency), birdsongs have waveforms that vary over time in envelope and frequency, so the loudness necessarily varies as a function of time. Previous studies have given the RT of a 1-kHz pure tone with various rise-times (i.e., the duration required for a signal to approach a maximum amplitude) [10,11]. These studies reported that the RT decreased when the rise-time was below 50 msec (i.e., steeper), but did not change considerably when the rise-time was higher than 50 msec. Furthermore, below 50 msec, the RT decreased only by 10–30 msec. This is because the total length of the pure tone was longer than the critical duration (CD), which is defined as the duration in which loudness remains constant. The CD has mostly been reported to range from 100 to 200 msec [12,13].

How can the loudness of a time-varying signal be evaluated? To evaluate the loudness of a non-steady-state sound, Namba et al. (2008) [14] introduced the idea of the sound exposure level ($L_{AE}$) [dB]. The $L_{AE}$ is the summation of the sound intensity within a certain duration, generally set at 1 sec. An impulsive sound is defined as a brief burst of sound pressure with a duration that is usually less than 1 sec [15], so that study used an amplitude-modulated and steady-state 1-kHz pure tone with durations from 80 to 1280 msec. Regardless of the envelope shape, the $L_{AE}$ (i.e., total energy) is highly correlated with the loudness of short and impulsive sounds ($r = 0.985$).

Another approach is measuring the instantaneous loudness of time-varying sound ($N$) [sone] using the Zwicker loudness model [16]. The algorithm for calculating time-varying loudness follows that for stationary sounds, but considers the nonlinear decay of the human hearing system, temporal summation, and forward masking. Although $N$ is a function of time, the purpose of this calculation is to estimate the overall loudness perceived from a time-varying sound. A percentile loudness $N5$, which is the loudness below 95% of the calculated instantaneous loudness, is generally used [16], but other represented values, such as the mean of the loudness level over time [17] and root mean cubed loudness [18] are recognized as being highly correlated with the overall loudness.

To describe the detectability of the birdsong using a physical characteristic in the present study, we measured RTs while participants listened to birdsongs of six species whose spectral characteristics were different, and we attempted to estimate the RTs for birdsongs based on the time-functional sound exposure level ($L_{AE}(t)$) and instantaneous loudness ($N(t)$). We hypothesized that the RTs for birdsongs would be determined based on the duration between the sound onset and the time at which the $L_{AE}$ or $N$ reached a certain value. By conducting individual hearing threshold tests before the RT measurements, we were able to adjust the output levels in the RT measurements to low sensation levels (SLs). The RT for a sound with a lower SL is largely influenced by the frequency [8,9], so we expected that the RTs would vary depending on the type of birdsong.

To realize a realistic situation in a public space, the participants in this study were divided into younger and older age groups. The RT measurement was conducted for both a silent condition and noisy conditions to assess the remarkableness of birdsong in environmental noise. Since the effects of the age and background noise stopped short of incorporating to the RT prediction model in the current study, we just reported the findings on the changed RT by the different ages and background noise to improve the model in the further studies. Thus, the aims of this study were (1) to predict the RT for birdsongs in the silent condition and (2) to evaluate differences in RT due to the effects of participant age (or hearing ability) and background noise.

## 2. Methods

### *2.1. RT measurement in the silent condition*

### 2.1.1. Participants

We recruited 20 younger (15 men and 5 women; age: 21.7 ± 0.1 years) and 20 older (8 men and 12 women; age: 74.3 ± 4.1 years) participants to complete an RT task. Before the RT measurement, all participants completed hearing tests conducted using an audiometer (AA-77A, RION, Kokubunju, Japan). Fig. 1 compares audiograms obtained by the younger and older participants. The mean hearing threshold levels (MHTL: averaged hearing threshold for 500 Hz, 1 kHz, and 2 kHz pure tones) were 2.75 ± 2.81 dB HL and 20.92 ± 8.09 dB HL for the younger and older groups, respectively. The younger participants all had normal hearing. The older participants were classified as having slight or mild hearing loss [19], but none used a hearing aid.

### 2.1.2. Birdsongs

Six instances of natural birdsong (Jay, Cuckoo, Himalayan cuckoo, Japanese grosbeak, Japanese white-eye, and Japanese bush warbler) had previously been recorded using a dummy head microphone (KU100, Neumann, Berlin, Germany) with a sampling rate of 48 kHz and a sampling resolution of 24 bits [20-23]. Fig. 2 shows the waveform of each birdsong recording with a normalized SPL. The songs of the Cuckoo and Japanese bush warbler were amplified quickly and slowly over time, respectively. Fig. 3 shows the spectra of the birdsongs. The FFT size and window type were 4096 samples and Hanning, respectively, and all spectra calculated during each birdsong recording were averaged. The sound energy of the Jay’s song was distributed in a broad band, i.e., it was noisy, while the other birdsongs had particular spectral peaks (Cuckoo: 678 Hz, Himalayan cuckoo: 420 Hz, Japanese grosbeak: 3208 Hz, Japanese white-eye: 4425 kHz, and Japanese bush warbler: 2110 Hz).

To determine the output level of the birdsongs for our RT measurement, we measured each participant’s hearing threshold for the birdsong stimuli using a transformed up-down procedure [24]. We employed a two-alternative forced-choice (2AFC) procedure. The same birdsong stimuli were presented to the participant’s left and right ears via a calibrated headphone (HD598, Sennheiser, Wedemark, Germany), with the signal passing through an AD/DA converter (UA101, Roland, Hamamatsu, Japan). The original birdsong stimuli were generated using a laptop computer, and the output and participant responses were recorded using numerical analytical software (Matlab, Mathworks, Natick, US). The hearing threshold tests were conducted in a soundproof chamber. Finally, the output level was determined at 10, 20, and 30 dB SL, based on the hearing threshold of each participant. The birdsong at 10 dB SL was heard faintly, while the birdsong at 30 dB SL was heard clearly but was not perceived as loud. The standard deviations of SPL at 0dB SL were 5.41 dB and 13.84 dB for the younger and older participants.

### *2.1.3. Apparatus*

For the RT measurement, the headphone (binaural presentation), AD/DA converter, and computer were the same as those used in the previous hearing threshold test. The participants were instructed to tap on an empty carton using their fingers as soon as they detected the birdsong stimuli, as shown in

Fig. 4. They could identify the birdsongs because they had completed the previous hearing threshold test. A subminiature piezotronics accelerometer (352A21; PCB Piezotronics, New York, US) was attached to the reverse side of the tapping surface, and the vibration produced at the moment of tapping was registered by a pre-amplifier (Type 2690, Brüel & Kjær, Nærum, Denmark) that sent a signal to an input channel of the AD/DA converter. A looped output containing the birdsong was sent to the AD/DA converter while the tapping vibration was recorded simultaneously on the stereo channel. Then, the difference in timing (i.e., RT) between the onset of the birdsong stimuli and the initiation of tapping was calculated.

### *2.1.4. Procedure*

In each session, the six birdsong stimuli were presented four times in random order with the interval between two successive stimuli ranging from 3 to 12 sec. This prevented the participants from predicting the onset timing. After some training, the participants were able to detect all 24 birdsong stimuli in one session, at which point they were progressed to new sessions where the sound output levels were varied, i.e., 10, 20, and 30 dB SL. The order of these sessions was randomized. The RT values were averaged over four trials. When calculating the RT, birdsong stimuli that did not elicit a reaction were excluded. This experiment was conducted at the National Institute of Advanced Industrial Science and Technology (AIST), Japan, and the institutional ethics committee approved the experimental protocol.

### 2.1.5. Estimation of RT

When the RT can be evaluated by acoustical features of the birdsongs, the most detectable birdsong might be selected without such a psychoacoustical experiment. In this study, the RT was estimated on the basis of the $L_{\mathrm{AE}}(t)$ and $N(t)$. Fig. 5 shows the $L_{\mathrm{AE}}(t)$ of the six birdsongs with sound pressure adjusted to 0 dB SL. The SPL at 0 dB SL was equivalent to the average values in the younger participants. The birdsong onset time is indicated as 0 msec. The $L_{\mathrm{AE}}(t)$ was calculated using the following equation:

$$L_{AE}(t) = 10\log\int_0^t \frac{P_A^2(s)}{p_0^2}ds \qquad (1)$$

where $P_{\mathrm{A}}(\mathrm{s})$ is an A-weighted sound pressure [Pa] and $p_0$ is a minimum audible pressure ($2.0 \times 10^{-5}$ Pa). The songs of the Jay, Cuckoo, and Himalayan cuckoo increased the $L_{\mathrm{AE}}(t)$ with a steeper slope, while the songs of the Japanese grosbeak, Japanese white-eye, and Japanese bush warbler increased it more gradually.

For smoothing the curves of $L_{\mathrm{AE}}(t)$, fitting by a power function,

$$L_{AE}(t) \approx at^b + c, \qquad (2)$$

where $a$, $b$, and $c$ are coefficients, were applied as shown in Fig. 6. We hypothesized that the participants would detect the birdsong when the $L_{\mathrm{AE}}(t)$ reached a certain value. Fig. 6 demonstrates the values for the birdsongs at 10, 20, and 30 dB SL as gray dotted lines. The lines for 10 and 20 dB SL were 20 and 10 dB higher than the line for 30 dB SL, respectively. The moment at which the $L_{\mathrm{AE}}$

reached a certain value was defined as $t_{\mathrm{att}}$. Fig. 6 shows an example of the $t_{\mathrm{att}}$ for the Japanese bush warbler song. For example, if the certain $L_{\mathrm{AE}}$ for 10, 20, and 30 dB SL are given as 25, 15, and 5 dB, respectively (dotted lines in Fig. 6), the $L_{\mathrm{AE}}$ required for detection of the birdsongs becomes 35 dB, because the origins of these curves were adjusted to 0 dB SL and are shifted higher by 30 dB for the birdsongs in 30 dB SL. When the required $L_{\mathrm{AE}}$ is defined as $L_{\mathrm{att}}$, the $t_{\mathrm{att}}$ can be obtained from Eq. (2) and the logarithmical value was

$$\log t_{att} = \frac{1}{b}\log\frac{L_{att}-c}{a}. \qquad (3)$$

We calculated the correlation coefficients between the actual RTs obtained in the trials and the $t_{\mathrm{att}}$ by changing the $L_{\mathrm{att}}$, and suitable $L_{\mathrm{att}}$ values were determined based on the maximum correlation coefficient.

In contrast, the $N(t)$ was calculated using a computational program written in ISO standard [16]. Fig. 7 shows the N($t$) of the six birdsong stimuli with onset times at 0 sec. Colored lines indicate the three averaged output levels corresponding to 10, 20, and 30 dB SL in the younger participants. The temporal resolution of the $N(t)$ was 2 msec, and band-dependent smoothing filters and a nonlinear temporal decay algorithm were added to the Zwicker method for evaluating stationary sounds. In a similar manner to the $L_{\mathrm{AE}}(t)$, we hypothesized that the participants would detect the birdsong when the $N(t)$ reached a certain value, $N_{\mathrm{att}}$, as demonstrated in Fig. 7f. The time required to reach $N_{\mathrm{att}}$ was defined as $t_{\mathrm{att}}$, and suitable $N_{\mathrm{att}}$ values were determined based on the maximum correlation between $t_{\mathrm{att}}$ and RT, as with the $L_{\mathrm{AE}}(t)$.

### *2.2. RT measurement in the noisy condition*

In the noisy condition, the participants, apparatus, and procedure were the same as those in the silent condition. The six birdsong stimuli were the same, but the output level was set at 30 dB SL for all trials, which was the level at which the sounds could be heard clearly in the silent condition. We used a computer to combine the birdsong stimuli at 30 dB SL with white noise and 1/6-octave bandpass noise filtered with a 5th order Butterworth filter, and presented the resulting stimuli using headphones. The center frequencies of the bandpass noise stimuli were 1, 2, and 4 kHz, for a total of 4 different types of noise. Because environmental noises in public spaces include many different cases, the used noises were represented in broader one (white noise) and narrower one (bandpass noise) to examine the spectral masking effect that could happen. The output levels of the noise stimuli were controlled according to the signal-to-noise ratio (SNR). For the white noise, the SNRs were 0, -10, and -20 dB, and for the bandpass noise stimuli the SNRs were -20, -30, and -40 dB. Therefore, 12 noise stimuli (four kinds of noise × three SNRs) were used in total. The minimum SNRs (i.e., -20 dB for white noise and -40 dB for bandpass noise) were enough to hear slightly the birdsongs in these noises. Regardless of the presence or absence of the birdsong, the noise stimuli continued to sound in the background during the experiment.

As with the silent condition, each birdsong stimulus appeared four times in a session, and the RT was defined on the basis of the averaged value, excluding the undetected birdsong stimuli. The sessions were composed for 12 noises, and the order of sessions was randomized.

## 3. Results

### *3.1. RT in the silent condition*

Fig. 8a shows the averaged RT in the 20 younger participants. For clarity, the standard deviations are listed in Table A1 (APPENDIX A). The percentage of missing trials (absence of response) was 5.3%. Regardless of the type of birdsong, the RT increased as the SL decreased. In other words, the reaction times were shorter for more clearly audible birdsong stimuli. In terms of the differences between the types of birdsong, we separated them into three groups on the basis of RT. The songs of the Jay, Cuckoo, and Himalayan cuckoo had short RTs, i.e., less than 0.75 sec. The RTs for the Japanese grosbeak and Japanese white-eye were in the middle, and ranged between 0.75 sec and 0.83 sec. The RT for the Japanese bush warbler was the longest, i.e., greater than 0.83 sec. Because few distributions of RT were normal, the statistical differences were examined according to a non-parametric analysis (Friedman test). As the results, the birdsong type ($\chi^2_5 = 97.93$, $p < 0.01$) and SL ($\chi^2_2 = 50.01$, $p < 0.01$) had significant difference on the 20 younger participants’ RT.

Fig. 8b shows the averaged RT in the 20 older participants, and Table A1 gives the standard deviations. The percentage of missing trials was 4.7%. Although the RT data roughly agreed with that for the younger participants, the overall RT values were longer in the older participants. According to Friedman test, the birdsong type ($\chi^2_5 = 118.22$, $p < 0.01$) and SL ($\chi^2_2 = 39.69$, $p < 0.01$) had significant difference on the 20 older participants' RT, and the difference between the younger and older participants was also significant ($\chi^2_1 = 50.73$, $p < 0.01$).

When estimating the RT based on $L_{\mathrm{AE}}(t)$, we found the maximum correlation between the measured RT in Fig. 8a and the $t_{\mathrm{att}}$ in Fig. 6 when the $L_{\mathrm{att}}$ was 37 dB for the younger participants, although this relationship was logarithmical, as follows:

$$RT \approx 0.14 \log t_{att} + 1.05. \tag{4}$$

Because the $t_{\mathrm{att}}$ can be described by the coefficients of the power function for each birdsong in Eqs. (3) and (4) can be changed to

$$RT \approx \frac{0.14}{b} \log \frac{37-c}{a} + 1.05. \tag{5}$$

For the younger participants, the estimation accuracy was high, as shown in Fig. 9a ($r = 0.977$, averaged error = 0.0293 sec).

When estimating the RT based on $N(t)$, we found the maximum correlation between the measured RT in Fig. 8a and the $t_{\mathrm{att}}$ in Fig. 7 when the $N_{\mathrm{att}}$ was 0.08 sone for the younger participants. As with the $L_{\mathrm{AE}}(t)$, the measured RT and the $t_{\mathrm{att}}$ had a logarithmic relationship, so the RT can be expressed by

$$RT \approx 0.094 \log t_{att} + 0.945. \quad (6)$$

The estimation accuracy was also high, although not to the extent of that for $L_{\mathrm{AE}}(t)$, as shown in Fig. 9b ($r$ = 0.848, averaged error = 0.0726 sec).

We used the same procedures for estimating RT for the data from the older participants. Using the $L_{\mathrm{AE}}(t)$ method, we identified the origin of the $L_{\mathrm{AE}}(t)$ for each birdsong on the basis of the averaged hearing thresholds among the 20 older participants (i.e., 0 dB SL). The recalculated $L_{\mathrm{AE}}$ curves were fitted by power functions, and the highest correlation was found at $L_{\mathrm{att}}$ = 55.5 dB. Thus, Eq. (5) was changed to:

$$RT \approx \frac{0.15}{b} \log \frac{55.5 - c}{a} + 1.30. \quad (7)$$

However, the estimation accuracy was lower than that for the younger participants ($r$ = 0.724, averaged error = 0.128 sec). Using the $N(t)$ method, the instantaneous loudness $N(t)$ was recalculated according to the SLs of the older participants, and the highest correlation was found at $N_{\mathrm{att}}$ = 0.5 sone. Thus, the estimation formula was

$$RT \approx 0.14 \log t_{att} + 1.27, \quad (8)$$

and the estimation accuracy was similar to that for the $L_{\mathrm{AE}}(t)$ method ($r$ = 0.703, averaged error = 0.140 sec).

### *3.2. RT in the noisy condition*

Fig. 10 shows the averaged RT values in the noisy condition among the 20 younger and 20 older participants in the left and right columns, respectively. The standard deviations are listed in Table A1. To compare these data with those from the silent condition, we added the RTs for the stimuli at 30 dB SL to the figure. When white noise was presented simultaneously with the birdsong stimuli (Figs. 10a-1 and a-2), the change in RT according to the SNR was similar to the change in SLs in the silent condition (Fig. 8). The RTs remained short for the songs of the Jay, Cuckoo, and Himalayan cuckoo, and remained long for the song of the Japanese bush warbler. The RTs for the Japanese grosbeak and Japanese white-eye varied widely between these two categories. Unlike in the silent condition, Friedman test revealed no significant difference between the younger and older participants ($\chi^2_1 = 1.62$, $p = 0.20$). For the older participants, RTs in the white noise with a SNR of 0 dB were slightly faster than in the silent condition, except for in trials in which the Japanese white-eye birdsong was combined (Fig. 10a-2).

When bandpass noise was added, the spectral configuration of the noise strongly affected the RT. When the bandpass noise had a center frequency of 1 kHz, the RTs for the Cuckoo were longer for both younger and older participants (Figs. 10b-1 and b-2), even though the Cuckoo was in the short RT group in the silent and white noise conditions. This was likely because the 1-kHz bandpass noise (1/6 octave) masked much of the spectral energy of the Cuckoo (Fig. 3). The same increase in RT was observed for the Japanese bush warbler and Japanese white-eye songs when presented with bandpass noise with a center frequency of 2 kHz (Fig. 10c) and 4 kHz (Fig. 10d), respectively. The thick gray circles in Figs. 10c-2 and d-2 indicate that more than 10 participants failed to hear the birdsongs in all four trials. Thus, the plotted RTs in the circles represent only the values for participants with stronger hearing abilities, as they provided the averaged RT data.

Contrary to our expectations, some RTs were shorter when the bandpass noises were presented. For example, when bandpass noise with a center frequency of 1 or 2 kHz was presented to younger participants, the RTs for Japanese white-eye remained in the faster reacting category regardless of the high SNR (Figs. 10b-1 and c-1). Furthermore, the RTs of the older participants for the Cuckoo and Himalayan cuckoo decreased when bandpass noise with a center frequency of 2 or 4 kHz was added (Figs. 10c-2 and d-2).

## 4. Discussion

### *4.1. Kind of birdsong*

Both for the younger and older participants, the RTs for the Cuckoo were the fastest except when presented with 1-kHz bandpass noise. A previous study measured auditory evoked fields (AEFs) via magnetoencephalography while participants listened to several birdsong stimuli, including that of the Cuckoo [20,21]. The researchers found that the AEF amplitudes were larger for the Cuckoo than for other birds. In psychological experiments measuring subjective salience, the Cuckoo song was perceived as relatively salient when all birdsong stimuli were presented with an equal sound pressure level (SPL) and loudness [21,22]. Thus, our RT data are consistent with these previous physiological and psychological experiments, indicating that the Cuckoo’s song is relatively easy to detect. Furthermore, Japanese participants preferred the Cuckoo’s song in psychoacoustics experiments, making it a good choice for use as a public signal [23]. Following the Cuckoo, the RTs were relatively high for the songs of the Himalayan cuckoo and Jay, while that for the Japanese bush warbler was the longest. This is at odds with previous studies reporting that the song of the Japanese bush warbler is subjectively salient compared with other birdsongs [20,21].

Although we presented the birdsongs with a low SL for the purpose of observing the spectral pattern in RT [8,9], the initial sound intensity was a contributing factor in determining the length of the RT [10,11]. The RTs for the Japanese grosbeak and Japanese white-eye varied widely between the faster and slower RT groups according to the SL (Fig. 8), and the same pattern was observed for the cumulative sound intensity $L_{\mathrm{AE}}$ (Fig. 5). Therefore, the birdsongs with amplitudes that increased with a steeper slope had a faster RT.

### *4.2. Accuracy of estimated RT and the effect of age*

Examining the sound exposure level as a function of time $L_{\mathrm{AE}}(t)$ revealed that the RT shifted according to the SLs. Furthermore, the estimated RT values derived from the $L_{\mathrm{AE}}(t)$ model (Fig. 6) were in agreement with the measured RT, and the estimated and measured RT values were highly correlated in the younger participants (Fig. 9a). The $L_{\mathrm{AE}}$ can be used to evaluate the loudness of a sound with high accuracy [14], and is also able to explain the reaction times for birdsong stimuli. From Eq. (5), the younger participants detected the birdsongs with an $L_{\mathrm{AE}}$ that reached 37 dB in the silent condition. In contrast, the RTs for the older participants were longer than 132 msec, calculated using averaged difference among birdsongs and SL (Fig. 8b), and the required $L_{\mathrm{AE}}$ ($L_{\mathrm{att}}$) to detect birdsongs in this group was 55.5 dB (Eq. (7)). In previous research, hearing-impaired listeners with poorer hearing than the participants in this study had RTs that were 50–70 msec longer for loudness-controlled pure tones compared with normal listeners [6,25,26]. Thus, the difference in RT was greater for the case of time-variable birdsong. The difference in the $L_{\mathrm{att}}$ between the two groups in the present study was 18.5 dB. Because the difference in the MHTLs between the two groups was 18.17 dB, it is possible to speculate that the difference in the $L_{\mathrm{att}}$ was caused by varying hearing acuity.

The accuracy of the estimated RT based on the $L_{\mathrm{AE}}(t)$ model was substantially reduced for the older participants. The MHTL of the older participants (20.92 ± 8.09 dB HL) had a wider standard deviation than that of the younger participants (2.75 ± 2.81 dB HL), so it might not have been appropriate to calculate the SPL at 0 dB SL for the $L_{\mathrm{AE}}(t)$ curve according to the averaged values among the 20 older participants. When the SPL at 0 dB SL was adjusted on the basis of individual hearing thresholds and the $L_{\mathrm{AE}}(t)$ fitting curves were reshaped individually, the accuracies improved in 10 of the older participants ($r \geq 0.9$: 4 participants, $0.9 > r \geq 0.8$: 4 participants, $0.8 > r > 0.724$: 2 participants).

The instantaneous loudness $N(t)$ model [16] was inferior to the $L_{\mathrm{AE}}(t)$ model in terms of accuracy, and the RTs shorter than 0.75 sec were especially difficult to estimate (Fig. 9b). The faster RTs were obtained for birdsongs with a sudden onset, i.e., the Jay, Cuckoo, and Himalayan cuckoo. The Zwicker loudness model, which measures the origin source of current instantaneous loudness, is unsuitable for estimating the loudness of impulsive sound [27,28]. Thus, there were more estimation errors for faster RTs. The loudness required to detect the presence of birdsong ($N_{\mathrm{att}}$) was 0.08 sone for the younger participants in the silent condition. Because this corresponds to 3.56 phon when converted into a loudness level ($phon = 40 + 10 \log_2 Sone$), it is reasonable that they detected the birdsong at a time at which the SPL was slightly supra-threshold. In some previous RT experiments using pure tones with a wide range of SPL values, a stable reaction could be obtained with SPLs that were above the threshold by 3–5 dB [6,8,9]. In contrast, the required loudness level for the older participants was 30 phon (0.5 sone).

### *4.3. Noise effect*

In the noisy condition, the spectral masking of the 1/6 octave-band noise clearly affected the RT. For example, the bandpass noise with a center frequency of 1 kHz prolonged the RT for the Cuckoo stimulus, which had spectral power distributed around 1 kHz, as shown in Figs. 3 and 10b. The increment of RT in the SNR-40dB conditions from the silent (30 dB SL) was the largest for the Cuckoo (0.40 sec) in the younger participants. For the older participants, while the increments were larger for the Cuckoo (0.33 sec) and Japanese bush warbler (0.47 sec), and it can be expected that their auditory filters might be broader than the younger participants. In the same manner, bandpass noises with center frequencies of 2 and 4 kHz prolonged the RTs of the Japanese bush warbler and Japanese white-eye, respectively. In the younger participants, the increments of RT for the Japanese bush warbler (0.69 sec) and Japanese white-eye (0.65 sec) were the largest by adding the bandpass noises with center frequencies of 2 and 4 kHz, respectively. Various studies have examined RT with a wide range of noisy and tonal backgrounds [29-34], and Chocholle and Greenbaum (1966) [30] and Emmerich et al. (1976) [31] showed that the RTs for pure tones increased due to masking when the pure tones and the background noise were closed spectrally. Therefore, the onsets of the birdsongs in the present study were obscured by the bandpass noises that passed through the same auditory filter. The spectral masking effect was particularly strong for the birdsongs that were undetected by more than 10 of the older participants (gray thick circles in Figs. 10c-2 and d-2), because the hearing levels at 2 and 4 kHz were lower than that at 1 kHz (Fig. 1).

In contrast to the expanding RTs, adding the bandpass noise shortened some of the RTs. Although spectral masking occurred when the spectral energies of the signal and noise were overlapped, noise can speed a reaction to a signal when the spectral energies are separated [31]. Fig. 11 shows the differences between the RTs in the silent condition (30 dB SL) and noisy conditions. We only obtained positive values, i.e., the RT was faster in the noisy condition. Regardless of the addition of background noise, the younger and older participants were both able to detect the song of the

Himalayan cuckoo more quickly in the noisy conditions. It is well established that low-frequency sounds tend to mask high-frequency sounds. In this study, the song of the Himalayan cuckoo had the lowest spectral distribution, and the spectral peak was 420 Hz, as shown in Fig. 3. Thus, this stimulus was not masked, but instead became easier to detect. For the older participants, the same tendency could be observed for the Cuckoo song, which had a spectral peak at 678 Hz, when presented with the bandpass noises with center frequencies of 2 and 4 kHz (Figs. 11c and 11d, respectively). Because the older participants experienced this benefit, the background noise may have passed through an auditory filter in a band lower than that of the stimulus, making both sounds easier to detect. Such enhancement is close to "*stochastic resonance*", in which signal transmission is boosted in a nonlinear system by adding optimal noise [34-39]. A previous study reported that the stochastic procedure could be used to explain RTs for sounds with a variety of sound intensities and durations [40]. The other possibility is that the cognitive performance (recognition of birdsongs in this study) was improved in an "*inverted-U shaped manner*" when an individual is under optimal stress [41-43]. The RT for visual cognitive tasks was shorter by adding controlled auditory signals [44-46].

However, such spectral nonlinearities were not observed when we added white noise (Fig. 10a), and the increases in RT with decreasing SNR paralleled the increase in RT as the SL decreased in the silent condition (Fig. 8). Unlike the silent condition, the RTs in the white noise condition were not significantly different between the younger and older participants. The older participants had faster RTs when we added white noise (Fig. 11a), and they are possible that the stochastic resonance elevated the intensity enough to detect it or that the optimal stress elevated the arousal of the participants following the inverted-U shape function.

## 5. Conclusion

In the silent condition with variable SLs, we observed RTs shorter than 0.75 sec for both the younger and older participants for the birdsongs of the Jay, Cuckoo, and Himalayan cuckoo, which have steeper slopes for increasing amplitudes. However, these birdsongs are different according to the individual bodies in the same species, so it is important to extract the acoustical characteristics which are dependent on the RT. In terms of accuracy for estimating RT, the $L_{\mathrm{AE}}(t)$ model was superior to the $N(t)$ model. The $L_{\mathrm{AE}}$(t) can be calculated using any birdsong with any SL, and the duration required to reach 37 dB on the fitted power function of the $L_{\mathrm{AE}}$(t) can be used to estimate the RT in younger participants, i.e., those with normal hearing, with high accuracy ($r$ = 0.977).

In the noisy condition, when the SNR was increased, spectral masking produced by narrow-band noise lengthened the RT and made the birdsongs more difficult to detect when the birdsong and noise were passing through the same auditory filter. In contrast, the RT slightly decreased when the spectral energies of the birdsong and noise did not overlap. Furthermore, white noise shortened the RT, especially in older participants. In this study, the RT estimation model including the effects of age and background noise was not accomplished. Especially, the nonlinearity of the noise effect (i.e., fasten

RT) disturbed highly accurate prediction. In the further study, the noise effect on RT will be examined in more detail using several bandpass noises with different bandwidth.

## CRediT authorship contribution statement

**Ryota Shimokura**: Conceptualization, Methodology, Software, Investigation, Analysis, Writing -original draft. **Yoshiharu Soeta**: Implementation of psychoacoustical test, Supervision.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the reported in this paper.

## Acknowledgments

The authors would like to thank the participants for their cooperation during the experiments and the Human Resources Centers for the Aged at Ikeda (Osaka). In addition, the authors would like to thank Ms. Aya Kanzaki and Ms. Ayaka Ariki, who helped to implement the psychoacoustical experiment. This work was supported by a Grant-in-Aid for Scientific Research (B) from the Japan Society for the Promotion of Science (115H02771). We thank Sydney Koke, MFA, from Edanz (https://www.edanz.com/ac) for editing a draft of this manuscript.

## APPENDIX A

Table A1 lists the standard deviations of the RTs among the younger and older participants. The averaged standard deviations in each condition were 0.05–0.07 sec higher in the older participant group than in the younger group. The standard deviations were relatively high for both the younger and older participants, likely because the RTs depended not only on perceptual responses but also on individual physical and cognitive responses [8,32,47]. Such individual differences may be easier to detect for time-variable sounds (birdsongs) compared with steady-state sounds (pure tones). Therefore, the changes in the averaged RTs shown in Figs. 8 and 10 capture common tendencies observed in all of the participants.

## References

[1] Suied C, Susini P, McAdams S, Patterson RD. Why are natural sounds detected faster than pips? J Acoust Soc Am. 2010;127EL:105-10.

[2] Ministry of Land, Infrastructure, Transport and Tourism. Guidelines for Facilitating the Move in Public Transportation System. 2007 (in Japanese).

[3] Wundt W. Grundzüge der Physiologischen Psychologie. 2nd ed. Engelman: Leipzig; 1874. German.

[4] Cattel J. The influence of the intensity of the stimulus on the length of reaction time. Brain. 1886;8:512-5.

[5] Piéron H. Nouvelles recherches sur l'analyse du temps de latence sensorielle en function des intensités excitatrices. Annee Psychol. 1920;22:58-142. French.

[6] Chocholle R. Variation des temps de réaction auditif en function de l'intensité à diverses frequencies. Annee Psychol. 1940;41:65-124. French.

[7] Marshall L, Brandt J. The relationship between loudness and reaction time in normal hearing listeners. Acta Otoloryngol. 1980;90:244-9.

[8] Kohfeld DL, Santee JL, Wallace ND. Loudness and reaction time: I. Percept. Psychophys. 1981;29:535-49.

[9] Epstein M, Florentine M. Reaction time to 1- and 4-kHz tones as a function of level. Ear and Hearing 2006:27;424-9.

[10] Warm JS, Foulke E. Effects of rate of signal rise and decay on reaction time to the onset and offset of acoustic stimuli. Percept Psychophys. 1970;7:159-60.

[11] Schlittenlacher J, Ellermeier W. Simple reaction time to the onset of time-varying sounds. Atten Percept Psychophys. 2015;77:2424-37.

[12] Florentine M, Buus S, Poulsen T. Temporal integration of loudness as a function of level. J Acoust Soc Am. 1996;99:1633-44.

[13] Buus S, Florentine M, Paulsen T. Temporal integration of loudness, loudness discrimination and the form of the loudness function. J Acoust Soc Am. 1997;101:669-80.

[14] Numba S, Kuwano S, Fastl H. Loudness of non-steady-state sound. Jpn Psychol Res. 2008;50:154-66.

[15] ISO 1996-1:2016, Acoustics -Description, measurement and assessment of environmental noise-Part 1: Basic quantities and assessment procedures, Geneva: International Organization for Standardization, 2016.

[16] ISO 532-1, Acoustics -Methods for calculating loudness- Part 1: Zwicker method, Geneva: International Organization for Standardization, 2017.

[17] Schlittenlacher J, Hashimoto T, Kuwano S, Namba S. Overall loudness of short time-varying sounds, Proceedings of Internoise; 2014 Nov 16-19; Melbourne, Australia, p. 1-4.

[18] Fiebig A, Sottek R. Contribution of peak events to overall loudness. Acta Acust United Acust. 2015;101:1116-29.

[19] Clark JG. Uses and abuses of hearing loss classification. American Speech-Language-Hearing Association. 1981;23:483-500.

[20] Soeta Y, Nakagawa S. Prediction of optimal auditory signals using auditory evoked magnetic responses. Building and Environment. 2015;94:924-9.

[21] Yano H, Soeta Y. Induced brain magnetic activities related to salient birdsong under noisy conditions. NeuroReport. 2021;32:82-7.

[22] Soeta Y, Ariki A. Subjective salience of birdsong and insect song with equal sound pressure level and loudness. Int J Environ Res Public Health. 2020;17:1-12.

[23] Soeta Y, Kagawa H. Subjective preference for birdsong and insect song in equal sound pressure level. Appl Sci. 2020;10:849-61.

[24] Levitt H. Transformed up-down methods in psychoacoustics. J Acoust Soc Am. 1971, 49:467-77.

[25] Seitz PF, Rakerd B. Auditory stimulus intensity and reaction time in listeners with longstanding sensorineural hearing loss. Ear and Hearing 1996;18:502-12.

[26] Florentine M, Buus S, Rosenberg M. Reaction-time data support the existence of softness imperception in cochlear hearing loss. In Auditory Signal Processing: Physiology, Psychoacoustics, and Models. Edited by Pressnitzer D, de Cheveigné A, McAdams S, Collet L. New York: Springer-Verlag; 2004.

[27] Rennies J, Verhey JL, Appell JE, Kollmeier B. Loudness of complex time-varying sounds? A challenge for current loudness models, Proceedings of Meetings on Acoustics. 2013;19:050189.

[28] Sottek R, Moll T. Loudness perception and modeling of impulsive sounds. Proceedings of Euro Noise; 2015 May 31-June 3; Maastricht, Netherlands; p. 1919-24.

[29] Raab DH, Grossberg M. Reaction time to changes in the intensity of white noise. J Exp Psychol. 1965;69:609-12.

[30] Chocholle R, Greenbaum H. La sonie de sons purs partiellement masqués. Étude comparatif par une méthode d' égalisation et par la méthode des temps de reaction. Journal de Psychologie Normal et Pathologique 1966;63:385-414. French.

[31] Emmerich DS, Pitchford LJ, Becker CA. Reaction time to tones in tonal backgrounds and a comparison of reaction time to signal onset and offset. Percept Psychophys. 1976:20;210-14.

[32] Kemp S, Irwin RJ. Reaction time to the start and end of weak signals in noise. Psychol Res. 1979;40:367-76.

[33] Kemp S. Reaction time to a tone in noise as a function of the signal-to-noise ratio and tone level. Percept Psychophys. 1984;36:473-6.

[34] Jaramillo F, Wiesenfeld K. Mechanoelectrical transduction assisted by Brownian motion: a role for noise in the auditory system. Nat Neurosci. 1998;1:384–8.

[35] Henry KR. Noise improves transfer of near-threshold, phase-locked activity of the cochlear nerve: evidence for stochastic resonance? J Comp Physiol A. 1999;184:577–84.

[36] Zeng FG, Fu QJ, Morse R. Human hearing enhanced by noise. Brain Res. 2000;869:251–5.

[37] Moss F, Ward LM, Sannita WG. Stochastic resonance and sensory information processing: a tutorial and review of application. Clin Neurophysiol. 2004;115:267–81.

[38] Ries DT. The influence of noise type and level upon stochastic resonance in human audition. Hear Res. 2007;228:136–43.

[39] Ward LM, MacLean SE, Kirschner A. Stochastic Resonance Modulates Neural Synchronization within and between Cortical Sources. PLoS ONE. 2010;5:e14371.

[40] Miller J, Ulrich R. Simple reaction time and statistical facilitation: A parallel grains model. Cogn Psychol. 2003;46:101-51.

[41] Yerkers RM, Dodson JD. The relation of strength of stimulus to rapidity of habit-formation. J Comp Neurol Psychol. 1908;18:459-482.

[42] Broadbent DE. A reformulation of the Yerkes-Dodson low. Br J Moth Stat Psychol. 1965;18:145-157.

[43] Mendl M. Performing under pressure: Stress and cognitive function. Appl Anim Behav Sci. 1999;65:221-244.

[44] Hershenson M. Reaction time as a measure of intersensory facilitation. J Exp Psychol. 1962;68:289-293.

[45] Berstein IH, Clark MH, Edelstein BA. Effects of an auditory signal on visual reaction time. J Exp Psychol. 1969;80:567-569.

[46] Herweg NA, Bunzeck N. Differential effects of white noise in cognitive and perceptual tasks. Frontiers in Psychology. 2015;6:1-16.

[47] Luce RD, Green DM. A neural timing theory for response times and psychophysics of intensity. Psychol Rev. 1972;79:14-57.

## FIGURE CAPTIONS

**Fig. 1.** Audiogram of 20 younger and 20 older participants (error bar: standard deviation).

**Fig. 2.** Waveforms of six birdsongs.

**Fig. 3.** Spectra of six birdsongs.

**Fig. 4.** Experimental conditions.

**Fig. 5.** Time-functional sound exposure levels ($L_{\mathrm{AE}}(t)$) of six birdsongs.

**Fig. 6.** Curves fitted by a power function for the calculation of sound exposure level and definitions of $t_{\mathrm{att}}$.

**Fig. 7.** Time varying loudness ($N(t)$) of six birdsongs and definitions of $N_{\mathrm{att}}$ and $t_{\mathrm{att}}$.

**Fig. 8.** Reaction times for six birdsongs in (a) younger and (b) older participants.

**Fig. 9.** Relationships between measured RT and estimated RT according to the (a) $L_{\mathrm{AE}}(t)$ model and the (b) $N(t)$ model.

**Fig. 10.** Reaction time in the noisy conditions: (a-1) younger and (a-2) older participants in the white noise condition, (b-1) younger and (b-2) older participants in the bandpass noise condition with a center frequency of 1 kHz, (c-1) younger and (c-2) older participants in the bandpass noise condition with a center frequency of 2 kHz, and (d-1) younger and (d-2) older participants in the bandpass noise condition with a center frequency of 4 kHz. Gray thick circles indicate that more than 10 participants failed to detect the birdsongs in all four trials, and the numbers by the circles indicate the number of participants who heard and reacted to the birdsongs.

**Fig. 11.** Differences in RTs under the silent condition (30 dBSL) and the noisy conditions: (a) white noise and bandpass noises with center frequencies of (b) 1 kHz, (c) 2 kHz, and (d) 4 kHz. The values were selected only in the cases of faster RTs in the noisy conditions.

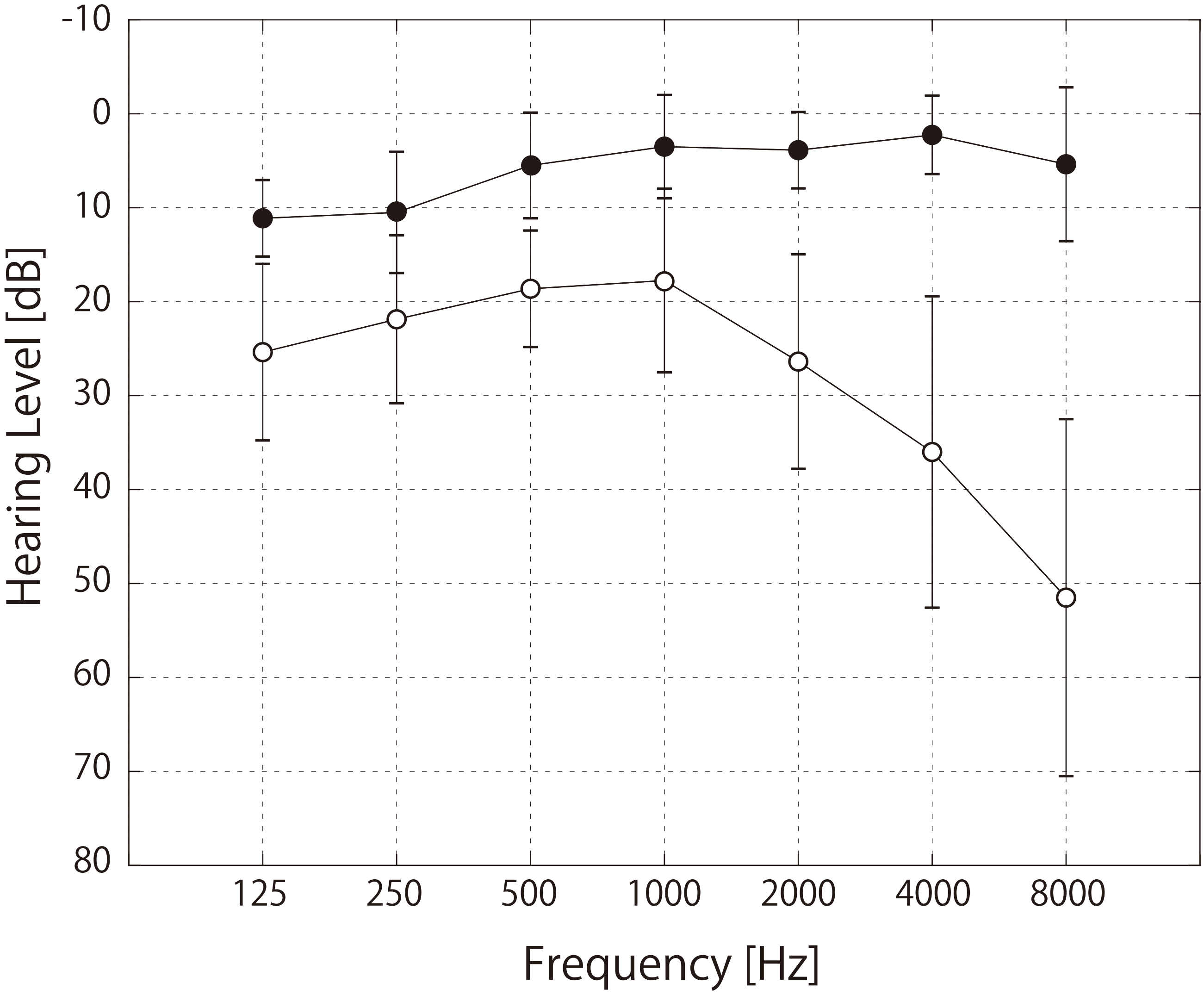

Figure 1

Shimokura and Soeta

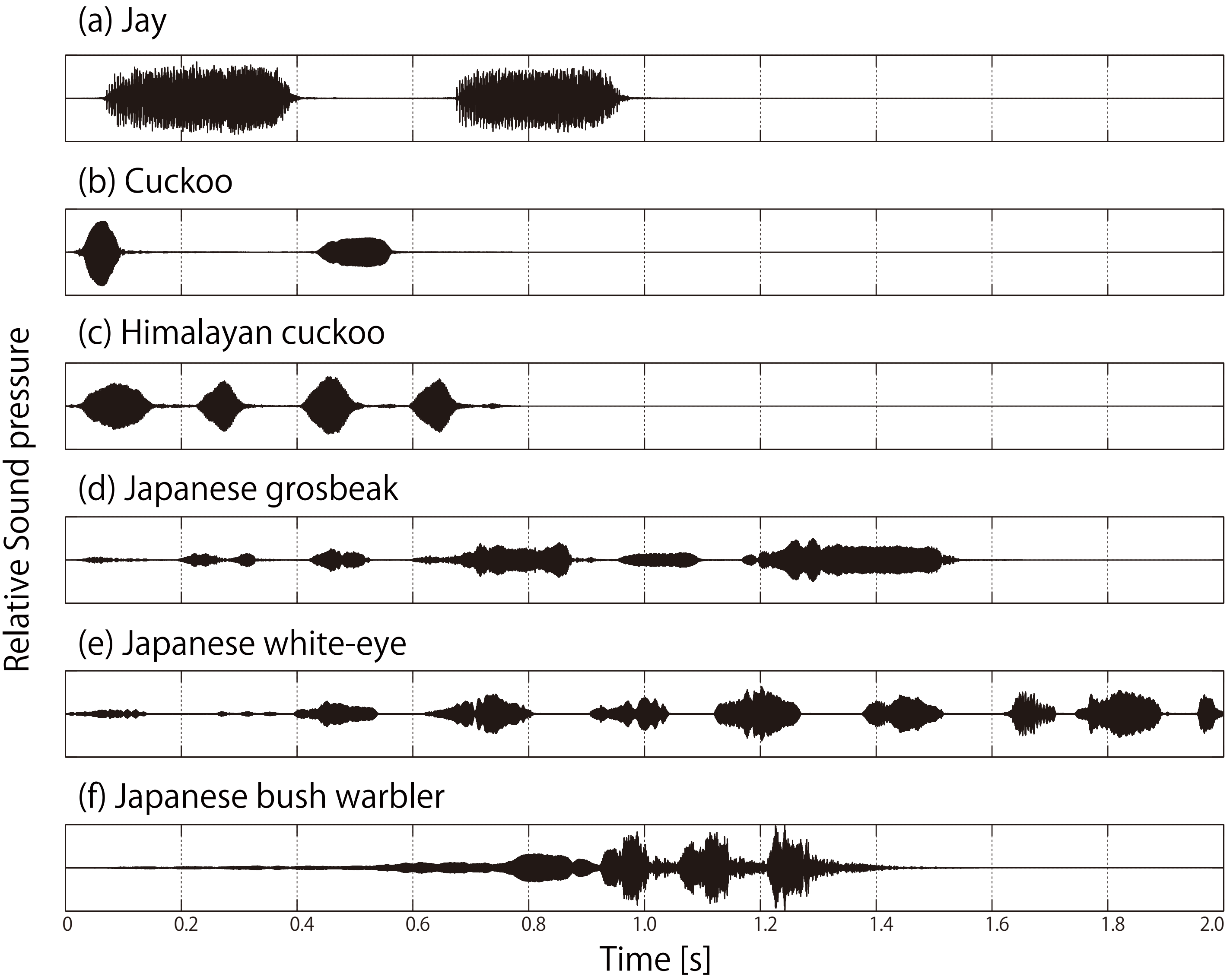

Figure 2

Shimokura and Soeta

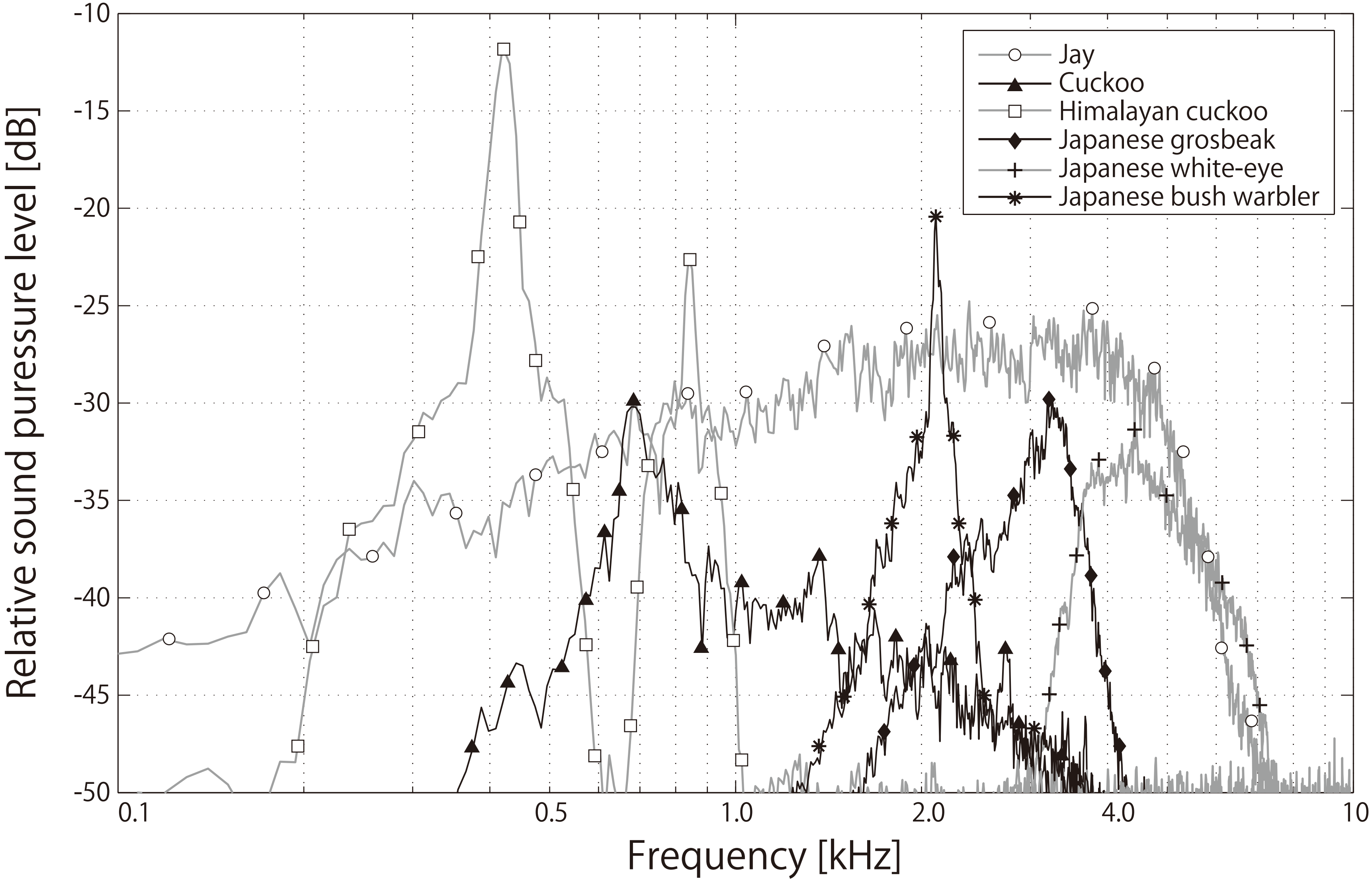

Figure 3

Shimokura and Soeta

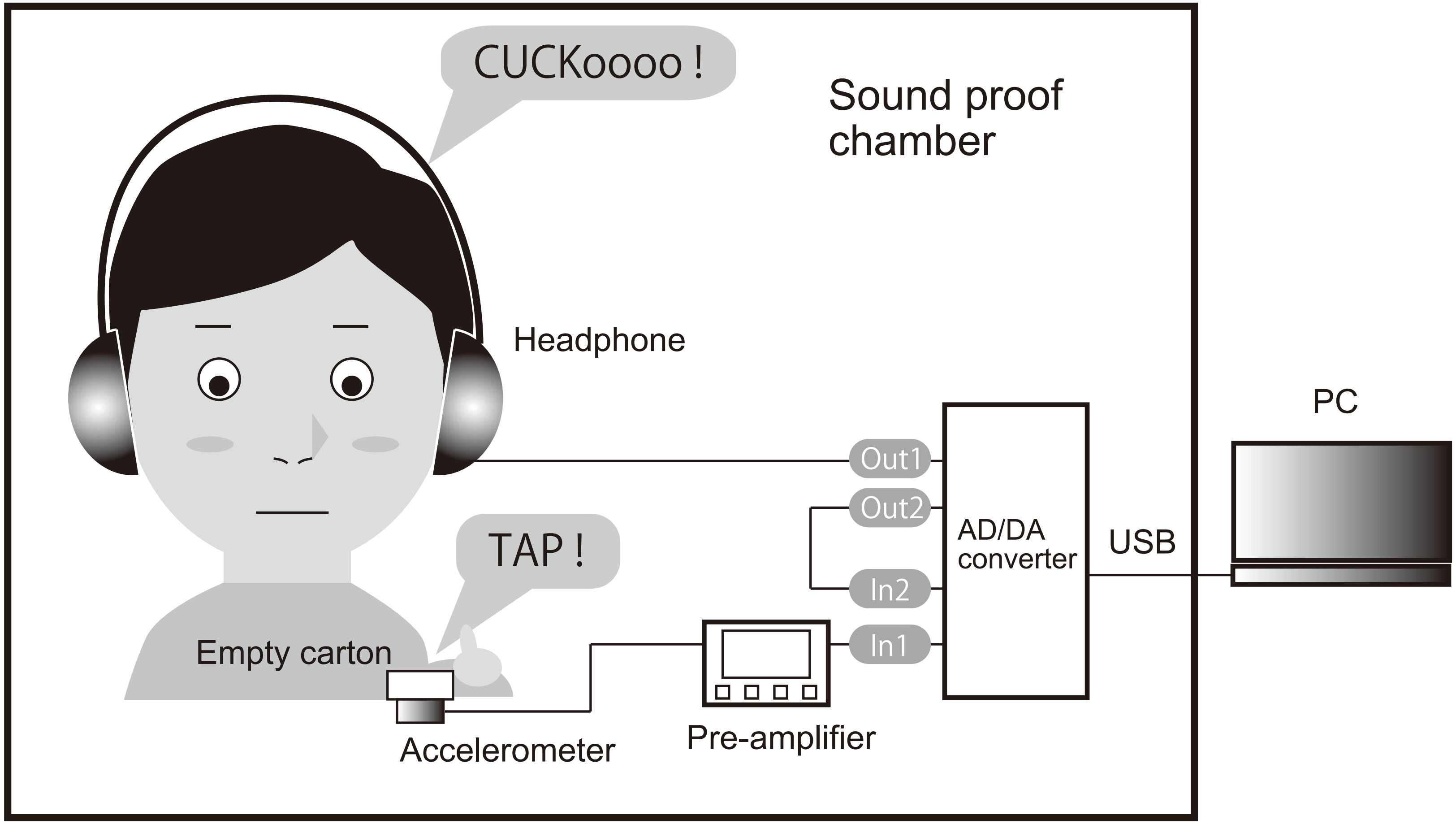

Figure 4

Shimokura and Soeta

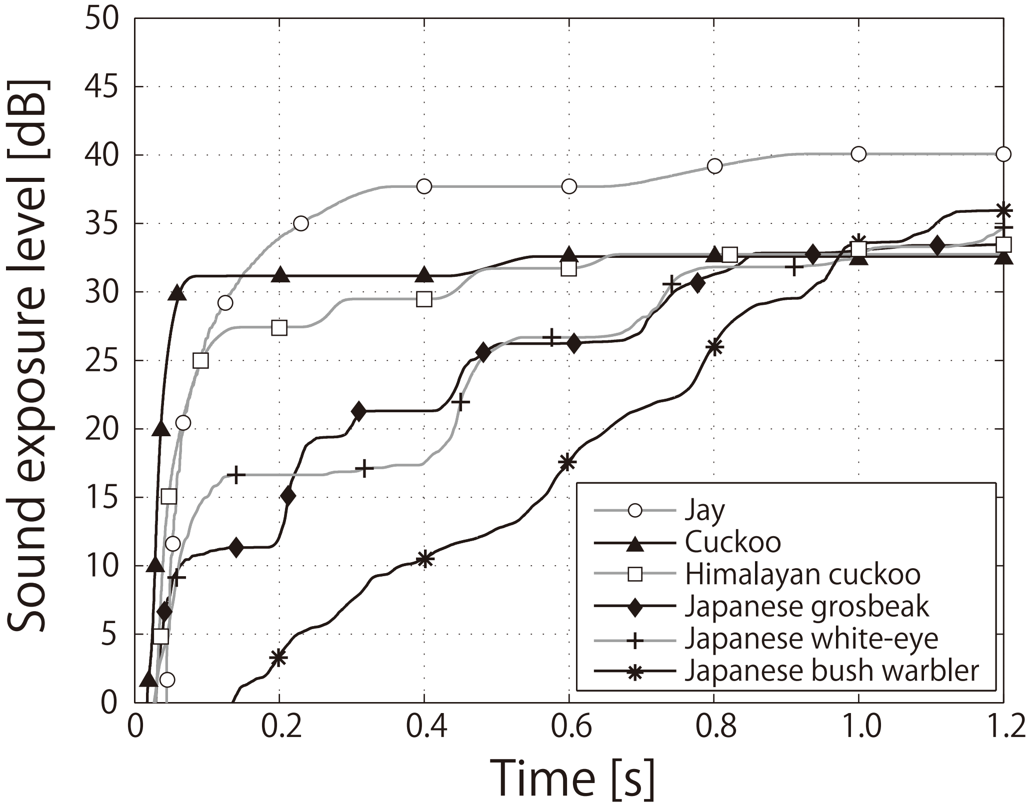

Figure 5

Shimokura and Soeta

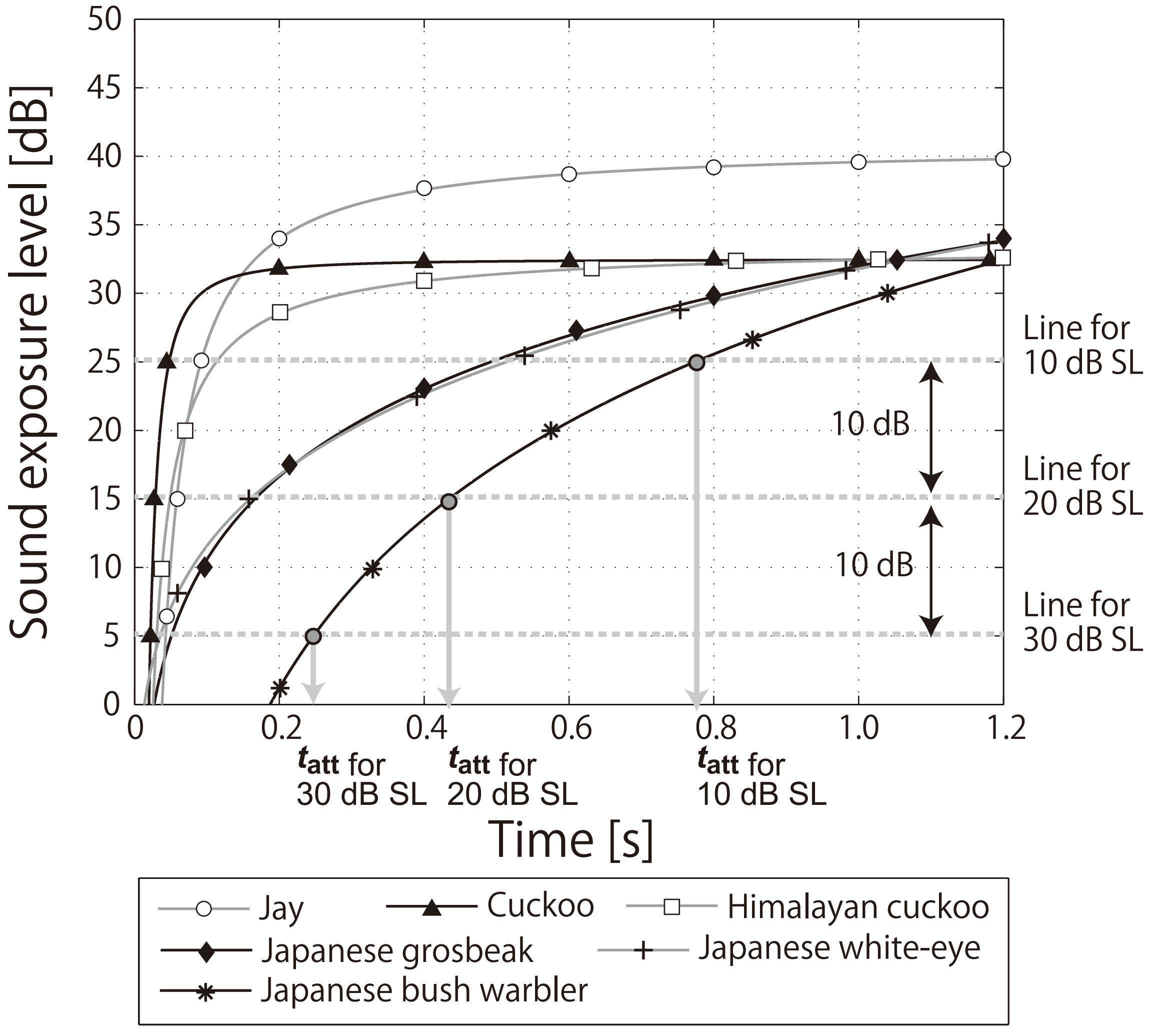

Figure 6

Shimokura and Soeta

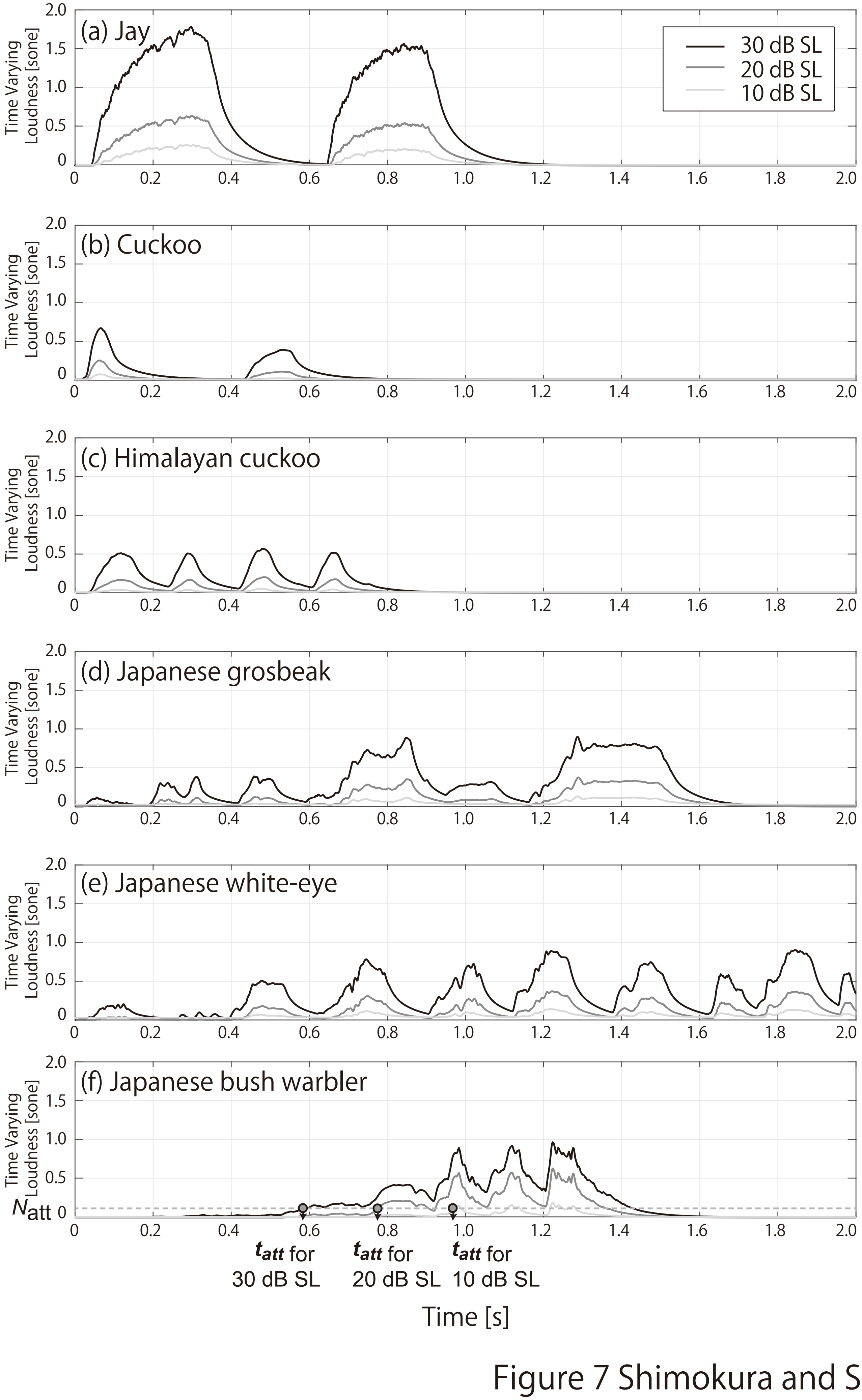

Figure 7 Shimokura and Soeta

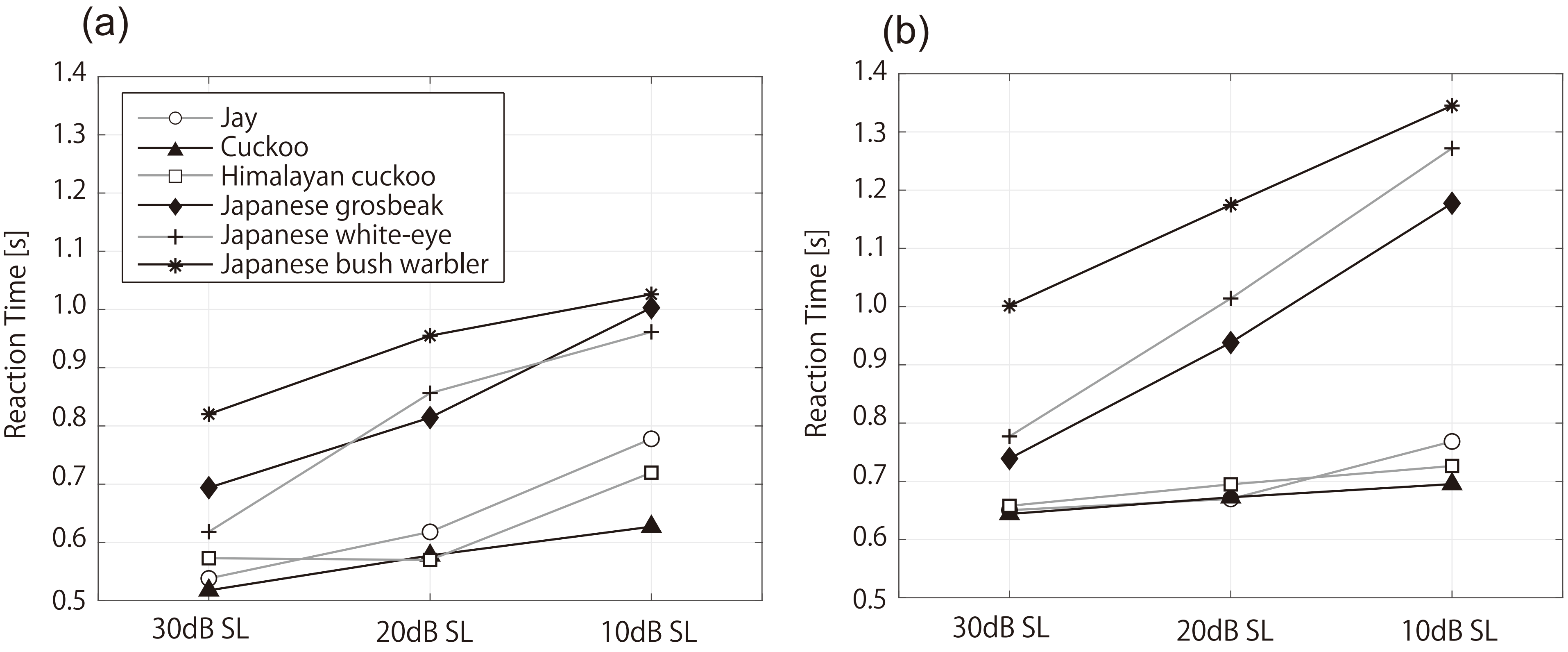

Figure 8

Shimokura and Soeta

Figure09

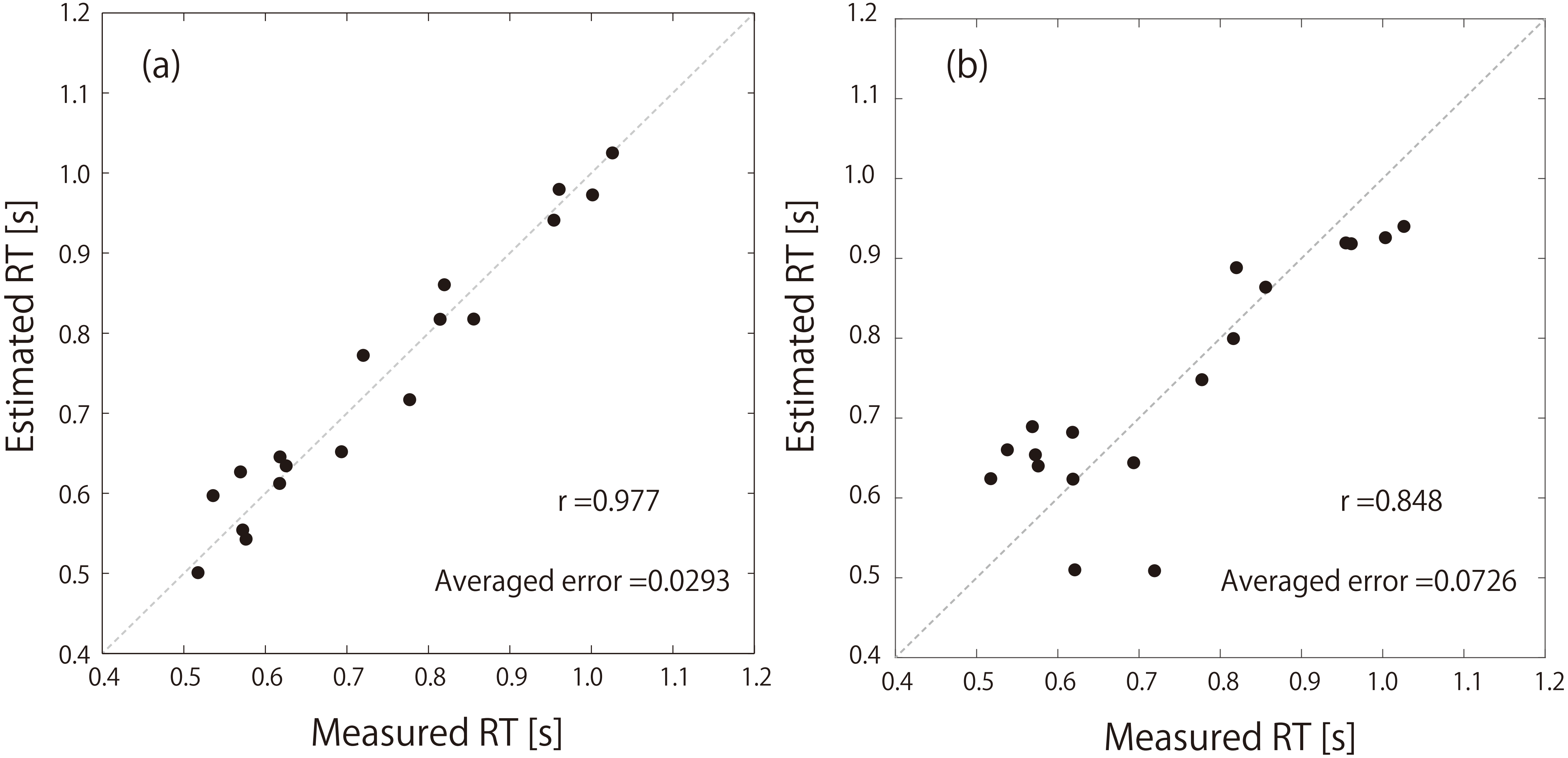

Figure 9

Shimokura and Soeta

Figure10

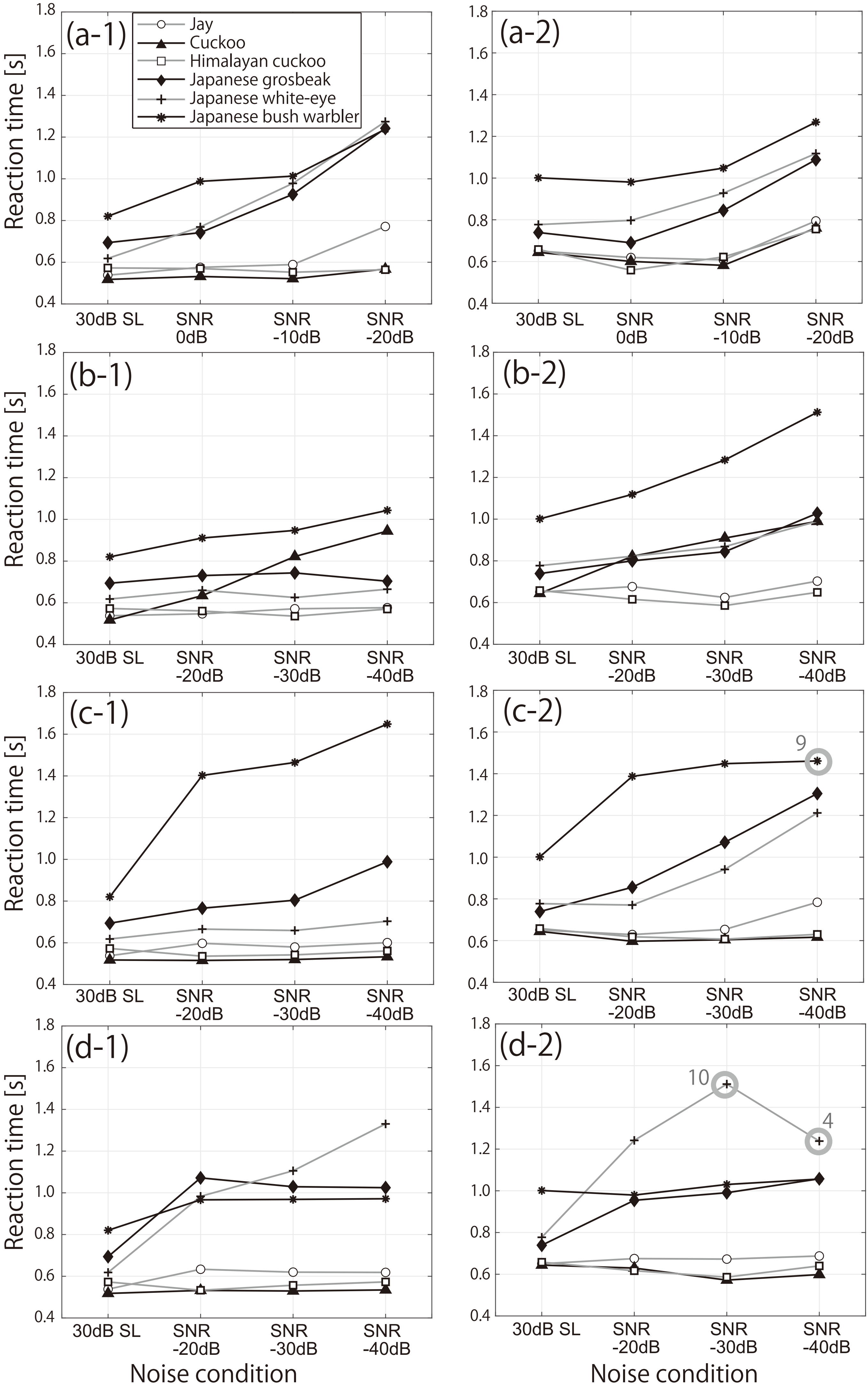

Figure 10 Shimokura and Soeta

Figure11

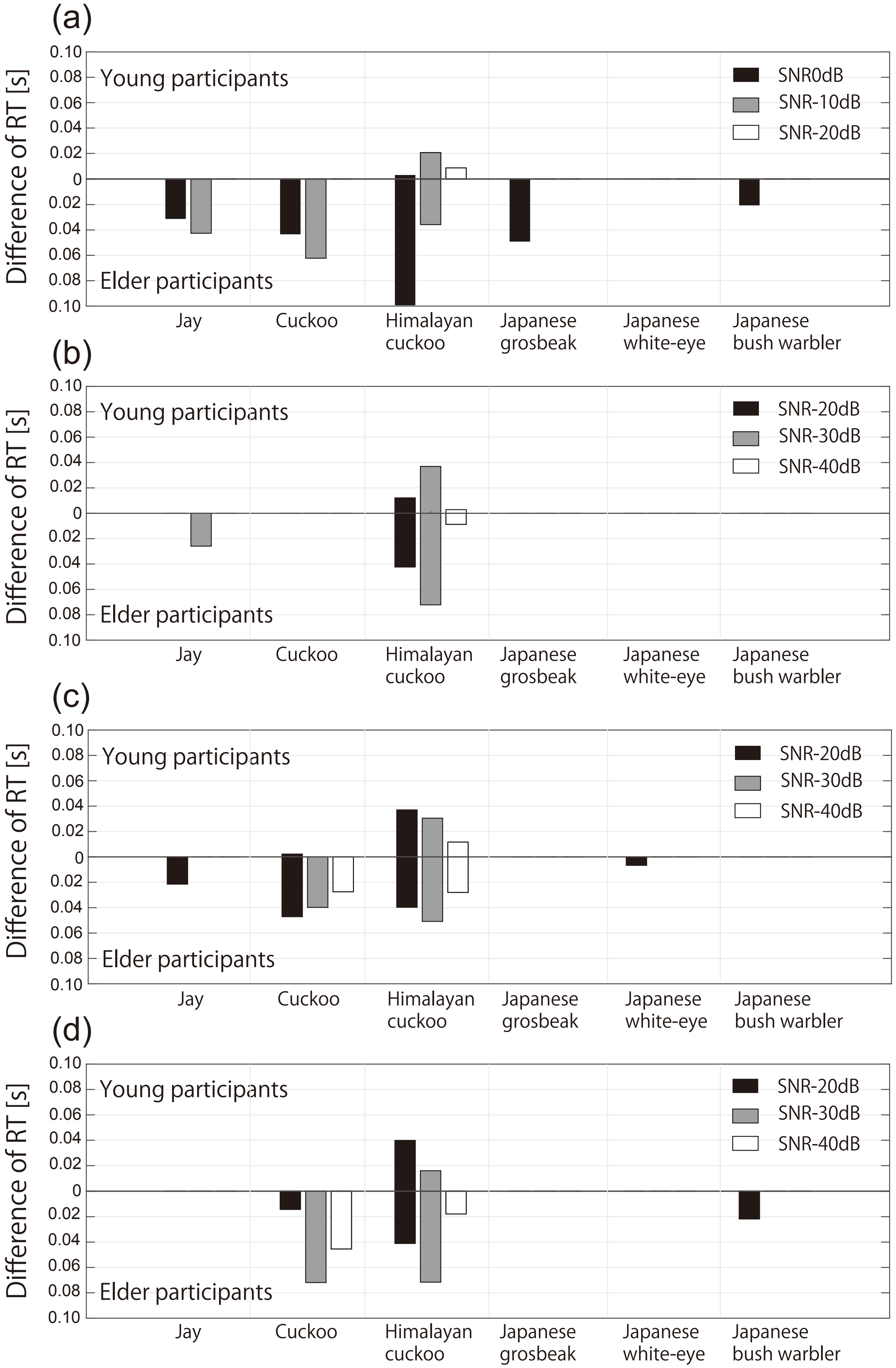

Figure 11 Shimokura and Soeta

Table A1 Standard deviation of RT in each condition

| | Younger participants | | | | | | | | Older participants | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Silent | | White noise | | | Bandpass noise | | | Silent | | White noise | | | Bandpass noise | | |
| | | | | | | 1 kHz | 2 kHz | 4 kHz | | | | | | 1 kHz | 2 kHz | 4 kHz |
| Jay | 30 dB SL | 0.19 | SNR0 dB | 0.19 | SNR-20 dB | 0.15 | 0.23 | 0.23 | 30 dB SL | 0.27 | SNR0 dB | 0.26 | SNR-20 dB | 0.29 | 0.26 | 0.21 |
| | 20 dB SL | 0.22 | SNR-10 dB | 0.18 | SNR-30 dB | 0.21 | 0.21 | 0.21 | 20 dB SL | 0.23 | SNR-10 dB | 0.25 | SNR-30 dB | 0.22 | 0.26 | 0.20 |
| | 10 dB SL | 0.24 | SNR-20 dB | 0.26 | SNR-40 dB | 0.19 | 0.17 | 0.20 | 10 dB SL | 0.23 | SNR-20 dB | 0.24 | SNR-40 dB | 0.29 | 0.27 | 0.32 |
| Cuckoo | 30 dB SL | 0.13 | SNR0 dB | 0.17 | SNR-20 dB | 0.18 | 0.17 | 0.14 | 30 dB SL | 0.22 | SNR0 dB | 0.23 | SNR-20 dB | 0.21 | 0.20 | 0.25 |
| | 20 dB SL | 0.15 | SNR-10 dB | 0.14 | SNR-30 dB | 0.26 | 0.13 | 0.15 | 20 dB SL | 0.21 | SNR-10 dB | 0.20 | SNR-30 dB | 0.29 | 0.22 | 0.18 |
| | 10 dB SL | 0.12 | SNR-20 dB | 0.14 | SNR-40 dB | 0.16 | 0.13 | 0.15 | 10 dB SL | 0.21 | SNR-20 dB | 0.27 | SNR-40 dB | 0.22 | 0.20 | 0.20 |
| Himalayan cuckoo | 30 dB SL | 0.15 | SNR0 dB | 0.18 | SNR-20 dB | 0.13 | 0.17 | 0.14 | 30 dB SL | 0.22 | SNR0 dB | 0.16 | SNR-20 dB | 0.19 | 0.21 | 0.23 |
| | 20 dB SL | 0.11 | SNR-10 dB | 0.15 | SNR-30 dB | 0.15 | 0.14 | 0.14 | 20 dB SL | 0.22 | SNR-10 dB | 0.17 | SNR-30 dB | 0.18 | 0.20 | 0.17 |
| | 10 dB SL | 0.16 | SNR-20 dB | 0.14 | SNR-40 dB | 0.15 | 0.15 | 0.15 | 10 dB SL | 0.21 | SNR-20 dB | 0.24 | SNR-40 dB | 0.20 | 0.19 | 0.27 |
| Japanese grosbeak | 30 dB SL | 0.24 | SNR0 dB | 0.20 | SNR-20 dB | 0.25 | 0.36 | 0.23 | 30 dB SL | 0.28 | SNR0 dB | 0.26 | SNR-20 dB | 0.30 | 0.33 | 0.28 |
| | 20 dB SL | 0.23 | SNR-10 dB | 0.23 | SNR-30 dB | 0.25 | 0.31 | 0.23 | 20 dB SL | 0.33 | SNR-10 dB | 0.33 | SNR-30 dB | 0.30 | 0.43 | 0.24 |
| | 10 dB SL | 0.35 | SNR-20 dB | 0.29 | SNR-40 dB | 0.25 | 0.32 | 0.21 | 10 dB SL | 0.32 | SNR-20 dB | 0.33 | SNR-40 dB | 0.34 | 0.44 | 0.31 |
| Japanese white-eye | 30 dB SL | 0.23 | SNR0 dB | 0.22 | SNR-20 dB | 0.21 | 0.28 | 0.28 | 30 dB SL | 0.32 | SNR0 dB | 0.30 | SNR-20 dB | 0.31 | 0.27 | 0.37 |
| | 20 dB SL | 0.22 | SNR-10 dB | 0.22 | SNR-30 dB | 0.21 | 0.25 | 0.37 | 20 dB SL | 0.35 | SNR-10 dB | 0.35 | SNR-30 dB | 0.35 | 0.38 | 0.30 |
| | 10 dB SL | 0.28 | SNR-20 dB | 0.32 | SNR-40 dB | 0.25 | 0.21 | 0.30 | 10 dB SL | 0.31 | SNR-20 dB | 0.40 | SNR-40 dB | 0.36 | 0.35 | 0.28 |
| Japanese bush warbler | 30 dB SL | 0.25 | SNR0 dB | 0.17 | SNR-20 dB | 0.22 | 0.18 | 0.21 | 30 dB SL | 0.30 | SNR0 dB | 0.29 | SNR-20 dB | 0.30 | 0.24 | 0.31 |
| | 20 dB SL | 0.26 | SNR-10 dB | 0.15 | SNR-30 dB | 0.23 | 0.16 | 0.23 | 20 dB SL | 0.29 | SNR-10 dB | 0.26 | SNR-30 dB | 0.28 | 0.26 | 0.30 |
| | 10 dB SL | 0.34 | SNR-20 dB | 0.28 | SNR-40 dB | 0.21 | 0.23 | 0.23 | 10 dB SL | 0.24 | SNR-20 dB | 0.24 | SNR-40 dB | 0.29 | 0.36 | 0.30 |