

Article

Not peer-reviewed version

Reaction Time to Amplitude-Modulated Tones Under Spectral Masking: Implications for Architectural Acoustic Design

[Ryota Shimokura](#) and [Yoshiharu Soeta](#) *

Posted Date: 17 March 2026

doi: 10.20944/preprints202603.1379.v1

Keywords: pure tone; envelope; reaction time; sound-exposure level



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Reaction Time to Amplitude-Modulated Tones Under Spectral Masking: Implications for Architectural Acoustic Design

Ryota Shimokura ¹  and Yoshiharu Soeta ^{2,*}

¹ Department of Systems Science, Graduate School of Engineering Science, Osaka University, Toyonaka 560-8531, Japan

² Molecular Biosystems Research Institute Life Science and Biotechnology, National Institute of Advanced Industrial Science and Technology (AIST), Ikeda, 563-8577, Japan

* Correspondence: y.soeta@aist.go.jp; Tel.: +81-50-3522-7155

Abstract

Detectability of auditory signals in built environments is a critical issue in architectural acoustics, particularly in public spaces where notification sounds must be perceived reliably under background noise. This study investigated reaction times (RTs) to amplitude-modulated pure tones under silent, white noise, and bandpass-noise conditions. Twenty young and twenty elderly participants responded to 1- and 2-kHz tones with flat, gentle, and steep onset envelopes. To describe perceptual detection in physically interpretable terms, a time-integrated sound-exposure level model, $L_{AE}(t)$, was applied. RT was defined as the moment when cumulative acoustic energy exceeded a criterion value relative to the hearing threshold. In silent conditions, RTs were accurately predicted by $L_{AE}(t)$, with onset-envelope shape influencing early energy accumulation. In noise conditions, RTs increased systematically with spectral proximity between target and masker, consistent with auditory filter theory. When spectral separation exceeded approximately four ERB numbers, masking effects were minimal and RT approached silent-condition values. These findings demonstrate that perceptual detection timing is governed by cumulative acoustic energy and spectral masking rather than instantaneous sound pressure level. The $L_{AE}(t)$ model provides a detection-oriented metric that complements conventional room-acoustic parameters and may support evidence-based design of perceptually robust auditory signals in architectural environments.

Keywords: pure tone; envelope; reaction time; sound-exposure level

1. Introduction

In architectural acoustics, the design of sound sources in built environments is not limited to achieving desirable reverberation characteristics or sound insulation performance; it also requires ensuring that auditory signals are perceptually effective within complex acoustic spaces. In public buildings such as railway stations, airports, and hospitals, auditory signs play a crucial role in guiding occupants, particularly visually impaired users. However, these signals are often presented in acoustically challenging environments characterized by background noise, spectral masking, and reflections. Therefore, understanding how listeners detect time-varying sounds under realistic acoustic conditions is essential for evidence-based acoustic design.

Reaction time (RT) provides a quantifiable index of auditory detectability that integrates perceptual, neural, and motor processes. In room-acoustic contexts, RT can serve as an objective measure of how effectively a sound source emerges from background noise and becomes perceptually salient. Although previous studies have investigated speech intelligibility and perception of loudness, fewer studies have examined the temporal dynamics of signal detection in noisy environments.

RTs contain several time components in different processes; therefore, they result from several time-consuming processes that cannot easily be separated. Luce (1986) described five possible pro-

cesses: (1) signal transduction into neural spikes, (2) transmission of spikes to the brain, (3) signal processing and motor programming for the target muscle (finger or mouth), (4) signal transmission to muscles, and (5) muscle contraction [1]. Because RT reflects the combined duration of these perceptual and neural processes, it is sensitive to the physical and spectral characteristics of the acoustic stimulus.

Consequently, previous studies investigating RT for auditory signals have examined how variations in stimulus properties influence response latency. The target sounds most commonly studied in reaction-time experiments were tones or bandpass noises, and the RTs associated with these sounds were discussed in terms of frequency, masking, and loudness [2–12]. In unmasked conditions, RTs did not vary with frequency for pure tones when the sensation levels (SLs) were comfortable to hear [7,8,10,12]. In masked conditions, RTs increased when the target tone frequencies were close to the masker frequencies [4,6]. These findings are consistent with auditory filter theory, whereby the perceived loudness of a tone is reduced when its frequency overlaps with that of the masker [13]. However, the aforementioned studies [4,6] investigated only few combinations of masker and maskee frequencies, which was insufficient to fully characterize the relationship between RT and spectral masking.

Some researchers have measured RT using amplitude-modulated tones [5,11]. As hearing stable tones is rare in real life, the findings of these studies can inform sound design, helping to identify sounds that effectively raise awareness. These studies reported that RT decreased as the onset slope increased but reached a plateau once the rise time exceeded approximately 50 ms. This plateau effect has been attributed to temporal integration mechanisms: when signal duration exceeds the critical duration (CD), loudness no longer increases with duration. The CD has been reported to range between 100 and 200 ms [14,15], suggesting that RT may be determined before loudness reaches its steady-state value.

In order to investigate easily detectable notification sounds, Shimokura and Soeta (2022) measured the RTs for several types of birdsongs that are often used as auditory signals for visually challenged people in public spaces in Japan [16]. Although the birdsongs had relatively stable pitch, their amplitude envelopes varied substantially. To account for this variation, they introduced a time-integrated sound-exposure level (L_{AE}) to estimate the RTs for the signals with time-varying loudness. The L_{AE} is the sum of the squared sound pressure over a period typically less than 1 s [17], and it quantifies the loudness of a non-steady state sounds well [18]. Time-integrated $L_{AE}(t)$ was then used to estimate the time at which sounds conforming to the concept of integrated loudness becomes perceptually noticeable [19–21]. RTs estimated using the $L_{AE}(t)$ model for birdsongs were more accurate than RTs estimated using the Zwicker loudness model for time-varying sounds with instantaneous loudness [13,22].

In their study on estimation of single RT, Miller and Ulrich (2003) proposed a parallel grain model (PGM), in which each signal is represented by a number of grains of information or activation. They termed the time required for the activation of a single grain to occur as an ‘activation time’ [23]. These activation times depended on the signal intensity (e.g., 10 ms for an intense signal and 40 ms for a weak signal in their examples) and were found to be correlated with RT when combined with the transmission time required for the activated grain to reach a decision center. Applying the concept of arrival time to the time-integrated intensity model suggests that the timing at which a tone is noticed may be determined by time-integrated intensity, with the integration beginning after the activation.

Temporal integration has often been modeled using a “leaky integrator,” which accumulates input over time while allowing gradual decay governed by a time constant. Such models are physiologically plausible and have been used to explain hearing thresholds through neural spike integration [24–26]. Two-time-constant models, whether arranged in series or parallel [27,28], provide improved predictions of hearing thresholds and loudness and have been standardized in ISO procedures [13]. Alternatively, Heil and Neubauer (2003) proposed a statistical neural model in which detection occurs when a criterion number of neural events is reached [29,30]. These approaches share a common principle with the $L_{AE}(t)$ framework: detection occurs when cumulative neural or acoustic energy

exceeds a threshold. In our previous study [16], the $L_{AE}(t)$ model was shown to predict reaction times more accurately than an instantaneous loudness model based on the Zwicker method, suggesting that cumulative energy integration better captures perceptual detection timing for time-varying sounds. However, whereas prior studies primarily focused on threshold estimation within the first 100 ms of stimulation, the present study examines RT for one-second signals, where extended temporal accumulation may further influence detection timing.

Spectral masking effects on RT have been reported previously. The RTs to tones increased due to masking by noise when the tone and background noise were spectrally close [4,6]. In contrast, Emmerich et al. (1976) reported that a tonal background could accelerate the RT to a tone when the spectral energies were distinct [6]. This acceleration of RT could even be observed for the RT to birdsongs, especially for the elderly participants [16], and has been discussed in association with “stochastic resonance” [31–36] and an “inverted-U-shaped manner” [37–39]. As mentioned above, Miller and Ulrich (2003) proposed the PGM to explain the attention timing, whereby grains of information were statistically stimulated not only by the target signal, but also by other auditory inputs [23]. Therefore, we can conclude that the noise energy spectrally separated from a target signal may also contribute to decision-making activation, even if it does not mask the target signal.

The aim of this study is to clarify how spectral characteristics of background noise influence RT to pure tones. Two pure tone at 1 and 2 kHz were masked by white noise and six bandpass noises positioned at varying spectral distances from the target frequency. Twenty young and twenty elderly participants reacted to those tones. The obtained RTs were analyzed using our proposed $L_{AE}(t)$ model, whereby the envelopes of the pure tones were modulated in three fade-in ways to create differences in the corresponding $L_{AE}(t)$ curves. Taking into account the activation time in the PGM [23], the cumulative start time of the $L_{AE}(t)$ curve occurred after the onset of the pure tones in several intervals. As in our previous study [16], RT was estimated as the moment when $L_{AE}(t)$ reached a criterion value. By introducing a time-integrated sound-exposure level model, $L_{AE}(t)$, we aim to provide a physically grounded and psychoacoustically interpretable framework that links acoustic energy integration, spectral masking, and perceptual detection timing. Such an approach contributes to the design of detectable auditory signals in architectural spaces.

2. Materials and Methods

2.1. Participants

Twenty young adults (9 men and 11 women; mean age: 27.2 ± 8.5 years) and twenty elderly adults (13 men and 7 women; mean age: 70.9 ± 4.1 years) participated in the experiment. All participants reported normal hearing for their age and had no history of neurological or auditory disorders. None of the participants used hearing-support devices. All participants provided informed consent prior to participation. Approval for the experimental protocol (Human 2020-0227L) was generated by the institutional ethics committee.

2.2. Acoustic Stimuli and Spectral Masking Conditions

Table 1 provides an overview of all signals and conditions. The target signals were pure tones of 1 and 2 kHz, with envelopes modulated in three types of fade-in envelope (flat, gentle, and steep slopes), as shown in Figure 1. The flat-slope signal did not modulate the envelope. The gentle- and steep-sloped signals had envelopes that increased at $0.5(t + 1)$ and t (t : time [s]), respectively. The signals had an initial duration of 1 s and a 5 ms Hanning taper at the onset to avoid clicking. The taper part was included in the signal duration. As the RTs in this study were expected to fall within the time duration, the taper was not applied at the offset, as shown in Figure 1. The presentation levels for the three signals were set at 10, 20, and 30 dB above the participant’s hearing threshold, as indicated in dBSL. According to the categorical unit of loudness, the pure tone at 10 dBSL was “very soft”, whereas the pure tone at 30 dBSL was “soft” [40].

Table 1. Summary list of signals and experimental conditions.

Signal	Frequency:	1 and 2 kHz
	Envelope:	Flat, Gentle and Steep
Silent condition	Signal level:	10, 20 and 30 dBSL
Noise condition	Signal level:	30 dBSL
	Noise type:	White noise and Six bandpass noises
	SNR:	High and Low

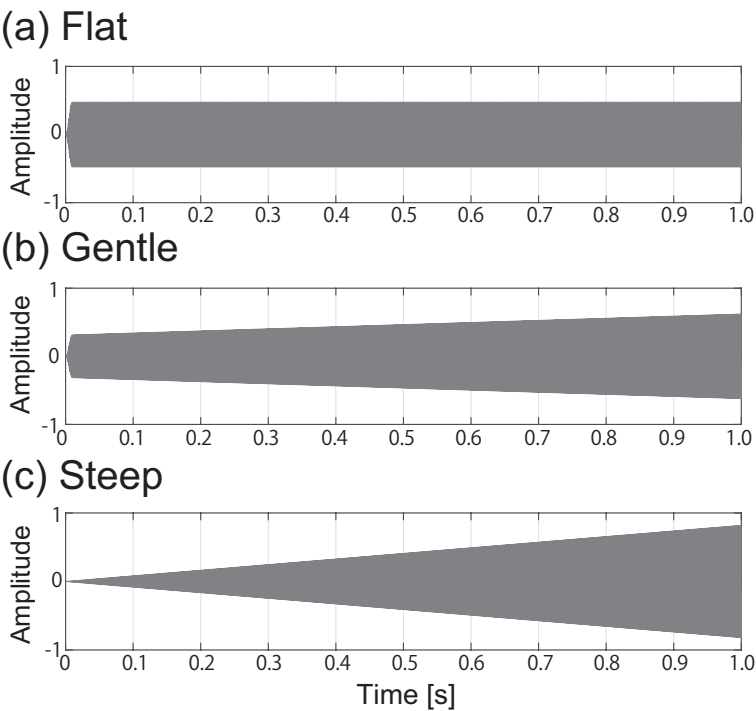


Figure 1. Waveforms of pure tones with (a) flat, (b) gentle, and (c) steep slopes.

Bandpass noise was generated using a Gamma-tone filter bank designed to approximate the human auditory filter (MATLAB, MathWorks, Natick, USA) [41]. White noise was passed through a bank of Gammatone filters that were equally spaced on an equivalent rectangular bandwidth (ERB) scale [42,43]. For example, a pure tone of 1 kHz will primarily activate the ERB filter with a number of 15 (center frequency: 1057 Hz). Figure 2 shows the spectra of the bandpass noise (solid lines) and the pure tone (dashed line) when they were calculated for 2048 samples and averaged over 1 s. Thus, the bandpass noise that masked the 1 kHz pure tone was set to numbers 10, 12, and 14 when shifted downwards, and 16, 18, and 20 when shifted upwards in the high-SNR condition (Figure 2a). In the low-SNR condition, the ERB numbers used were 9, 11, 13, 17, 19, and 21 (Figure 2b), as listed in Table 2. Preliminary experiments showed that when the bands closest to 1 kHz (i.e., 14 and 16) were used at a low SNR, participants failed to hear the targets several times. Therefore, we shifted the signals outward against the pure-tone frequencies by one ERB number. The ERB numbers are referred to as BD3, BD2, BD1, BU1, BU2, and BU3 in ascending order (BD: band down, BU: band up). The ERB numbers for a pure tone of 2 kHz were selected in the same manner and are listed in Table 2.

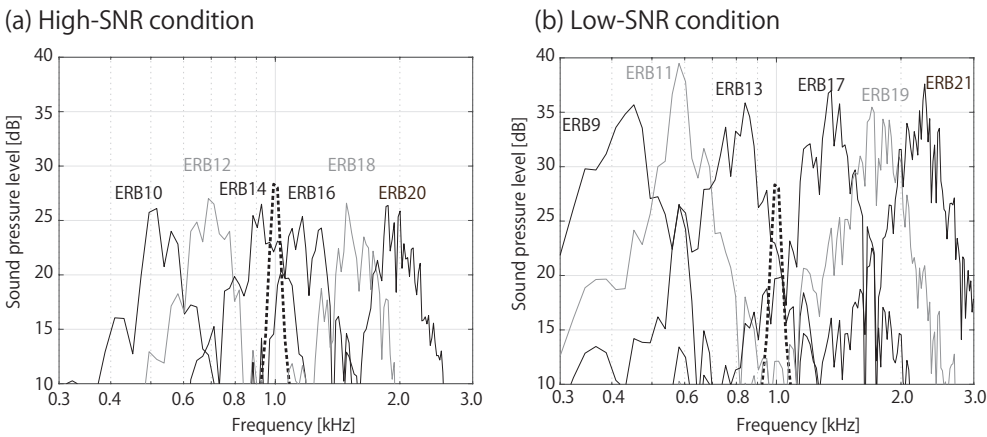


Figure 2. Spectra of bandpass noises (rigid lines) and a pure tone (dot line) in (a) high- and (b) low-SNR conditions.

Table 2. ERB number and center frequency of bandpass noise.

		1 kHz Pure tone					
		BD3	BD2	BD1	BU1	BU2	BU3
High-SNR	ERB number	9	11	13	17	19	21
	Frequency [Hz]	439	602	805	1371	1762	2247
Low-SNR	ERB number	10	12	14	16	18	20
	Frequency [Hz]	516	698	924	1206	1556	1991
		2 kHz Pure tone					
		BD3	BD2	BD1	BU1	BU2	BU3
High-SNR	ERB number	15	17	19	21	23	25
	Frequency [Hz]	1057	1371	1762	2247	2852	3603
Low-SNR	ERB number	14	16	18	22	24	26
	Frequency [Hz]	924	1206	1556	2533	3207	4045

2.3. Apparatus

The sensation level (SL) of the target signal and the signal-to-noise ratio (SNR) of the noise were controlled according to the individual’s hearing threshold. Therefore, prior to the RT measurements, we used a transformed up-down procedure [44] to determine the hearing thresholds for pure tones at 1 and 2 kHz, which served as the target signals. Threshold tests were conducted using a two-alternative forced-choice procedure with diotic listening in the soundproof chamber.

As there were no significant differences in hearing thresholds according to the envelope modulations of the pure tone (i.e., flat, gentle or steep slopes as shown in Figure 1), the threshold in the pure tones was represented with a flat slope. Finally, the level of the pure tone and SNR were determined using the measured hearing thresholds. The averaged sound-pressure levels at 0 dB SL were 1.2 ± 4.9 dB at 1 kHz and 2.4 ± 4.4 dB at 2 kHz for young participants and 23.7 ± 9.6 dB at 1 kHz and 27.1 ± 10.1 dB at 2 kHz for elderly participants.

Participants were instructed to tap an empty carton with their fingers as quickly as possible when they detected the target tone. In the RT measurements conducted under masking conditions, pure tones at 30 dBSL were presented against a background noise. Based on preliminary RT data obtained from four young and four elderly individuals (who did not participate in the present experiment), SNRs were determined to ensure reliable detection across participants. For the young group, SNRs of 0 and –20 dB were selected, whereas for the elderly group, SNRs of 10 and –10 dB were selected. These were categorized as high SNR (0 dB for young; 10 dB for elderly) and low SNR (–20 dB for young; –10 dB for elderly) conditions.

RT measurements were conducted under three acoustic conditions: silent, white-noise, and bandpass-noise conditions (Table 1). In the silent condition, 18 types of pure tone (two frequencies, three envelope shapes, and three presentation levels) were presented four times in a random order.

The inter-stimulus interval was randomly varied between 3 and 12 s to prevent participants from predicting the timing of the onset.

In the white noise and bandpass noise conditions, the target signals were limited to 30 dBSL, and the white and bandpass noises were presented continuously through headphones. To avoid the effects of fatigue, the experiment was separated into five sessions, each session lasting 6–7 minutes, with sufficient breaks between sessions. The RT values were averaged over four trials, and non-response trials were excluded.

2.4. Estimation of RT

Based on the RT measurements conducted using birdsong [16], a time-integrated $L_{AE}(t)$ was introduced to estimate the RT for pure tones. $L_{AE}(t)$ was calculated as follows:

$$L_{AE}(t) = 10 \log_{10} \int_0^t \frac{P(s)}{P_{min}} ds, \quad (1)$$

where $P(s)$ is the sound pressure [Pa] and P_{min} is the pressure at hearing threshold [Pa]. In this equation, 0 s represents both the start time of the integral and the onset of the signal. Figure 3 shows the $L_{AE}(t)$ curves for pure tones with flat, gentle, and steep slopes, with the sound pressure adjusted to 0 dB SL¹. Therefore, for these stimuli, the $L_{AE}(t)$ was 0 dBSL after one second. As shown in Figure 3, the integration curve of the pure tone with a steep envelope rose gently. Unlike the $L_{AE}(t)$ of birdsongs, the curves rose smoothly. Therefore, approximations by power functions were unnecessary and the curves could be expressed mathematically. The time-cumulative pressures on the flat (CP_F), gentle (CP_G), and steep (CP_S) slopes are expressed as follows:

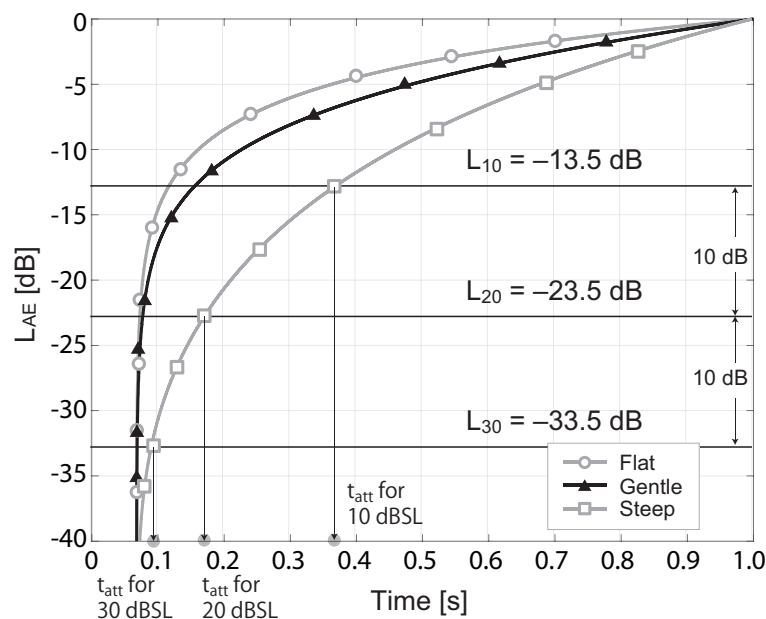


Figure 3. Time-functional sound-exposure levels ($L_{AE}(t)$) of pure tones with flat, gentle, and steep slopes and example of t_{att} for the steep-sloped tone at 10 (L_{10}), 20 (L_{20}), and 30 (L_{30}) dBSL.

$$CF_P(t) = \int_0^t \sin(2\pi fs)^2 ds, \quad (2)$$

$$CF_G(t) = \int_0^t \left\{ \frac{1}{2}(s+1) \sin 2\pi fs \right\}^2 ds, \quad (3)$$

¹ The $L_{AE}(t)$ curves shown in Figures 5 and 6 of the RT study with birdsongs [16] could not be divided by the sampling rate after the discrete integrated energy was calculated, and the $L_{AE}(t)$ was normalized by the minimum audible pressure (2.0×10^{-5} Pa). Consequently, the values on the vertical axis differ from those in Figure 3 of this paper.

$$CF_S(t) = \int_0^t (s \sin 2\pi fs)^2 ds, \quad (4)$$

where f is the frequency of the pure tone [Hz]. After solving the above integrals and using the envelopes of the sounds, $L_{AE}(t)$ in flat (L_{AEF}), gentle (L_{AEG}), and steep (L_{AES}) slopes can be approximated using the following equations:

$$L_{AEF}(t) \approx 10 \log_{10} t, \quad (5)$$

$$L_{AEG}(t) \approx 10 \log_{10} \frac{t^3 + 3t^2 + 3t}{7}, \quad (6)$$

$$L_{AES}(t) \approx 10 \log_{10} t^3, \quad (7)$$

As t approaches 1, $L_{AE}(t)$ converges to 0 dB. The frequency of the pure tone does not influence $L_{AE}(t)$.

This model hypothesizes that a listener responds when the cumulative energy of the target signal reaches a criterion corresponding to 0 dBSL (i.e., the hearing threshold). Accordingly, the RT to pure tones above 0 dBSL was estimated using t_{att} , defined as the time at which $L_{AE}(t)$ reaches a certain value. Figure 3 shows examples of determining the t_{att} of the steep-sloped signal. The certain values differ depending on the target level but increase linearly in dB units. Therefore, the values for the signals at 20 and 10 dBSL are 10 dB and 20 dB higher, respectively, than those at 30 dBSL ($L_{20} = L_{30} + 10$ and $L_{10} = L_{30} + 20$). When the target signal is close to threshold, a longer time is required, and the listener reacts later. If the listener continues to hear the target signal throughout the entire time (1 s), the cumulative energy will reach the SPL at their hearing threshold and they will clearly notice it.

In the noise condition, it is assumed that the amount of masking determines the detection criterion in the model. Previous masking studies have shown that the loudness of a pure tone is reduced by a masker within the same critical band, and the band sensation level is almost equal to the amount of masking [45,46]. The band sensation level was defined as the effective masking level within the auditory filter centered at the signal frequency. In this study, bandpass noises were generated using ERB-wide gammatone filters with center frequencies that were equally spaced on an ERB-number scale. To obtain the masking amount, the background noise was passed through the ERB filter corresponding to the signal frequency, and the sound energy within that filter was calculated. For example, the level of the white noise was 50 dBSL in the low-SNR condition for young participants. This follows from the fact that the pure tone was presented at 30 dBSL with an SNR of -20 dB, where SNR was defined as the difference between signal and noise levels. After generating white noise at 50 dBSL, the band sensation level for a pure tone of 1 kHz was quantified by the sound energy passing through the fifteenth ERB filter. The band sensation level for each bandpass noise was derived using the same procedure.

3. Results

3.1. RT in the Silent Condition

Figure 4 shows the average RT of the 20 young (upper panel) and 20 elderly (lower panel) participants. Regardless of the envelope shape, the RT increased as the SL decreased. In other words, the participants reacted faster to clearly audible pure tones. For the envelope, the RTs for the flat and gentle slopes were similar, whereas the RTs for the steep slopes were longer. These differences increased as SLs decreased. A two-way analysis of variance revealed significant effects of SL and envelope on the RT were significant for both tones and participant groups ($p < 0.01$ in all cases). As normal distributions were not observed for the RTs of 1 kHz and 2 kHz in each participant group, a non-parametric analysis (paired Wilcoxon signed-rank test) was conducted. No significant differences were observed between the RTs for 1 kHz and 2 kHz in young participants ($p = 0.85$), but significant differences were observed in elderly participants ($p < 0.05$). According to the unpaired Wilcoxon test,

elderly participants had significantly shorter RT than those in young participants ($p < 0.01$), as the SPLs of the target signals were higher at the same SL.

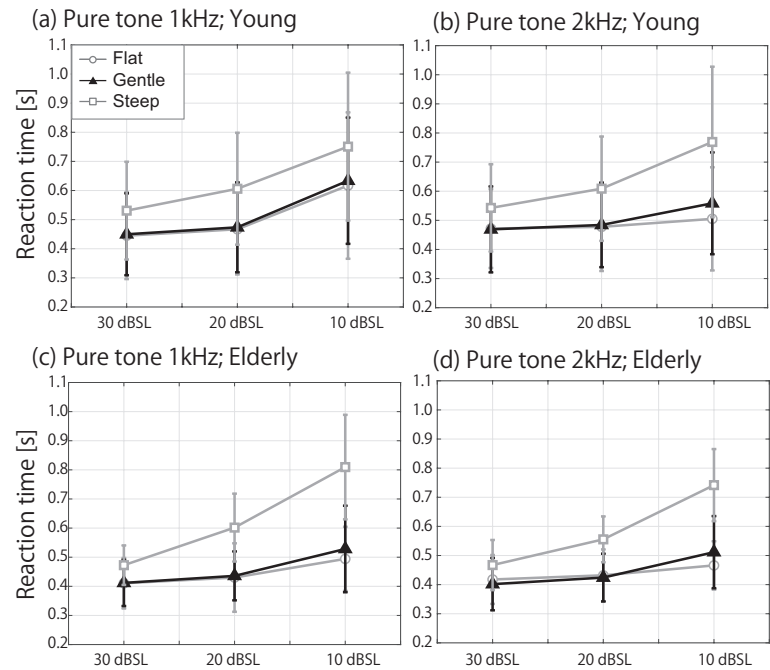


Figure 4. Averaged reaction times in the silent condition in each SL (error bar: standard deviation).

The data are in agreement with the model’s predictions. By varying the L_{30} in 0.5 dB steps, we identified the most likely values of L_{30} , L_{20} , and L_{10} . This produced a logarithmic t_{att} that correlated most highly with the measured RTs ($RT \approx a \log_{10} t_{att} + b$; a , b : constants). Furthermore, taking into account the potential for a delayed start time for integration, we changed the start time for the calculating $L_{AE}(t)$ in 10 increments (0–90 ms in 10 ms intervals). For example, when the start time is 50 ms, a zero is inserted for the first 50 ms before $L_{AE}(t)$ is calculated. We estimated the average RTs at both 1 and 2 kHz using a unique $L_{AE}(t)$ model for each young and elderly participant, as shown in Figure 4.

Table 3 shows the start time [s], L_{30} [dB], a , b [s], the correlation coefficient, and the averaged error [s] between the measured and estimated RTs when the correlation coefficient between the common-logarithmic t_{att} and the measured RT was the highest. The model parameters were fitted to the data, the best fit was achieved when a delayed integration starting 60 ms after onset was assumed for both young and elderly participants. The estimated L_{30} values were -33.5 dB and -38.0 dB for young and elderly participants, respectively. Estimation accuracy was high for both groups, with correlation coefficients of 0.96 for young participants and 0.99 for elderly participants, both statistically significant at $p < 0.01$. These values were comparable to the accuracy previously reported for RTs to birdsong in young participants ($r = 0.98$, $p < 0.01$) [16].

Table 3. Parameters when t_{att} s were the most correlated with RTs.

		Start time [sec]	L_{30} [dB]	a	b [sec]	Correlation coefficient	Averaged error [sec]
Silent condition	Young	0.06	-33.5	0.38	0.92	0.96	0.02
	Elder	0.06	-38	0.57	1.11	0.99	0.01
Noise condition	Young	0.08	-31	0.60	1.14	0.90	0.04
	Elder	0.02	-40	0.25	0.85	0.93	0.02

3.2. RT in Noise Conditions

Figure 5 shows the average RT under white-noise conditions. For reference, the average RT for the tone at 30 dBSL in the silent condition is included. RT increased slightly when background noise was added to the high-SNR condition. In contrast, the RTs in the low-SNR condition were considerably longer. In particular, responses to tones with steep slopes were notably slower, as were responses to tones with gentle slopes compared to those to tone with flat slopes. The prolongation of RTs under the low-SNR condition was more pronounced for young participants because the assigned SNR was lower for young participants (−20 dB) than for elderly participants (−10 dB).

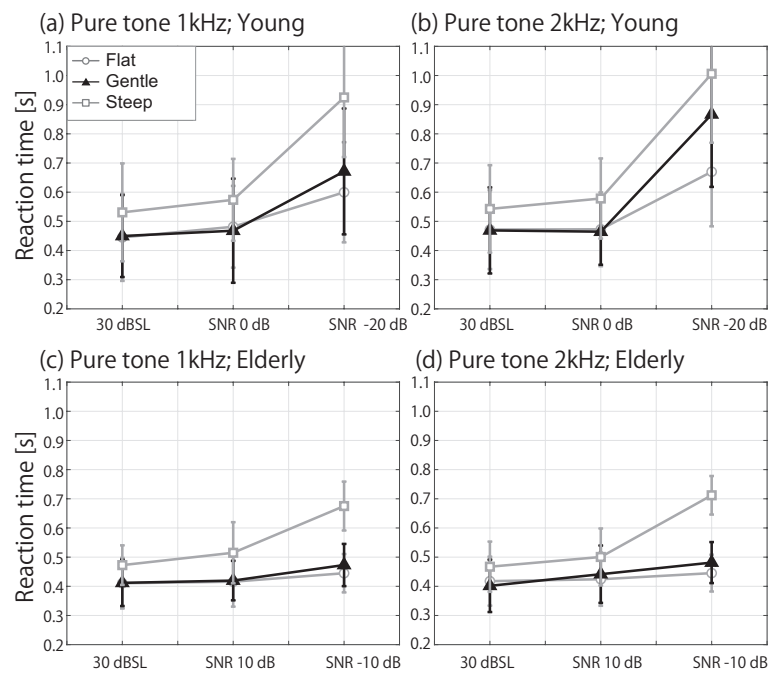


Figure 5. Averaged reaction times under white noise in each SNR (error bar: standard deviation).

In the bandpass noise condition, the RTs were arranged according to the ERB number for different SNRs (Figures 6 and 7). The vertical rigid line indicates the ERB number containing the pure tone frequency. The horizontal dotted lines show the RTs at 30 dBSL in the silent condition. When the ERB number of the bandpass noise approached the pure tone frequency, the RTs increased (i.e., BD1 and BU1). By contrast, bandpass noises that were spectrally distinct from the pure tone resulted in RTs similar to those observed in the silent condition (i.e., BD3, BD2, BU2, and BU3). This tendency was clearly observed for steeply sloped tones, with higher vertex RTs around the frequencies of pure tones were at a low SNR (Figure 7). In other words, if the spectral distance between the bandpass noise and the pure tone was further by four ERB numbers (1 kHz with an ERB number of 15 was below 11 and above 19, and 2 kHz with an ERB number of 20 was below 16 and above 24), the effect of the noise on RT could be minimized. RT increased more with the ERB number around the frequencies of pure tones for young participants, because the assigned SNR was lower for them than for elderly participants.

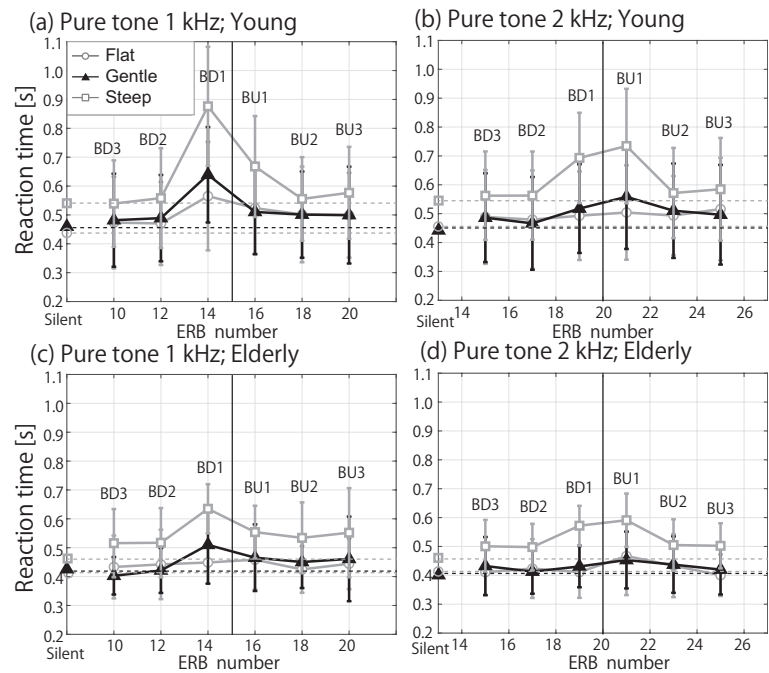


Figure 6. Averaged reaction times under the high-SNR bandpass noise for each ERB number (error bar: standard deviation).

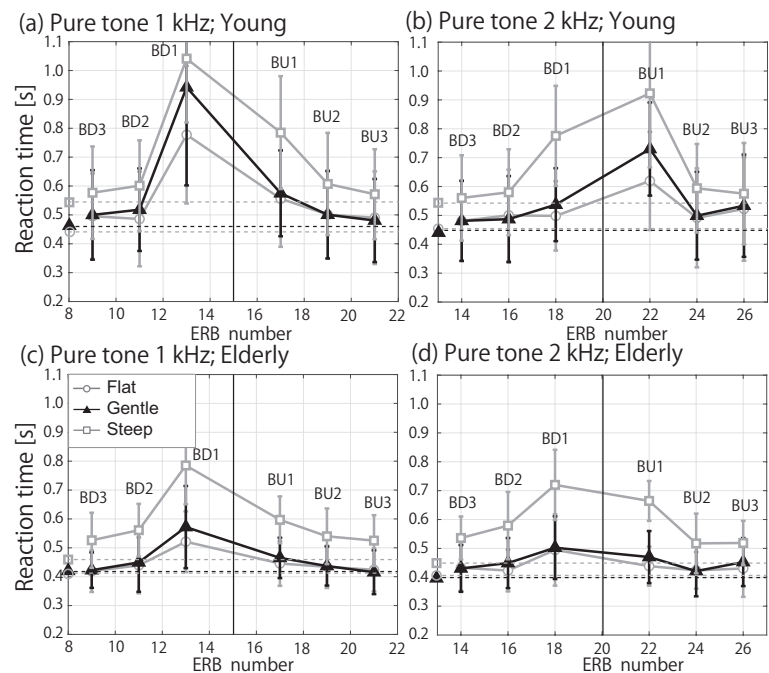


Figure 7. Averaged reaction times under the low-SNR bandpass noise for each ERB number (error bar: standard deviation).

The RTs for the background noise were estimated using the $L_{AE}(t)$ curves and the amount of masking. Figure 8 shows the relationship between RT and the masking amount calculated from white and bandpass noises. For the young participants (Figure 8a), RT remained low until a masking amount of 10.5 dB, after which it increased in two stages, at 22.0 and 29.4 dB. These three categories are referred to as “mask low,” “mask middle,” and “mask high.” In the high-SNR condition, the bandpass noises BD1 and BU1 were categorized as mask-middle, while in the low-SNR condition, the white and bandpass noises BD1 and BU1 were categorized as mask-high. All other noise conditions were categorized

as mask low. For elderly participants, the masking amounts in the mask-low, -middle, and -high categories were 0.37, 12.1, and 19.2 dB, respectively (Figure 8b).

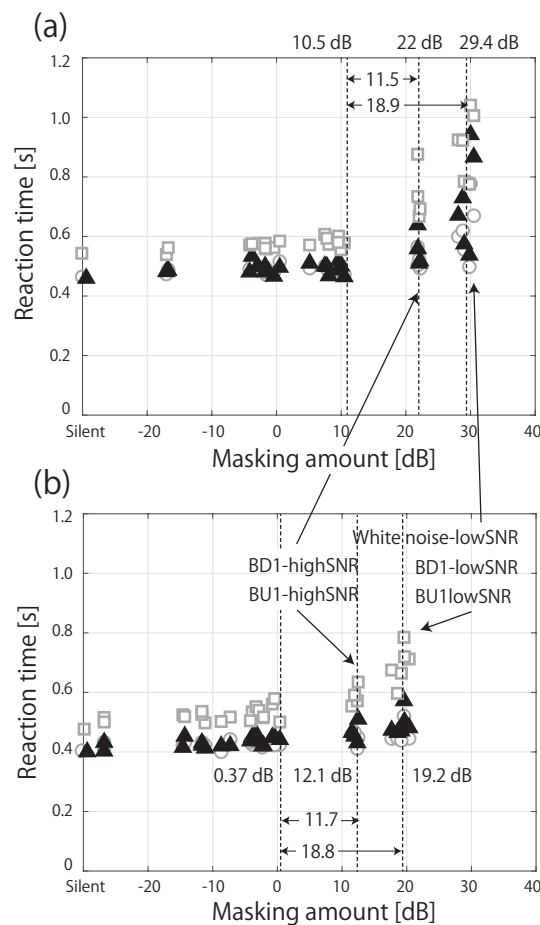


Figure 8. Relationships between reaction time and masking amount for (a) young and (b) elderly participants.

As the RTs in the mask-low category were similar to those in the silent condition, the noise conditions in this masking category can be considered to have a slight effect on RT. Therefore, if the $L_{AE}(t)$ value used to estimate the RTs in the mask-low category is assumed to be the same as L_{30} , the standard values for the mask-middle (L_{MM}) and mask-high (L_{MH}) categories should be $L_{30} + 11.5$ ($= 22.0 - 10.5$) dB and $L_{30} + 18.9$ ($= 29.2 - 10.5$) dB, respectively, for the young participants. As in the silent condition, $L_{AE}(t)$ estimation in the noisy condition can be performed by exploring the t_{atts} to reach L_{30} , L_{MM} , and L_{MH} . For the elderly participants, the L_{MM} and L_{MH} were $L_{30} + 11.7$ ($= 12.1 - 0.37$) dB and $L_{30} + 18.8$ ($= 19.2 - 0.37$) dB, respectively. For both sets of participants, the L_{MM} and L_{MH} were approximately 10 and 20 dB higher than L_{30} , respectively. Table 3 shows the estimation results and Figure 9 illustrates comparison between the measured and estimated RTs for each category. As this model estimated the average RTs in the three categories for pure tones at 1 and 2 kHz, the estimated RTs (dotted lines) were the same for pure tones at these frequencies. As shown in Figure 9, the estimation accuracy was very high for both groups, with correlation coefficients of 0.90 for young participants and 0.93 for elderly participants, both statistically significant at $p < 0.01$. The start time for calculating $L_{AE}(t)$ was shorter for the elderly participants. For both groups, the estimated L_{30} values were close to those in the silent condition.

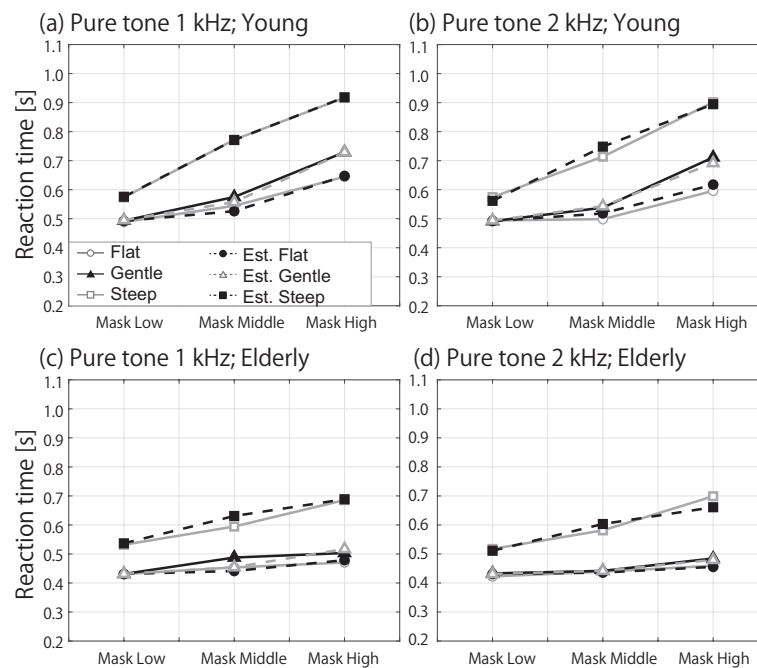


Figure 9. Measured (rigid lines) and estimated (dot lines) reaction times for mask-low, -middle, and -high categories.

4. Discussion

4.1. Temporal Integration Mechanism and the $L_{AE}(t)$ Model

As shown in Figure 4, the RTs to pure tones with steep slopes were more extended for the sound pressures in the lower SL. These results were consistent with the extensions of t_{att} in the L_{AE} curves (Figure 3). The $L_{AE}(t)$ model could then estimate the RTs to pure tones in the silent condition with very high level of accuracy ($r = 0.96$ for young participants and $r = 0.99$ for elderly participants). These estimation accuracies were similar to those for birdsong RT for young participants ($r = 0.98$) but higher for elderly participants ($r = 0.72$) [16]. Changing the targets to simple pure tones resulted in the $L_{AE}(t)$ model performing better for elderly participants. In the auditory temporal integration, the effective time constant (i.e., CD) just above the hearing threshold decreases to approximately 100 ms [29]. This means that little intensity accumulation occurs for determining the hearing threshold when the signal duration is longer than 100 ms. In this study, the presentation levels were below 30 dBSL (i.e., closed to the thresholds) and were one second long; the remaining 900 ms may therefore have had little influence on the shift in hearing threshold. However, L_{AE} curves that accumulated for one second were important for estimating the RTs of the one-second-long signals.

Unlike the models used in the birdsong experiment, the $L_{AE}(t)$ calculation was normalized to the hearing threshold, as shown in Equation 1. This improvement helps us to understand the model's meaning. Since the $L_{AE}(t)$ is calculated from the target signal at the auditory threshold, the L_{AE} at 1 s (the target's duration) is 0 dBSL. In other words, any participant could detect the target signal if they listened to it for its entire duration. The L_{AE} values required to notify the pure tones at 30, 20, and 10 dBSL were -33.5 , -23.5 , and -13.5 dB, for the young participants, respectively. Interestingly, the absolute values of the L_{AE} were almost equal to the SLs. Therefore, we can hypothesize that participants may notice the signal at 30 dBSL when the L_{AE} value is lower than their hearing threshold by 30 dB. Drawing lines for L_{40} and L_{50} in Figure 3 additionally shows that the RTs to the pure tones at 40 and 50 dBSL were 470.16 and 469.97 ms, respectively, for flat-sloped signals. For signals with a steep slope, the estimated RTs at 40 and 50 dBSL were 479.67 and 470.70 ms, respectively. The estimated RTs remained almost unchanged at higher SLs compared to 30 dBSL. The RTs converged at sufficiently high SLs, but differed from previous studies due to variations in the number and age of participants, instructions, and amount of training. However, there was a common tendency: the RTs increased rapidly when the SLs of the signals were less than 30 dB [10,11]. Even when the target signals were

narrow-band noise, the measured RTs were prolonged at the low levels [47]. As shown in Figure 3, the $L_{AE}(t)$ curves indicate a slight increase in the RT for an SL higher than 30 dBSL, and they may be applicable to any envelope modulation of the target signal.

Another characteristic of $L_{AE}(t)$ was its independence from the frequency of the pure tone. The $L_{AE}(t)$ curves can be approximated using equations 5, 6 and 7, which do not include frequency as a variable. In fact, there was little difference in the measured RTs between the 1 and 2 kHz pure tones. This independence from frequency has been observed in previous studies, not only for pure-tone targets [10], but also for narrowband noise targets [47].

4.2. Influence of Spectral Masking

As in previous studies, RT increased with louder background noise [9,48,49]. The effect of noise on the RT may be influenced by the spectral masking of the pure tone, as shown in Figures 6 and 7, in a manner similar to that observed for loudness [45,46]. Using the $L_{AE}(t)$ model to apply the masking amount to obtain t_{att} produced a highly accurate estimate of RT for both participants, as shown in Figure 9. For instance, the young participants' RT did not increase significantly up to a masking amount of 10.5 dB (mask low), and they reacted more slowly to background noise that was 11.5 dB (mask middle) and 18.9 dB (mask high) louder than the mask-low category (Figure 8). As the most likely L_{30} in the noise condition was -31 dB (Table 3), the L_{MM} and L_{MH} were -19.5 and -12.1 dB, respectively. If the SNR decreased by a further 10 dB (i.e., to SNR -30 dB), the L_{AE} would reach almost 0 dBSL. In the preliminary experiments, almost all the participants could not hear the pure tones in the white noise under an SNR of -30 dB; therefore, the $L_{AE}(t)$ model could explain the upper limit of the masking amount.

The RTs under the noise condition were influenced not only by the noise energy but also by the spectral distance from the target signal's frequency. These spectral effects can be seen in the measured RTs in the bandpass noise (Figures 6 and 7). The bandpass noises used in this measurement were synthesized using a Gammatone filter scaled by auditory critical bands. Thus, the bandpass noises of BD1 and BU1, which overlapped with the pure-tone band spectrally (ERB 15 for 1 kHz and ERB 20 for 2 kHz), increased their masking amounts, resulting in longer RTs. However, bandpass noise that was spectrally distant (BD3, BD2, BU2, and BU3) produced RTs that were similar to those in the silent condition. In other words, the masking effect had little impact on RT. When the center frequency of the noise was more than four ERBs away from the pure tone frequency, the masking effect was almost nullified. Although some studies have reported that the small amounts of RT data measured under noise only indicate the possibility of spectral masking of RTs [4,6], our structured results based on the Gammatone filters clearly demonstrate the spectral RT behaviors along the frequency range.

4.3. Start Time for Calculating L_{AE}

To maximize the correlation between the measured and estimated RTs, the most suitable start time for calculating L_{AE} was 60 ms after the onset of the signal. As the Hanning taper of 5 ms was applied at the onset of the pure tones, the start time may include the taper length. We hypothesized that the start time might be caused by the activation time; however, we may need to reconsider this interpretation. A limitation of the $L_{AE}(t)$ model is that the t_{att} approaches to 0 s as the signal becomes more intense, as illustrated in Figure 3 and equation 5. As the t_{att} and RT were logarithmically related ($RT \approx a \log_{10} t_{att} + b$), the RT diverges to $-\infty$ in this case. This problem can be solved by changing the start time of integration. For example, considering the 60-ms delay, the time-cumulative pressure of a pure tone with a flat slope (CPF) changes to

$$CPF(t) = \int_{0.06}^t (\sin 2\pi fs)^2 ds, \quad (8)$$

and the L_{AEF} can be approximated to

$$L_{AEF}(t) \approx 10 \log_{10}(t - 0.06), \quad (9)$$

According to equation 9, the t_{att} approaches, but does not fall below, 60 ms regardless of the intensity of the signal. For young participants and flat-sloped pure tones, the estimated RT was 460 ms, and the RT for pure tones with a flat slope is likely to be around 460 ms if the presentation level exceeds 30 dBSL, as shown in Figures 4a and 4b. Interestingly, the start time defines a possible minimum RT in this condition. As discussed above, the delayed start time of integration is also consistent with the previous studies [10,11,47] showing that the response time peaks around 30 dBSL.

4.4. Age-Related Effects

Although the L_{30} value for young participants was almost the same as the SL of the target signal, the L_{30} value for elderly participants was lower, as shown in Table 3. According to our model, elderly participants noticed the sound more quickly than younger participants. Although the presentation level was normalized by the individual hearing threshold, the measured RTs were significantly faster in elderly participants. As the presentation levels were determined based on the SL, the SPL of the pure tone for the elderly participants was approximately 24 dB higher than that for the young participants. Simply aligning the SLs of young and elderly participants made it difficult to compare their RTs. The larger variance of hearing thresholds in the elderly participant group is another issue that needs to be addressed when discussing the effect of ageing on RT.

The start times for both participant groups were the same in the silent condition (60 ms), whereas in the noise condition there was a significant difference between the two groups (80 ms for young participants and 20 ms for elderly participants). For pure tones with a flat slope, the minimum RTs were 488 ms for young participants and 430 ms for elderly participants. This difference in RTs can be seen in Figure 9. In order to enable the elderly participants to react to the signals as many times as possible, the assigned SNRs were higher than for the young participants. However, such experimental adjustments made discussions on the effects of ageing difficult, as with the silent conditions.

RT experiments using birdsong have reported that some bandpass noises can shortened the RTs, particularly for elderly participants [16]. In this study, the bandpass noise in BD3, BD2, BU2, and BU3 shortened the RTs for 38% of the responses from young participants and 39% of the responses from elderly participants. However, this tendency was not distinctive for elderly participants. One possible explanation is that the spectral overlap between the target tone and masker was insufficient to reveal age-related differences in auditory filtering. To investigate this possibility, additional RT measurements should be conducted using noisy targets whose spectra span multiple auditory filters. Under such conditions, individual audiograms may have a stronger influence on RTs and may reveal age-related differences. An elderly person often has a sensorineural hearing loss, which broadens their auditory filter and make it difficult to separate a target signal from a background noise [50]. Therefore, before conducting such RT measurements, participants' audiograms should be examined and notched-noise masking tests should also be conducted to estimate individual auditory filter bandwidths.

4.5. Practical Implications for Architectural Sound Design

From the perspective of architectural acoustics, the present findings have practical implications for the design of perceptually effective auditory signals in built environments. In transportation facilities, hospitals, and other public buildings, notification sounds must be detectable within complex acoustic fields characterized by background noise and room reflections. The results indicate that detectability is determined not solely by overall sound pressure level but by cumulative acoustic energy relative to spectral masking within auditory critical bands.

Conventional architectural acoustic metrics—such as reverberation time (RT60), clarity (C50/C80), and speech transmission index (STI)—primarily describe spatial acoustic characteristics. The LAE(t)-based approach complements these metrics by providing a detection-oriented parameter grounded in temporal energy accumulation. Designing notification sounds that (1) minimize spectral overlap with dominant environmental noise components and (2) promote rapid early energy accumulation may enhance perceptual salience without increasing overall sound levels.

Of the birdsongs researched, the cuckoo's call, which is often used in Japanese public spaces, was the easiest to notice subjectively [51] and had the shortest t_{att} [16]. In a noisy environment, the sign sounds should differ from the dominant frequency of the noise source by at least four ERB numbers to minimize the masking effects. For example, the sound energy of train noise is distributed below 500 Hz (ERB number 10) [52], so sign sounds in stations should be design using sounds above 924 Hz (ERB number: 14). High-frequency sounds are unpleasant, so the sign sounds with the lowest possible frequency should be selected according to the noise frequency.

5. Conclusions

The $L_{AE}(t)$ model is suitable for estimating RTs for pure tones with different amplitude-modulated envelopes. When the $L_{AE}(t)$ is calculated from pure tones at the listener's threshold, the t_{att} that is highly correlated with the RTs can be approximated to a moment at an L_{AE} lower than 0 dB (the hearing threshold) by almost the SL of the pure tone. For example, RTs measured from a pure tone at 30 dB SL can be estimated at the t_{att} at which L_{AE} reaches -33.5 dB for young participants. However, note that the calculation of $L_{AE}(t)$ begins 60 ms (start time) after the onset of the signal.

When the amplitude-modulated pure tones overlap with the background noise, the RT can be estimated using the $L_{AE}(t)$ model, as in the silent condition. However, the value used to specify t_{att} corresponds to the masking amount against the target pure tone. Therefore, as with subjective loudness, RT is influenced by spectral masking. Differently from the subjective loudness, the pure tones must be at least four ERBs away from the noise frequency to minimize the effect of spectral masking on RT.

Future studies should examine RT estimation using a wider diversity of acoustic stimuli and more complex acoustic environments. In particular, targets with richer spectral and temporal structures and signals that span multiple auditory filters should be investigated to better understand how cumulative acoustic energy interacts with auditory filtering in realistic listening situations. In addition, incorporating individual audiograms and auditory-filter bandwidths into the $L_{AE}(t)$ framework may improve the prediction accuracy of RTs and help clarify age-related differences in auditory processing.

Author Contributions: Ryota Shimokura: Conceptualization, Methodology, Software, Investigation, Analysis, Writing -original draft. Yoshiharu Soeta: Implementation of psychoacoustic tests, Supervision.

Funding: This study was supported by a Grant-in-Aid for Scientific Research (B) from the Japan Society for the Promotion of Science (JP22H03916).

Institutional Review Board Statement: Approval for the experimental protocol (Human 2020-0227L) was generated by the ethics committee of National Institute of Advanced Industrial Science and Technology (AIST).

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: The authors thank the participants for their cooperation during the experiments and the Human Resources Centers for the Aged at Ikeda (Osaka). In addition, the authors thank Ms. Takako Nakazawa, who helped implement the psychoacoustic experiments.

Conflicts of 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.

References

1. Luce, R.D. *Response times: Their role in inferring elementary mental organization.*; Oxford University Press, 1986.
2. Chocholle, R. Variation des temps de réaction auditif en fonction de l'intensité à diverses frequencies. *Annee. Psychol.* **1940**, *41*, 65–124.
3. Burke, K.S.; Crestone, M.E. Shutts, R.E. Hearing loss and reaction time. *Arch. Otolaryngol.* **1965**, *81*, 49–56.

4. Chocholle, R.; Greenbaum, H. La sonie de sons purs partiellement masqués. Étude comparatif par une méthode d'égali sation et par la méthode des temps de reaction. *Journal de Psychologie Normal et Pathologique* **1966**, *63*, 385–414.
5. Warm, J.S.; 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*, 159–160.
6. Emmerich, D.S.; Pitchford, L.J.; Becker, C.A. Reaction time to tones in tonal backgrounds and a comparison of reaction time to signal onset and offset. *Percept. Psychophys.* **1976**, *20*, 210–214.
7. Marshall, L.; Brandt, J. The relationship between loudness and reaction time in normal hearing listeners. *Acta Otolaryngol.* **1980**, *90*, 244–249.
8. Kohfeld, D.L.; Santee, J.L.; Wallace, N.D. Loudness and reaction time: I. *Percept. Psychophys.* **1981**, *29*, 535–549.
9. 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*, 473–476.
10. Epstein, M.; Florentine, M. Reaction time to 1- and 4-kHz tones as a function of level. *Ear and Hearing* **2006**, *27*, 424–429.
11. Schlittenlacher, J.; Ellermeier, W. Simple reaction time to the onset of time-varying sounds. *Atten. Percept. Psychophys.* **2015**, *77*, 2424–2437.
12. Schlittenlacher, J.; Ellermeier, W.; Avci, G. Simple reaction time to the onset of time-varying sounds. *Atten. Percept. Psychophys.* **2017**, *79*, 628–636.
13. ISO 532-1, Acoustics-Methods for calculating loudness-Part 1: Zwicker method, Geneva: International Organization for Standardization, 2017.
14. Florentine, M.; Buus, S.; Poulsen, T. Temporal integration of loudness as a function of level. *J. Acoust. Soc. Am.* **1996**, *99*, 1633–1644.
15. 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*, 669–680.
16. Shimokura, R.; Soeta, Y. Estimation of reaction time for birdsongs and effects of background noise and listener's age. *Appl. Acoust.* **2022**, *194*, 1–11 (108785).
17. 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.
18. Numba, S.; Kuwano, S.; Fastl, H. Loudness of non-steady-state sound. *Jpn. Psychol. Res.* **2008**, *50*, 154–166.
19. Munson, W.A. The growth of auditory sensation. *J. Acoust. Soc. Am.* **1947**, *19*, 584–591.
20. Florentine, M.; Buus, S.; Poulsen, T. Temporal integration of loudness as a function of level. *J. Acoust. Soc. Am.* **1996**, *99*, 1633–1644.
21. Buus, S.; Florentine, M.; Poulsen, T. Temporal integration of loudness, loudness discrimination, and the form of the loudness function. *J. Acoust. Soc. Am.* **1997**, *101*, 669–680.
22. Glasberg, B.R.; Moore, B.C. A model of loudness applicable to time-varying sounds. *J. Acoust. Soc. Am.* **2002**, *50*, 331–342.
23. Miller, J.; Ulrich, R. Simple reaction time and statistical facilitation: A parallel grains model. *Cogn. Psychol.* **2003**, *46*, 101–151.
24. Plomp, R.; Bouman, M.A. Relationship between hearing threshold and duration for tone pulse. *J. Acoust. Soc. Am.* **1959**, *31*, 749–758.
25. Zwislöcki, J.J. Theory of temporal auditory summation. *J. Acoust. Soc. Am.* **1960**, *32*, 1046–1060.
26. Zwislöcki, J.J. Temporal summation of loudness: An analysis. *J. Acoust. Soc. Am.* **1969**, *46*, 431–441.
27. Poulsen, T. Loudness of tone pulses in a free field. *J. Acoust. Soc. Am.* **1981**, *69*, 1786–1790.
28. Hots, J.; Rennie, J.; Verhey, J.L. Influence of time constants and comparison on the prediction of temporal integration of loudness. *Proc. Conference on Acoustics AIA-DAGA.* **2013**, 1266–1268.
29. Heil, P.; Neubauer, H. A unifying basis of auditory thresholds based on temporal summation. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 6151–6156.
30. Heil, P.; Matysiak, A.; Neubauer, H. A probabilistic Poisson-based model accounts for an extensive set of absolute auditory threshold measurement. *Hear. Res.* **2017**, *353*, 135–161.
31. Jaramillo, F.; Wiesenfeld, K. Mechano-electrical transduction assisted by Brownian motion: a role for noise in the auditory system. *Nat. Neurosci.* **1998**, *1*, 384–388.
32. Henry, K.R. Noise improves transfer of near-threshold, phase-locked activity of the cochlear nerve: evidence for stochastic resonance? *J. Comp. Physiol. A.* **1999**, *184*, 577–584.
33. Zeng, F.G.; Fu, Q.J.; Morse, R. Human hearing enhanced by noise. *Brain. Res.* **2000**, *869*, 251–255.

34. Moss, F.; Ward, L.M.; Sannita, W.G. Stochastic resonance and sensory information processing: a tutorial and review of application. *Clin. Neurophysiol.* **2004**, *115*, 267–281.
35. Ries, D.T. The influence of noise type and level upon stochastic resonance in human audition. *Hear. Res.* **2007**, *228*, 136–143.
36. Ward, L.M.; MacLean, S.E.; Kirschner, A. Stochastic resonance modulates neural synchronization within and between cortical sources. *PLoS ONE* **2010**, *5*, e14371.
37. Yerkes, R.M.; Dodson, J.D. The relation of strength of stimulus to rapidity of habit-formation. *J. Comp. Neurol. Psychol.* **1908**, *18*, 459–482.
38. Broadbent, D.E. A reformulation of the Yerkes-Dodson law. *Br. J. Math. Stat. Psychol.* **1965**, *18*, 145–157.
39. Mendl, M. Performing under pressure: Stress and cognitive function. *Appl. Anim. Behav. Sci.* **1999**, *65*, 221–244.
40. Heeren, W.; Hohmann, V.; Appell, J.E.; Verhey, J.L. Relation between loudness in categorical units and loudness in phons and sones. *J. Acoust. Soc. Am.* **2013**, *133*, EL314–319.
41. <https://jp.mathworks.com/help/audio/ref/gammatonefilterbank-system-object.html> (accessible in 19th February 2026).
42. Moore, B.C.J.; Glasberg, B.R. Derivation of auditory filter shapes from notched-noise data. *Hear. Res.* **1990**, *47*, 103–138.
43. Hartmann, W.M. *Signals, Sound, and Sensation*. Springer Science & Business Media. p. 251, 2004.
44. Levitt, H. Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. Am.* **1971**, *49*, 467–477.
45. Garner, W.R.; Miller, G.A. The masked threshold of pure tones as a function of duration. *J. Exp. Psychol.* **1947**, *37*, 293–303.
46. Hawkins, J.E.; Stevens, S.S. The masking of pure tones and of speech by white noise. *J. Acoust. Soc. Am.* **1950**, *22*, 6–13.
47. Wagner, E.; Florentine, M.; Buus, S.; McCormack, J. Spectral loudness summation and simple reaction time. *J. Acoust. Soc. Am.* **2004**, *116*, 1681–1686.
48. Raab, D.H.; Grossberg, M. Reaction time to changes in the intensity of white noise. *J. Exp. Psychol.* **1965**, *69*, 609–612.
49. Kohfeld, D.L.; Goedecke, D.W. Intensity and predictability of background noise as determinants of simple reaction time. *Bulletin of the Psychonomic Society* **1978**, *12*, 129–132.
50. Glasberg, B.R.; Moore, B.C.J. Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments. *J. Acoust. Soc. Am.* **1986**, *79*, 1020–1033.
51. 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*, 8858, 12 pages.
52. Shimokura, R.; Soeta, Y. Characteristics of train noise in above-ground and underground stations with side and island platforms. *J. Sound Vib.* **2011**, *330*, 1621–1633.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.