Speech Communication 134 (2021) 1–11

Contents lists available at ScienceDirect

Speech Communication

journal homepage: www.elsevier.com/locate/specom

Effect of prior exposure on the perception of Japanese vowel length contrast in reverberation for nonnative listeners

Eri Osawa [a,*], C.T. Justine Hui [b], Yusuke Hioka [b], Takayuki Arai [a]

[a] Department of Information and Communication Sciences, Sophia University, Tokyo, 102-8544, Japan
[b] Acoustics Research Centre, Department of Mechanical Engineering, University of Auckland, Auckland, 1142, New Zealand

ARTICLE INFO

Keywords:
Speech perception
Reverberation
Japanese length contrast
Nonnative listener

ABSTRACT

While reverberation often degrades speech intelligibility, previous studies have shown that prior exposure to reverberation can reduce its adverse effects on speech perception. The current study investigated the effect of prior exposure to reverberation on the perception of nonnative speech sounds. We compared the results from two experiments, one in "blocked presentation" where the target words with the same amount of reverberation were presented to participants consistently, and another in "random presentation" where the amount of reverberation added to the target words changed between each trial. A Japanese minimal pair,/ie/ 'house' -/iie/ 'no', where vowel length creates a phonemic difference, was used as the target. The results for native listeners showed that their responses did not differ significantly between the blocked and random presentations. On the other hand, the effect of presentation type was significant in terms of the responses from the nonnative listeners. The results showed that nonnative listeners did not respond differently between the anechoic and reverberant conditions in the blocked presentation. However, there was a significant difference between the anechoic and reverberant conditions in the random presentation. The results from the nonnative listeners suggest that they try to obtain information of reverberation from the exposure since they could not use top-down processing effectively as much as native listeners.

1. Introduction

We often have conversations under noisy and reverberant environments in our day-to-day life. Researchers have investigated the effects of reverberation on speech perception for many decades. However, there are still open questions about how we perceive speech in adverse environments.

In an indoor environment, a sound usually travels from a speaker to a listener via multiple transmission paths, due to reflections on various surfaces in the environment. Apart from the direct sound traveling directly from the speaker to the listener, early reflections, which are defined as reflections reaching the listener within 50 ms after the direct sound arrives, are known to increase speech intelligibility by reinforcing the direct sound (Bradley et al., 2003). On the other hand, late reverberation, which arrives later than the early reflections as "tails" of the direct sound, is known to decrease speech intelligibility. This late reverberation, especially when it is long, causes difficulties for listeners to identify speech sounds correctly (Knudsen, 1929; Gelfand and Silman, 1979; Nábělek et al., 1989; Hodoshima et al., 2006; Osawa et al., 2018). Bolt and MacDonald (1949) discussed two effects of reverberation on speech intelligibility: self-masking and overlap-masking effects. The self-masking effect describes temporal distortion of the sound itself caused by late reverberation making the sound being perceived as longer than its original length. The overlap-masking effect describes masking of the following sounds by the tail of the current sound caused by its reverberation. Nábělek et al. (1989) investigated the effects of reverberation on the intelligibility of English consonants by examining the identification of English consonants in /-ɑt/ and /s-ɑt/ contexts. The results showed that listeners misheard consonants in the /s-ɑt/ context more frequently than the /-ɑt/ context. Since errors unique to the /s-ɑt/ context were observed, the study suggested that coarticulatory cues between the preceding /s/ and the target consonants reduced the intelligibility on top of overlap-masking and self-masking effects. Osawa et al. (2020) reported an effect of self-masking on speech perception using Japanese vowel length contrast in the word-final position. The results showed that "long vowel" responses increased in the reverberant condition since native Japanese listeners use vowel duration as the primary cue to distinguish between short and long vowels. Since the offset of the target vowel in the word-final position became unclear in the reverberant condition because of the

---

[*] Corresponding author.
E-mail addresses: eri1989.16.11@gmail.com (E. Osawa), yusuke.hioka@ieee.org (Y. Hioka), arai@sophia.ac.jp (T. Arai).

https://doi.org/10.1016/j.specom.2021.07.009
Received 5 October 2020; Received in revised form 29 June 2021; Accepted 22 July 2021

self-masking effect, native Japanese listeners hesitated to determine the choice between short and long vowels in the reverberant condition.

Despite such detrimental effects of reverberation, we usually can converse without too much difficulty in reverberant environments. Our auditory system can reduce the detrimental effects of reverberation. Previous studies have shown that binaural listening can reduce the negative effects of reverberation (Nabelek and Pickett, 1974; Gelfand and Hochberg, 1976; Nábělek and Robinson, 1982). Gelfand and Hochberg (1976) showed both normal-hearing and hearing impaired listeners benefit from binaural listening, and Nábělek and Robinson (1982) also reported that both young and elderly listeners benefit from binaural hearing.

Another mechanism to overcome the effects of reverberation is perceptual adaptation to listening environments. Previous studies have shown that listeners can adapt their perception to an environment by being exposed to the environment (Watkins, 2005; Brandewie and Zahorik, 2010; Srinivasan and Zahorik, 2013; Zahorik and Anderson, 2013). Watkins (2005) examined the effect of a carrier phrase preceding the target word on the distinction between /sir/ and /stir/ in reverberation. The results showed that the number of responses /stir/ increased when the same amount of reverberation was added to both target word and the carrier phrase. This suggests that since listeners were able to obtain the characteristics of the reverberation inherent to the environment from the carrier phrase, they managed to identify /t/ overlapped by /s/ in the reverberant environment. Brandewie and Zahorik (2010) examined the effect of the presence of a carrier sentence preceding the target word. The results showed that there was improvement in speech intelligibility by prior exposure to reverberation. These studies suggest that we may be able to learn the characteristics of reverberant environments from the prior exposure allowing listeners to find perceptual cues more easily.

This adaptation is also observed in the perception of non-speech sounds in reverberation. When listeners identify material types (e.g., wood and metal) from their impact sounds in a reverberant environment, prior exposure to the reverberation of the environment helps identification of the material types. Koumura and Furukawa (2017) showed that presentation context affected material type identification in reverberation. They examined the difference in material type identification between two presentation contexts. One was called "blocked condition" where the stimuli with the same amount of reverberation were presented to participants consistently. The other was called "random condition" where the amount of reverberation in the stimuli changed between each trial. The results indicated that listeners could find acoustic characteristics of each materials (i.e., wood, metal, and glass) better in the blocked condition than in the random condition. This indicated that prior exposure to the acoustics of the environment helped listeners with retrieving information regarding the stimuli in reverberation.

While the effect of prior exposure to reverberation on the perception of native speech has been understood reasonably well, to the best of our knowledge, the effect on perception of nonnative speech sounds has not been investigated. Listeners frequently mishear nonnative speech sounds in reverberation (Nábělek and Donahue, 1984; Takata and Nábělek, 1990; Masuda, 2016; Osawa et al., 2018, 2020). In addition, even though nonnative listeners could identify phonemes of the target language correctly under non-reverberant condition, they were not able to identify the phonemes as accurately as native listeners under the reverberant conditions (Nábělek and Donahue, 1984; Takata and Nábělek, 1990). These previous studies suggest nonnative listeners would experience more difficulty in understanding an utterance in reverberation compared to native listeners. Hence, we hypothesize that nonnative listeners in reverberation would struggle with understanding an utterance regardless of their prior exposure to the reverberation.

In the current study, we investigated whether prior exposure to reverberation can reduce the detrimental effects of reverberation on nonnative speech perception. We conducted an experiment using two different presentation types to investigate the effect of prior exposure: blocked and random presentations. As the target stimuli, we chose to examine Japanese vowel length contrast where temporal information was crucial to differentiating between the contrasts. Previous studies on adaptation to a reverberant environment (Watkins, 2005; Brandewie and Zahorik, 2010; Srinivasan and Zahorik, 2013; Zahorik and Anderson, 2013) have not investigated whether prior exposure to reverberation would allow listeners to estimate the original length of a speech sound distorted by reverberation. In addition, nonnative Japanese listeners often have difficulties in acquiring Japanese vowel length contrast. The previous studies (Nábělek and Donahue, 1984; Takata and Nábělek, 1990; Masuda, 2016; Osawa et al., 2018, 2020) have shown that nonnative Japanese listeners often have difficulties in listening to some structure which does not exist in their native languages in reverberation. Therefore, by using Japanese vowel length contrast, the current study examined how prior exposure to reverberation would affect the perception of speech sound duration in reverberation and whether the exposure would help nonnative listeners compensate for distortion of durational feature caused by reverberation. The experiment was conducted by recruiting nonnative Japanese listeners who have learned or are currently learning Japanese as the participants, and native Japanese listeners as the control group.

2. Method

2.1. Participants

Ten nonnative Japanese listeners (five male and five female) participated in the experiment. Their age ranged from 19 to 36 with the mean age being 25.5 years. Their native languages were English, Mandarin, Cantonese, Korean, and Thai. Since they all lived in Auckland, New Zealand when they participated in the experiment, they did not use Japanese in their daily lives. They all had taken formal Japanese classes in high school or university. Their mean length of learning Japanese was 5.4 years (SD = 2.7 years).

Eleven native Japanese listeners (four male and seven female) also participated in the experiment as the control group. Their age ranged from 31 to 49 with the mean age being 37.1 years. They all also lived in Auckland, New Zealand when they participated in the experiment.

All participants reported that they did not have any hearing disorder.

2.2. Stimuli

The target was a minimal pair of Japanese words differing in vowel length, /ie/ 'house' and /iie/ 'no'. Vowels have the steady-state region which has a certain duration while consonants are transient and shorter than vowels in general. Listeners often have difficulty in estimating the original length of a vowel in reverberation since the vowel energy often causes longer reverberation, and it is difficult to find the offset of the steady-state region of a vowel prolonged by reverberation. Therefore, reverberation affects perceived duration of vowels more than consonants. Even native Japanese listeners had a difficulty in distinguishing between short and long vowels in the word-final position in reverberation while they could distinguish between singleton and geminate consonants in the reverberant condition as well as in the non-reverberant condition (Osawa et al., 2020). In the current study, we selected the vowel length contrast in the word-medial position as the target since (Osawa et al., 2020) showed that even native listeners could not distinguish the vowel length contrast in the word-final position categorically in a reverberant condition. We selected the minimal pair since they are familiar words even for nonnative listeners. Learners of Japanese would learn these words in the beginning of an introductory course in general. According to the NTT psycholinguistic databases (Amano and Kondo, 1999), the word

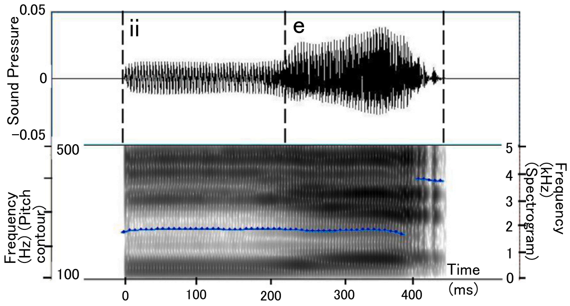

Fig. 1. Speech waveform and pitch contour of the longest speech sample (i.e., step 11). The vertical axis on the left side of the bottom figure shows the frequency range of the pitch contour. That on the right side shows the frequency range of the spectrogram. The horizontal axis indicates time (ms).

familiarity rating for /ie/ is 6.4 and that for /iie/ is 6.0 out of 7, with 7 being “most familiar word” and 1 being “least familiar word”.

The stimuli used varied in the duration of the vowel along the durational continuum, where “ie” and “iie” were on the two extreme. To create the durational continuum of /ie/ - /iie/, the duration of /ii/ in /iie/ was reduced stepwise using Praat (Boersma, 2001). The vowel duration was reduced by two or three glottal cycle steps to make the time interval between steps in the continuum consistent. The intervals between steps ranged from 8 to 12 ms with the mean being 10.9 ms. Eleven synthetic speech samples were generated along a durational continuum of /ie/ - /iie/. The vowel length of the longest speech sample of the continuum (i.e., step 11) was 220 ms whereas that of the shortest speech sample (i.e., step 1) was 120 ms.

To avoid drastic pitch change after reducing the vowel duration, the token was produced as a word without pitch rising. Although /iie/ is a word with an accent pattern of LHH (low-to-high-to-high) generally, it is frequently produced as a word without a pitch rising (e.g., HHH). Fig. 1 shows the speech waveform and pitch contour of the longest speech sample (i.e., step 11). The line in the bottom of Fig. 1 shows the pitch contour. The flat pitch region of /ii/ was reduced to create the durational continuum.

The token was spoken by one of the authors who is a speaker of Tokyo standard Japanese in her early 30’s. The recording was conducted in a sound-proof room at Sophia University, Japan. The token was digitally recorded at a sampling rate of 48 kHz with 24 bits/sample using a Marantz digital recorder (PMD661MK II).

In the current study, reverberant stimuli were rendered by a sound reproduction system (details are discussed in Section 2.3) which was able to replicate acoustics of various environments. An Ambisonics technique was adopted for the sound reproduction, which was realized by generating the stimuli using a first-order Ambisonics microphone array (Rode NT-SF1). Room impulse responses (RIRs) were measured in various rooms using the microphone array by playing a swept sine signal from a loudspeaker (Genelec 8020D), and the measured RIRs were convolved with the signal of the speech samples to create the stimuli, i.e., speech samples with reverberation. The RIRs were measured under three different reverberant conditions, summarized in Table 1. As the values of the reverberation time (RT) suggest, Room 2a and 2b were more reverberant environment than Room 1. The RIRs of Room 2a and 2b were measured in the same room but at different distances between the source and the microphone (the microphone array was located at a farther distance from the loudspeaker in Room 2b than in Room 2a) in the room. This is reflected in the RTs of Room 2a and 2b being close to each other, whereas, C50s of Room 2b is smaller than that of Room 2a. The signal-to-overlap-masking ratio (SOR) (Arai et al., 2007) indicates the effect of the tail of preceding sound caused by reverberation was more significant in Room 2 (Room 2b in particular) than in Room 1.

Table 1
The reverberation time (RT) (s) , clarity (C50) (dB), signal-to-overlap-masking ratio (SOR) (dB) (Arai et al., 2007) for each reverberant condition.

	RT (s)	C50 (dB)	SOR (dB)
Room 1	0.44	3.87	−1.77
Room 2a	0.83	4.53	−2.38
Room 2b	0.89	1.92	−2.58

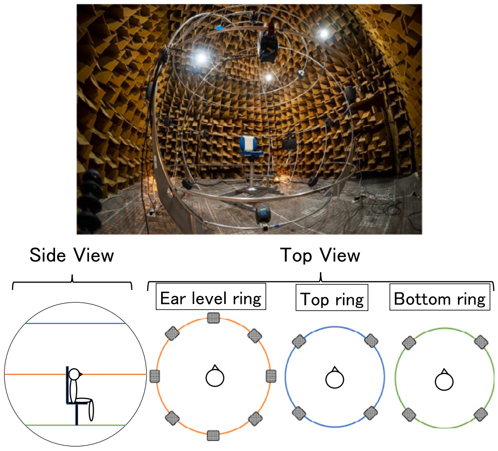

Fig. 2. Sound reproduction system. The bottom figure shows the placement of the loudspeakers.

Participant No.: 1

Trial 0

Fig. 3. Graphical user interface.

2.3. Sound reproduction system

The experiments under both the anechoic and reverberant conditions were conducted using the sound reproduction system built in the anechoic chamber of the Acoustics Laboratory at the University of Auckland, New Zealand as shown in Fig. 2. The sound reproduction system consists of a 16 channel loudspeaker (Genelec 8020D) array in a spherical configuration wherein the participant was seated at the center of the sphere. Of the 16 loudspeakers, eight of them were placed along a circle located at the height of participant’s ear, four of them were placed along a circle at 45 degrees elevation above the participant’s ear level, and another four were placed along a circle at 45 degrees below the participant’s ear level, as shown in Fig. 2. All loudspeakers were facing towards the center of the sphere (i.e. participant’s position).

All stimuli were stored in a PC, and were presented via an audio interface (MOTU 16 A) using a digital audio workstation (Reaper). Stimuli from the anechoic condition were presented only from the loudspeaker located in front of the participant’s seat (0 degree elevation) whereas stimuli of the reverberant conditions were presented through all 16 loudspeakers using a commercially available Ambisonics decoder (Harpex Berge and Barrett, 2010). We adjusted the sound pressure level of the stimuli to 53.7 ± 1 dBA.

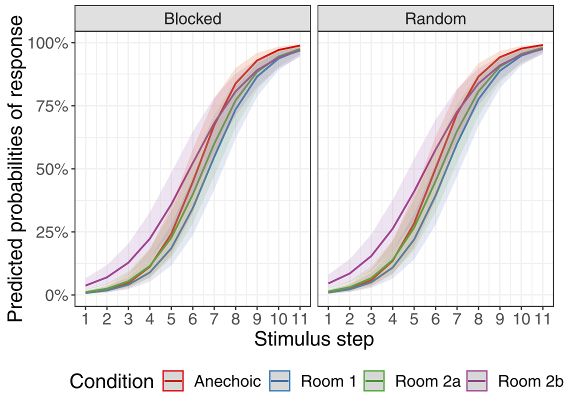

Fig. 4. Predicted probability of "long vowel" responses as a function of stimulus steps for native listeners under anechoic (red line), Room 1 (blue line), Room 2a (green line), Room 2b (purple line) conditions. The left figure shows the results of the blocked presentation and the right figure shows the results of the random presentation. The shaded areas indicate the 95% confidence interval. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.4. Procedure

We used two presentation types in the current study. One was a "blocked presentation" where the target words with the same amount of reverberation were presented to participants consistently where the three different reverberant conditions (i.e., Room 1, 2a, and 2b) and the anechoic condition were conducted separately. Since the participants were exposed to the same amount of reverberation for several minutes, we regard the blocked presentation as the condition which allows participants to have prior exposure to the reverberant conditions. The other presentation type was a "random presentation" where the reverberant condition of the target words changed between each trial. That is, the participants listened to the stimuli with four different conditions (i.e., Room 1, 2a, 2b, and the anechoic conditions) randomly. Since the participants were not exposed to a particular reverberant condition until the stimulus was presented, we regard the random presentation as the condition without prior exposure. There were five test sessions (i.e., the four conditions in the blocked presentation, and the random presentation). The participants could take a break anytime they felt fatigue.

A graphical user interface (GUI) for Reaper was developed in Python for controlling the experiment as well as allowing the participants to enter their responses. The participants were asked to identify whether the presented word included a short or long vowel without thinking too much. We asked the participants not to think too long after they listened to the stimuli. However, we did not require them to make speeded responses. The next stimulus would not be played until the participants submitted their current response. They were asked to give their response by pressing the labeled button on the GUI using a mouse. The target words in romaji, which is representations of Japanese sounds in Roman letters, were displayed on the button. There was also a "submit" button on the GUI. Their responses were not submitted until they pressed the submit button. Therefore, if the participants pressed a wrong button accidentally, they could correct their answer. Fig. 3 shows the GUI for the experiment. The button allowed participants to confirm whether their responses were their final answer. Since each stimulus was presented to the participants only five times, we created the submit button to prevent the participants from making errors. There was no feedback to their responses in the test session.

In total, 440 trials (11 interval steps × 4 acoustic conditions × 2 presentation types × 5 repetitions) were conducted for each participant. In addition, the participants carried out a practice session to familiarize themselves with the procedure prior to the start of the test. The practice session was for confirming the procedure, and to show the participants how to submit their responses, where the stimuli would come after they had submitted their responses. The participants listened to five stimuli and gave responses as they would in the test session. There was no feedback on their responses in the practice session as well.

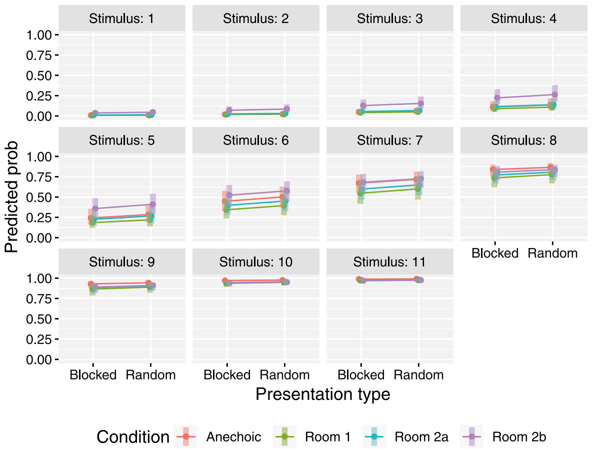

Fig. 5. Predicted probability of "long vowel" responses to each stimulus under the anechoic (red line), Room 1 (green line), Room 2a (blue line) and Room 2b (purple line) conditions for native listeners. The horizontal axis shows the presentation types (i.e., the blocked and random presentations). The error bars indicate the 95% confidence interval. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.5. Statistical methods

We used a generalized linear mixed model (GLMM) to model the perceptual responses. For the modeling, we used the *lme4* package in R Studio (R Core Team, 2019). A GLMM was built for fitting of the perception of each listener group using *glmer* function of the *lme4* package. A random factor `participant`, and fixed factors of `stimuli`, `acoustic condition`, and `presentation type` were used in the model. The dependent variable was the response to the stimuli (i.e., /ie/ or /iie/).

In addition, we conducted post-hoc tests using the *emmeans* package in R Studio to examine any significant interactions, where *p*-value were adjusted using the Tukey method.

3. Results

3.1. Native listeners

Fig. 4 shows the distinction between /ie/ and /iie/ for native listeners in the blocked and random presentations. The vertical axis of the plot indicates the predicted probability of "long vowel" responses in percentage as a function of stimulus steps in the horizontal axis for the anechoic (red line), Room 1 (blue line), Room 2a (green line), and Room 2b (purple line) conditions.

From the statistical analysis, we found the interaction between stimulus step and acoustic condition to be significant ($\chi^2 = 23.008$, df $=$ 10, $p < 0.0001$). Table 2 shows the results of the post-hoc comparisons of "long vowel" response log-odds ratio among acoustic conditions. The multiple comparisons were corrected by Tukey method. Similarly, Fig. 5 shows the predicted probability of "long vowel" responses to each stimulus under each acoustic condition in the blocked and random presentations. The results of the post-hoc comparisons showed that there were significant differences between the anechoic and some of the reverberant conditions in the predicted probability of "long vowel" responses to the longer stimuli (i.e., from step 8 to step 11). Native

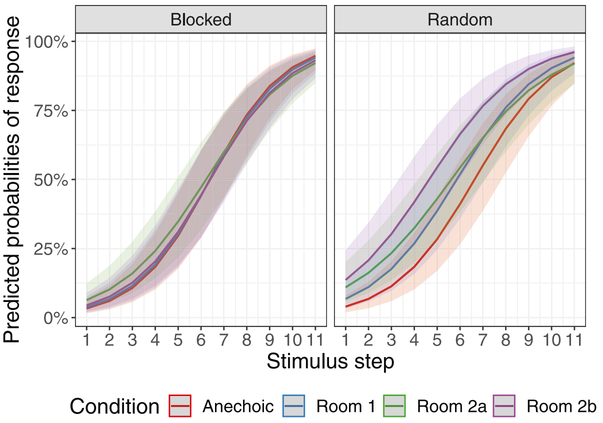

Fig. 6. Predicted probability of "long vowel" responses as a function of stimulus steps for nonnative listeners under anechoic (red line), Room 1 (blue line), Room 2a (green line), and Room 2b (purple line) conditions. The left figure shows the results for the blocked presentation and the right figure shows the results for the random presentation. The shaded areas indicate the 95% confidence interval. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

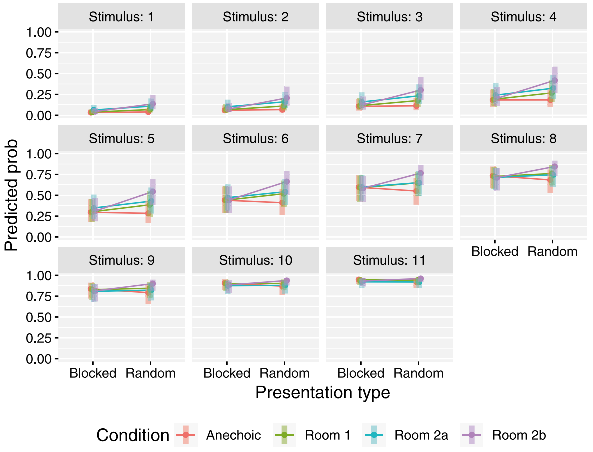

Fig. 7. Predicted probability of "long vowel" responses to each stimulus under the anechoic (red line), Room 1 (green line), Room 2a (blue line) and Room 2b (purple line) conditions for nonnative listeners. The horizontal axis shows the presentation types (i.e., the blocked and random presentations). The error bars indicate the 95% confidence interval. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

listeners gave less "long vowel" responses to the longer stimuli in the reverberant conditions than in the anechoic condition. In terms of the identification of the shorter stimuli (i.e., from step 1 to step 5), there were significant differences between the Room 2b condition and the others. Native listeners gave more "long vowel" responses to the shorter stimuli in the Room 2b condition than in the other conditions.

The interactions between stimulus step and presentation type and between acoustic condition and presentation type were not significant [(χ^2 = 1.07, df = 11, p = 0.3008), (χ^2 = 3.6597, df = 13, p = 0.3006)]. In addition, there was no significant interaction between stimulus step, acoustic condition, and presentation type (χ^2 = 6.0262, df = 14, p = 0.1972).

The results showed that there was no significant difference between the blocked and random presentations for native listeners.

3.2. Nonnative listeners

Fig. 6 shows the distinction between /ie/ and /iie/ for nonnative listeners in the blocked and random presentations. Fig. 6 indicates the predicted probability of "long vowel" responses as a function of stimulus steps for anechoic (red line), Room 1 (blue line), Room 2a (green line), and Room 2b (purple line) conditions.

We found the interaction between stimulus step and acoustic condition (χ^2 = 8.5251, df = 10, p = 0.03632), between stimulus step and presentation type (χ^2 = 5.7815, df = 11, p = 0.0162), and between acoustic condition and presentation type (χ^2 = 20.561, df = 14, $p < 0.001$) to be significant. Table 3 shows the results of post-hoc comparisons of "long vowel" response log-odds ratio among acoustic conditions in the random presentation. The multiple comparisons were corrected by Tukey method. There was no contrast which was significant in the blocked presentation. Similarly, Fig. 7 shows the predicted probability of "long vowel" responses to each stimulus under each acoustic condition in the blocked and random presentations.

The results showed that there was always a significant difference between the Room 2b and some of the other acoustic conditions in the predicted probability of "long vowel" responses in the random presentation. Nonnative listeners gave more "long vowel" responses in the Room 2b condition than the other acoustic conditions when the amount of reverberation in the target changed between each trial although their perceptual performance under all acoustic conditions did not differ from each other in the blocked presentation.

We conducted another post-hoc test to investigate the difference between the blocked and random presentations in terms of the participants' responses. Table 4 shows the acoustic conditions where there was a significant difference between the blocked and random presentations in the predicted probability of "long vowel" responses. The results revealed that there was a significant difference between the blocked and random presentations in distinguishing all the stimuli under the Room 2b condition.

There was no significant interaction between stimulus step, acoustic condition, and presentation type (χ^2 = 6.2169, df = 17, p = 0.1015). In contrast to native listeners, the results showed that there was a significant difference between the blocked and random presentations for nonnative listeners, i.e., the results showed the effect of prior exposure on the distinction for nonnative listeners.

4. Discussion

4.1. Native listeners

The results showed that there was no significant difference between the blocked and random presentations in the distinction between /ie/ and /iie/ for the native Japanese listeners. This could be interpreted that native listeners could adapt their perception to an environment by listening to the whole target word with reverberation. There is a possibility that native listeners could adapt their perception to an environment and determine the choice during the process that they listened to /i/, /e/ overlapped by /i/, and /e/.

The results also suggest that the native listeners could use top-down processing effectively to distinguish the stimuli. We usually estimate what we would hear from language experience and knowledge when we have a conversation in adverse environments. The reverberant conditions used in the current study were common environments in our daily lives. Language experience in reverberant environments for the native listeners would help fast speech processing in the reverberant conditions.

Zsiga (2012) suggested that the effect of top-down processing could be understood that we have word-level prototypes for speech processing. We would use the word-level prototype of a certain word as a reference to identify a target. Regarding Japanese vowel length contrast, native listeners would use information regarding the length of the target vowel and the adjacent phonemes within a word to compare the stimuli with the prototype since native Japanese listeners use the duration of adjacent phonemes besides the duration of the target vowel to distinguish between short and long vowels (Fujisaki and Sugito,

Table 2
Pairwise comparisons of “long vowel” response log-odds ratio among acoustic conditions (Anechoic, Room 1, Room 2a, and Room 2b) by stimulus for native listeners.

Stimuli	Contrast	Estimate	SE	z.ratio	p.value	Stimuli	Contrast	Estimate	SE	z.ratio	p.value
1	Anechoic - Room 1	−0.03	0.36	−0.07	0.9999	7	Anechoic - Room 1	0.53	0.14	3.88	**0.0006**
	Anechoic - Room 2a	−0.37	0.35	−1.07	0.7077		Anechoic - Room 2a	0.32	0.14	2.31	0.0955
	Anechoic - Room 2b	−1.6	0.31	−5.1	**<.0001**		Anechoic - Room 2b	−0.04	0.14	−0.31	0.99
	Room 1 - Room 2a	−0.34	0.34	−1.01	0.7423		Room 1 - Room 2a	−0.21	0.13	−1.65	0.3498
	Room 1 - Room 2b	−1.57	0.31	−5.11	**<.0001**		Room 1 - Room 2b	−0.57	0.13	−4.46	**<.0001**
	Room 2a - Room 2b	−1.23	0.3	−4.17	**0.0002**		Room 2a - Room 2b	−0.36	0.13	−2.78	**0.0277**
2	Anechoic - Room 1	0.07	0.3	0.22	0.9961	8	Anechoic - Room 1	0.62	0.17	3.72	**0.0011**
	Anechoic - Room 2a	−0.26	0.29	−0.88	0.8132		Anechoic - Room 2a	0.43	0.17	2.55	0.0533
	Anechoic - Room 2b	−1.34	0.26	−5.12	**<.0001**		Anechoic - Room 2b	0.22	0.17	1.29	0.5673
	Room 1 - Room 2a	−0.32	0.29	−1.12	0.6753		Room 1 - Room 2a	−0.19	0.15	−1.24	0.602
	Room 1 - Room 2b	−1.41	0.26	−5.42	**<.0001**		Room 1 - Room 2b	−0.4	0.15	−2.63	0.0422
	Room 2a - Room 2b	−1.08	0.25	−4.38	**0.0001**		Room 2a - Room 2b	−0.21	0.16	−1.37	0.5204
3	Anechoic - Room 1	0.16	0.24	0.65	0.9151	9	Anechoic - Room 1	0.71	0.21	3.36	**0.0043**
	Anechoic - Room 2a	−0.14	0.24	−0.6	0.9319		Anechoic - Room 2a	0.54	0.22	2.54	0.0545
	Anechoic - Room 2b	−1.08	0.21	−5.08	**<.0001**		Anechoic - Room 2b	0.48	0.21	2.25	0.1095
	Room 1 - Room 2a	−0.3	0.24	−1.27	0.5808		Room 1 - Room 2a	−0.17	0.19	−0.87	0.8204
	Room 1 - Room 2b	−1.24	0.21	−5.8	**<.0001**		Room 1 - Room 2b	−0.24	0.19	−1.24	0.6029
	Room 2a - Room 2b	−0.94	0.2	−4.62	**<.0001**		Room 2a - Room 2b	−0.07	0.19	−0.35	0.9855
4	Anechoic - Room 1	0.25	0.19	1.3	0.5653	10	Anechoic - Room 1	0.81	0.27	3.04	**0.0126**
	Anechoic - Room 2a	−0.03	0.19	−0.15	0.9989		Anechoic - Room 2a	0.66	0.27	2.47	0.0649
	Anechoic - Room 2b	−0.82	0.17	−4.86	**<.0001**		Anechoic - Room 2b	0.74	0.26	2.82	**0.0245**
	Room 1 - Room 2a	−0.28	0.19	−1.47	0.4556		Room 1 - Room 2a	−0.15	0.24	−0.61	0.9299
	Room 1 - Room 2b	−1.07	0.17	−6.22	**<.0001**		Room 1 - Room 2b	−0.07	0.23	−0.29	0.9914
	Room 2a - Room 2b	−0.79	0.16	−4.85	**<.0001**		Room 2a -Room 2b	0.08	0.24	0.33	0.9877
5	Anechoic - Room 1	0.34	0.15	2.25	0.1109	11	Anechoic - Room 1	0.9	0.32	2.8	**0.0265**
	Anechoic - Room 2a	0.09	0.15	0.59	0.9347		Anechoic - Room 2a	0.77	0.32	2.4	0.0776
	Anechoic - Room 2b	−0.56	0.14	−4.12	**0.0002**		Anechoic - Room 2b	1	0.31	3.18	**0.0079**
	Room 1 - Room 2a	−0.26	0.15	−1.7	0.3222		Room 1 - Room 2a	−0.12	0.29	−0.43	0.9742
	Room 1 - Room 2b	−0.9	0.14	−6.48	**<.0001**		Room 1 - Room 2b	0.1	0.28	0.35	0.985
	Room 2a - Room 2b	−0.65	0.13	−4.85	**<.0001**		Room 2a - Room 2b	0.22	0.28	0.79	0.8603
6	Anechoic - Room 1	0.44	0.13	3.34	**0.0046**						
	Anechoic - Room 2a	0.2	0.13	1.57	0.394						
	Anechoic - Room 2b	−0.3	0.12	−2.44	0.0691						
	Room 1 - Room 2a	−0.23	0.13	−1.84	0.2547						
	Room 1 - Room 2b	−0.74	0.12	−6.01	**<.0001**						
	Room 2a - Room 2b	−0.5	0.12	−4.19	**0.0002**						

1977). There is a possibility that the native listeners may process the tails added by reverberation and a phoneme overlapped by the tail of the preceding phoneme by referring to the prototype. The native listeners may be able to estimate the length of the tail of each phoneme in the target word, and estimate the original form of the target word by using information about the length in the prototype, for example, the ratio of the length of the target vowel to the adjacent phonemes in the word.

The difference between the results of the current study and the previous studies suggests that the effect of prior exposure might depend on perceptual cues. The previous studies examined the identification of consonants (Watkins, 2005) or sentences (Brandewie and Zahorik, 2010; Srinivasan and Zahorik, 2013) while the current study investigated perception of the length of a segment. Prior exposure would be effective for the participants to become able to perceive some characteristics of speech signals which were distorted by reverberation. On the other hand, listeners might have difficulties in changing the prediction of the original length of a segment which was distorted by reverberation by exposure to reverberation.

Some factors would have affected the effect of prior exposure in distinguishing between the vowel length contrast in reverberation. Brandewie and Zahorik (2011) reported that there was no effect of prior exposure to reverberation in their study. They examined the effect of a carrier sentence preceding the target on speech intelligibility in reverberation using the modified rhyme test. The results showed that the effect of prior exposure to reverberation was observed only for a portion of the participants. They discussed that listeners could obtain additional information from the anechoic stimuli. Since the anechoic and reverberant stimuli without a carrier sentence were presented randomly in the condition without prior exposure, listeners could use the anechoic stimuli as a reference to identify speech sounds with reverberation. They suggested that there would be no difference between responses with and without prior exposure if the anechoic stimuli were included in the stimuli presented to the listeners randomly. In the current study, since the anechoic stimuli were also included in the stimuli of the random presentation, the anechoic stimuli might have helped the listeners with distinguishing the reverberant stimuli.

Another factor that could explain why there was no difference in performance between blocked and random presentation is the language background of the participants. Watkins (2005), which examined the effect of a carrier sentence on the distinction between /sir/ and /stir/ in reverberation, showed that listeners could estimate the amount of reverberation and find /t/ despite /s/ overlapping /t/ by hearing a carrier sentence preceding to the target. In Watkins (2005), the participant group consisted of native and nonnative English listeners, and the nonnative listeners spoke English fluently. In the current study, we analyzed the data for native and nonnative listeners separately. Previous studies have shown that nonnative listeners were not able to identify target phonemes as accurately as native listeners under reverberant conditions, even though the nonnative listeners could identify the phonemes correctly under the non-reverberant condition (Nábělek and Donahue, 1984; Takata and Nábělek, 1990). This implies there is a possibility that the ability to compensate degraded speech by reverberation between native and nonnative listeners is different even if the nonnative listeners are able to speak the target language fluently.

Our results from the native listeners shows that the predicted probability of “long vowel” responses to the shortest stimulus (i.e., step 1) differed between the Room 2b condition and the other acoustic conditions, where the participants identified the shortest stimulus as a long vowel word more frequently in the Room 2b condition than

Table 3
Pairwise comparisons of "long vowel" response log-odds ratio among acoustic conditions (Anechoic, Room 1, Room 2a, and Room 2b) by stimulus in the random presentation for nonnative listeners.

Stimuli	Contrast	Estimate	SE	z.ratio	p.value	Stimuli	Contrast	Estimate	SE	z.ratio	p.value
1	Anechoic - Room 1	−0.56	0.28	−2.04	0.1734	7	Anechoic - Room 1	−0.41	0.17	−2.47	0.0654
	Anechoic - Room 2a	−1.09	0.27	−4.11	**0.0002**		Anechoic - Room 2a	−0.42	0.16	−2.57	**0.0497**
	Anechoic - Room 2b	−1.34	0.27	−5.06	**<.0001**		Anechoic - Room 2b	−0.98	0.17	−5.71	**<.0001**
	Room 1 - Room 2a	−0.53	0.25	−2.10	0.1533		Room 1 - Room 2a	−0.01	0.16	−0.06	0.9999
	Room 1 - Room 2b	−0.78	0.25	−3.11	**0.0103**		Room 1 - Room 2b	−0.57	0.17	−3.32	**0.0050**
	Room 2a - Room 2b	−0.25	0.24	−1.05	0.7211		Room 2a - Room 2b	−0.56	0.17	−3.32	**0.0050**
2	Anechoic - Room 1	−0.54	0.24	−2.23	0.1155	8	Anechoic - Room 1	−0.38	0.18	−2.10	0.1526
	Anechoic - Room 2a	−0.98	0.23	−4.21	**0.0001**		Anechoic - Room 2a	−0.31	0.18	−1.72	0.3136
	Anechoic - Room 2b	−1.28	0.23	−5.53	**<.0001**		Anechoic - Room 2b	−0.92	0.19	−4.84	**<.0001**
	Room 1 - Room 2a	−0.44	0.22	−2.01	0.1858		Room 1 - Room 2a	0.08	0.18	0.42	0.9750
	Room 1 - Room 2b	−0.74	0.22	−3.40	**0.0038**		Room 1 - Room 2b	−0.54	0.19	−2.79	**0.0274**
	Room 2a - Room 2b	−0.30	0.21	−1.45	0.4713		Room 2a - Room 2b	−0.61	0.19	−3.25	**0.0064**
3	Anechoic - Room 1	−0.51	0.21	−2.44	0.0702	9	Anechoic - Room 1	−0.36	0.21	−1.73	0.3101
	Anechoic - Room 2a	−0.87	0.20	−4.27	**0.0001**		Anechoic - Room 2a	−0.20	0.20	−0.97	0.7684
	Anechoic - Room 2b	−1.22	0.20	−6.02	**<.0001**		Anechoic - Room 2b	−0.86	0.22	−3.98	**0.0004**
	Room 1 - Room 2a	−0.35	0.19	−1.84	0.2536		Room 1 - Room 2a	0.16	0.21	0.79	0.8604
	Room 1 - Room 2b	−0.71	0.19	−3.69	**0.0013**		Room 1 - Room 2b	−0.50	0.22	−2.29	0.1011
	Room 2a - Room 2b	−0.35	0.18	−1.92	0.2184		Room 2a - Room 2b	−0.67	0.22	−3.10	**0.0106**
4	Anechoic - Room 1	−0.49	0.19	−2.64	**0.0416**	10	Anechoic - Room 1	−0.33	0.24	−1.40	0.5016
	Anechoic - Room 2a	−0.76	0.18	−4.21	**0.0001**		Anechoic - Room 2a	−0.08	0.23	−0.36	0.9836
	Anechoic - Room 2b	−1.16	0.18	−6.45	**<.0001**		Anechoic - Room 2b	−0.80	0.25	−3.24	**0.0066**
	Room 1 - Room 2a	−0.27	0.17	−1.57	0.3970		Room 1 - Room 2a	0.25	0.24	1.05	0.7182
	Room 1 - Room 2b	−0.67	0.17	−3.93	**0.0005**		Room 1 - Room 2b	−0.47	0.25	−1.86	0.2441
	Room 2a - Room 2b	−0.41	0.17	−2.45	0.0683		Room 2a - Room 2b	−0.72	0.25	−2.93	**0.0180**
5	Anechoic - Room 1	−0.46	0.17	−2.76	**0.0297**	11	Anechoic - Room 1	−0.31	0.27	−1.13	0.6733
	Anechoic - Room 2a	−0.64	0.16	−3.94	**0.0005**		Anechoic - Room 2a	0.03	0.27	0.10	0.9996
	Anechoic - Room 2b	−1.10	0.17	−6.63	**<.0001**		Anechoic - Room 2b	−0.74	0.28	−2.63	**0.0425**
	Room 1 - Room 2a	−0.18	0.16	−1.15	0.6574		Room 1 - Room 2a	0.34	0.27	1.24	0.5985
	Room 1 - Room 2b	−0.64	0.16	−3.98	**0.0004**		Room 1 - Room 2b	−0.43	0.29	−1.52	0.4266
	Room 2a - Room 2b	−0.46	0.16	−2.93	**0.0180**		Room 2a - Room 2b	−0.77	0.28	−2.77	**0.0290**
6	Anechoic - Room 1	−0.44	0.16	−2.71	**0.0339**						
	Anechoic - Room 2a	−0.53	0.16	−3.37	**0.0041**						
	Anechoic - Room 2b	−1.04	0.16	−6.38	**<.0001**						
	Room 1 - Room 2a	−0.10	0.16	−0.62	0.9270						
	Room 1 - Room 2b	−0.61	0.16	−3.76	**0.0010**						
	Room 2a - Room 2b	−0.51	0.16	−3.23	**0.0067**						

Table 4
Pairwise comparisons of "long vowel" log-odds ratio between the blocked and random presentations by stimulus and acoustic condition for nonnative listeners.

Stimuli	Condition	Estimate	SE	z.ratio	p.value	Stimuli	Condition	Estimate	SE	z.ratio	p.value
1	Anechoic	0.19	0.23	−0.83	0.4051	7	Anechoic	0.18	0.17	1.11	0.2693
	Room 1	0.62	0.22	−2.81	**0.005**		Room 1	0.25	0.17	−1.51	0.132
	Room 2a	0.60	0.21	−2.80	**0.0051**		Room 2a	0.22	0.16	−1.41	0.1593
	Room 2b	1.22	0.22	−5.67	**<.0001**		Room 2b	0.85	0.17	−5.07	**<.0001**
2	Anechoic	0.13	0.21	−0.61	0.54	8	Anechoic	0.25	0.17	1.42	0.1557
	Room 1	0.56	0.20	−2.77	**0.0056**		Room 1	0.19	0.17	−1.08	0.282
	Room 2a	0.53	0.19	−2.76	**0.0057**		Room 2a	0.16	0.17	−0.96	0.3375
	Room 2b	1.16	0.20	−5.89	**<.0001**		Room 2b	0.79	0.18	−4.44	**<.0001**
3	Anechoic	0.07	0.19	−0.34	0.7325	9	Anechoic	0.31	0.19	1.67	0.0957
	Room 1	0.50	0.19	−2.68	**0.0074**		Room 1	0.12	0.19	−0.67	0.5044
	Room 2a	0.47	0.18	−2.67	**0.0076**		Room 2a	0.10	0.18	−0.54	0.5864
	Room 2b	1.10	0.18	−6.04	**<.0001**		Room 2b	0.73	0.19	−3.79	**0.0001**
4	Anechoic	0.00	0.18	−0.02	0.9861	10	Anechoic	0.37	0.20	1.85	0.0649
	Room 1	0.44	0.17	−2.52	**0.0118**		Room 1	0.06	0.20	−0.31	0.7594
	Room 2a	0.41	0.16	−2.49	**0.0126**		Room 2a	0.04	0.20	−0.19	0.8534
	Room 2b	1.04	0.17	−6.08	**<.0001**		Room 2b	0.66	0.21	−3.81	**0.0015**
5	Anechoic	0.06	0.17	0.35	0.7256	11	Anechoic	0.43	0.22	1.97	**0.0486**
	Room 1	0.37	0.17	−2.26	**0.0237**		Room 1	0.00	0.22	0.00	0.9996
	Room 2a	0.35	0.16	−2.22	**0.0265**		Room 2a	0.03	0.22	0.12	0.9086
	Room 2b	0.97	0.16	−5.94	**<.0001**		Room 2b	0.60	0.23	−2.63	**0.0087**
6	Anechoic	0.12	0.16	0.74	0.4609						
	Room 1	0.31	0.16	−1.92	0.0553						
	Room 2a	0.29	0.16	−1.85	0.0651						
	Room 2b	0.91	0.16	−5.59	**<.0001**						

the others. This suggests that RT is not the only parameter to predict an effect on perceived duration in reverberant environments. Table 1 shows the signal-to-overlap-masking ratio (SOR) (Arai et al., 2007) which indicates the amount of the tail of the prolonged /i/. Since the amount of /i/ overlapping the following /e/ of the Room 2b condition was the largest of the acoustic conditions, listeners might have had

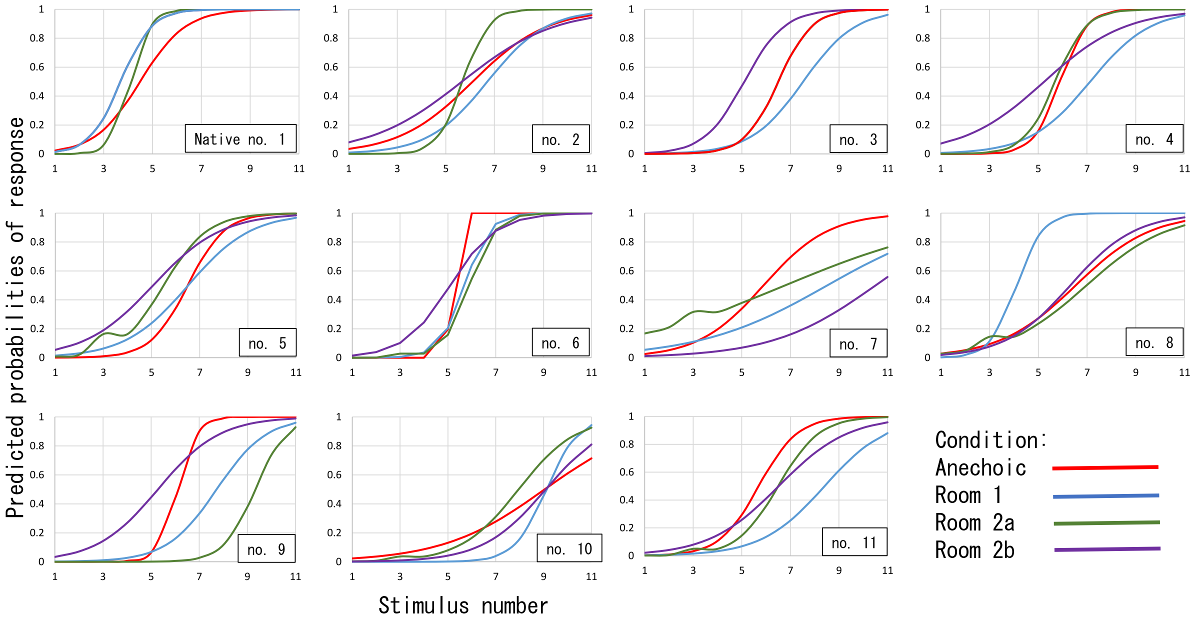

Fig. A.1. Individual performance for the native participants in the blocked presentation.

difficulty in finding the offset of /i/, i.e., the boundary between /i/ and /e/, in the Room 2b condition.

4.2. Nonnative listeners

In contrast to the native listeners, the results for nonnative listeners showed that prior exposure changed their distinction between /ie/ and /iie/ in the reverberant conditions. In the blocked presentation, distinction for nonnative listeners in the reverberant conditions did not differ significantly from that in the anechoic condition. This indicated that there was no effect of reverberation on their perception in the blocked presentation. The nonnative listeners would have tried to obtain information regarding the reverberation during the test session using bottom-up processing. They would have tried to compensate their perception with information from prior exposure since they could not use top-down processing as effectively as the native listeners. Nonnative listeners may have less experiences with various Japanese phonetics and phonology which would help listeners with understanding an utterance degraded by reverberation than native listeners. In addition, perceptual cues which nonnative listeners have learned is considered to be less robust than that of native listeners. Thus, nonnative listeners could not use the knowledge effectively as reference to understand an utterance degraded by reverberation. One of the nonnative participants mentioned that she could distinguish the stimuli better in the latter half of a session in the blocked presentation. Although there was no significant difference between early and late stimuli in the blocked presentation in the distinction of the stimuli, we observed a slight difference in their identification of the shortest stimulus. Early in the block, the nonnative listeners often identified the shortest stimulus as "long vowel". On the other hand, late in the block, that is, after exposure to reverberation, the predicted probability of "long vowel" responses to the shortest stimulus decreased. This suggests that nonnative listeners tended to identify the shortest stimulus, which was prolonged by reverberation, as "long vowel", because they might distinguish the length contrast depending on the perceived duration in the reverberant conditions early in the block. On the other hand, late in the block, the nonnative listeners could estimate the original duration of the target vowel from the prior exposure. Therefore, the nonnative listeners identified the shortest stimulus as "short vowel" frequently late in the block. This suggests that the nonnative listeners could learn the acoustics of the environment from exposure to reverberation. A stimulus word might be too short to obtain sufficient information to distinguish the nonnative length contrast under the reverberant conditions. In addition, they might not to be able to adapt their perception to the reverberation which changed between each trial.

In the random presentation, the distinction of the reverberant stimuli differed from that of the anechoic stimuli. There is a possibility that the preceding and following stimuli affected the distinction of a stimulus. In the experiment, the stimuli were presented in isolation, and they were presented immediately after the participants gave their response. Thus, the intervals between each trial were short. Watkins (2005) showed that listeners rarely able to recognize degraded speech sounds correctly when the target word was presented with different amount of reverberation from that of the carrier phrases. Since the preceding and the following stimuli might have similar effect to that of the carrier phrases used in Watkins (2005), the nonnative listeners might have been confused. In contrast to the native listeners who could use the anechoic stimuli as a reference in the random presentation, nonnative listeners might not have been able to obtain additional information from the anechoic stimuli.

The results showed that there was no difference in the distinction of the anechoic stimuli between the blocked and random presentations for nonnative listeners. According to the Native Language Magnet (NLM) theory developed by Kuhl and Iverson (1995), a prototype of each category, which is regarded as a good example of a category, pulls in stimuli surrounding the prototype (i.e., magnet effect). Pisoni and Lively (1995) suggested that a category consists of many exemplars and a prototype could be regarded as an average of the exemplars. Therefore, we need various inputs to build a robust prototype of a category. Osawa et al. (2020) suggested that nonnative listeners might follow the opposite process to native listeners to build a prototype. In the process of learning a nonnative language, learners first learn a good example (i.e., the prototype) of nonnative phonemes in the pronunciation of a teacher or teaching materials. Since learners might initially acquire prototypes with little exposure to various pronunciations, they have not learned which differences they can disregard between the prototype and candidates. In the current study, nonnative listeners were able to distinguish the anechoic stimuli in the random presentation as well as in the blocked presentation. This indicated that they were able to distinguish the anechoic stimuli without prior exposure as well as with prior exposure. This suggests that they are able to identify the stimuli immediately if they are not degraded by reverberation. Since

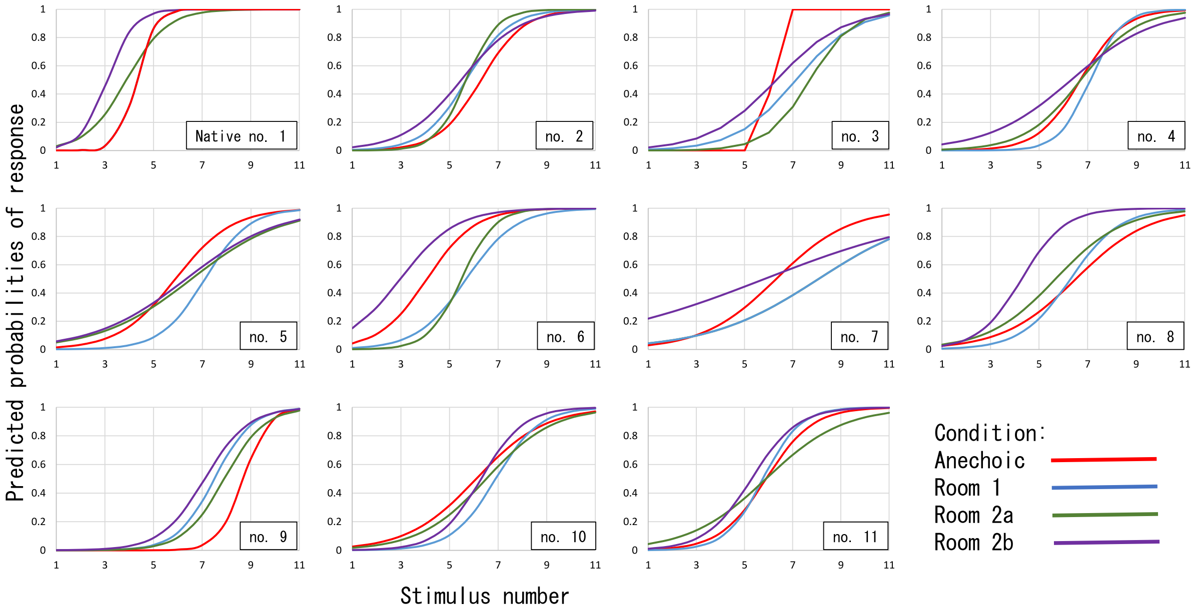

Fig. A.2. Individual performance for the native participants in the random presentation.

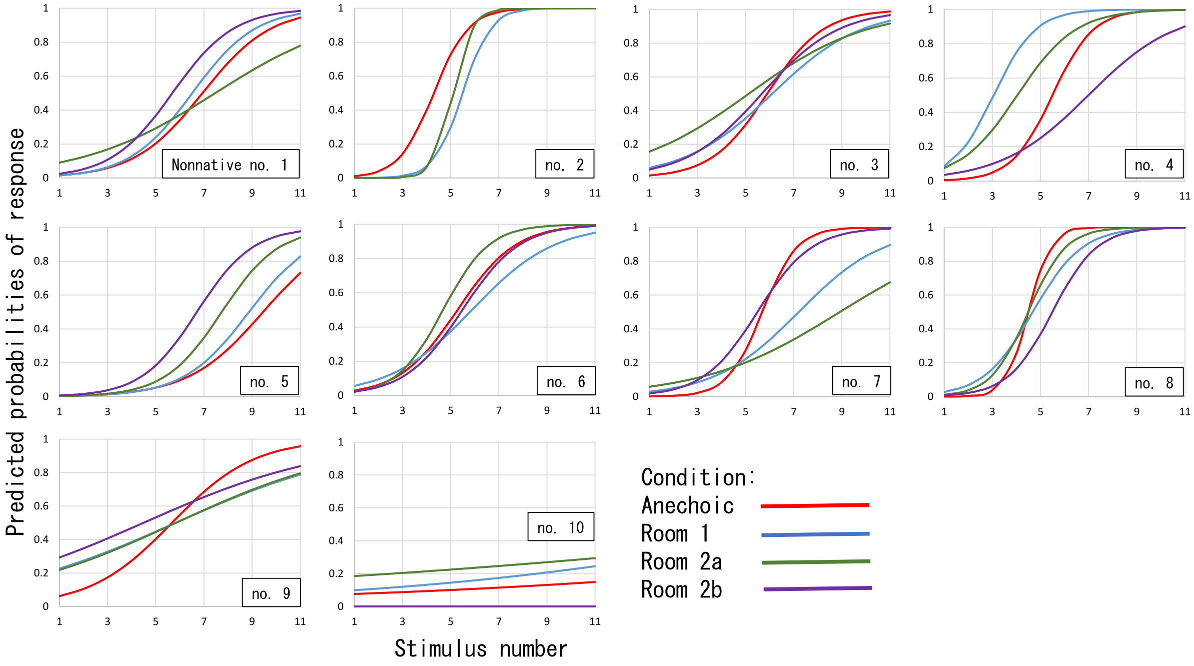

Fig. A.3. Individual performance for the nonnative participants in the blocked presentation.

the anechoic stimuli were the sounds closest to the prototype in the current setting, they were able to identify the stimuli immediately. On the other hand, the reverberant stimuli were not familiar sounds for the nonnative listeners. Thus, they might not be able to determine whether they could disregard the differences between the prototypes and the reverberant stimuli.

5. Conclusion

In the current study, we investigated the effect of prior exposure on the distinction of Japanese vowel length contrast with reverberation for nonnative listeners. The results for native listeners showed that their responses did not differ significantly between the blocked and random presentations. On the other hand, the results for nonnative listeners showed that there was a significant difference between the anechoic and reverberant conditions in the random presentation but not in the blocked presentation. The results for nonnative listeners indicated that there was an effect of prior exposure on distinguishing between vowel lengths, suggesting an attempt to obtain information of the environment from prior exposure. The results suggest that there is a difference between native and nonnative listeners in time required to complete adaptation of their perception to environments.

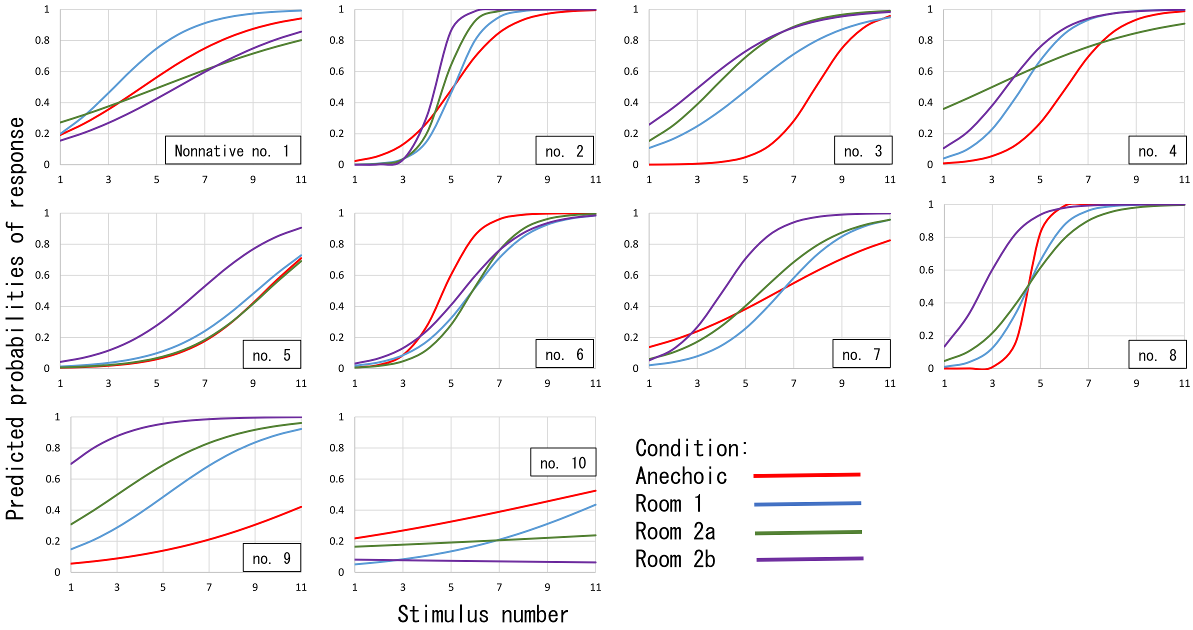

Fig. A.4. Individual performance for the nonnative participants in the random presentation.

CRediT authorship contribution statement

Eri Osawa: Conception and design of study, Acquisition of data, Analysis and/or interpretation of data, Writing – original draft, Writing – review & editing. **C.T. Justine Hui:** Conception and design of study, Acquisition of data, Analysis and/or interpretation of data, Writing – original draft, Writing – review & editing. **Yusuke Hioka:** Acquisition of data, Analysis and/or interpretation of data, Writing – original draft, Writing – review & editing. **Takayuki Arai:** Conception and design of study, Analysis and/or interpretation of data, Writing – original draft, Writing – review & editing.

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 work reported in this paper.

Acknowledgments

This research was supported in part by the Japan Society for the Promotion of Science (JSPS) KAKENHI (Grant number: 19K23083) and Sophia University Special Grant for Academic Research. All authors approved the version of the manuscript to be published.

Appendix. Individual performance

See Figs. A.1–A.4.

References

Amano, S., Kondo, T., 1999. NTT database series, Nihongo-no Goitokusei: Lexical properties of Japanese. Sanseido, Tokyo.

Arai, T., Murakami, Y., Hayashi, N., Hodoshima, N., Kurisu, K., 2007. Inverse correlation of intelligibility of speech in reverberation with the amount of overlap-masking. Acoust. Sci. Technol. 28 (6), 438–441.

Berge, S., Barrett, N., 2010. High angular resolution planewave expansion. In: Proc. of the 2nd International Symposium on Ambisonics and Spherical Acoustics May, pp. 6–7.

Boersma, P., 2001. Praat, a system for doing phonetics by computer. Glot. Int. 5 (9), 341–345.

Bolt, R., MacDonald, A., 1949. Theory of speech masking by reverberation. J. Acoust. Soc. Am. 21 (6), 577–580.

Bradley, J.S., Sato, H., Picard, M., 2003. On the importance of early reflections for speech in rooms. J. Acoust. Soc. Am. 113 (6), 3233–3244.

Brandewie, E., Zahorik, P., 2010. Prior listening in rooms improves speech intelligibility. J. Acoust. Soc. Am. 128 (1), 291–299.

Brandewie, E., Zahorik, P., 2011. Adaptation to room acoustics using the modified rhyme test. In: Proceedings of Meetings on Acoustics 161ASA. Acoustical Society of America, 050007.

Fujisaki, H., Sugito, M., 1977. Onsei no butsuriteki seishitsu (physical characteristics of speech sounds. Iwanami Koza Nihongo 5: On'in 63–106.

Gelfand, S., Hochberg, I., 1976. Binaural and monaural speech discrimination under reverberation. Audiology 15 (1), 72–84.

Gelfand, S.A., Silman, S., 1979. Effects of small room reverberation upon the recognition of some consonant features. J. Acoust. Soc. Am. 66 (1), 22–29.

Hodoshima, N., Arai, T., Kusumoto, A., Kinoshita, K., 2006. Improving syllable identification by a preprocessing method reducing overlap-masking in reverberant environments. J. Acoust. Soc. Am. 119 (6), 4055–4064.

Knudsen, V.O., 1929. The hearing of speech in auditoriums. J. Acoust. Soc. Am. 1 (1), 56–82.

Koumura, T., Furukawa, S., 2017. Context-dependent effect of reverberation on material perception from impact sound. Sci. Rep. 7 (1), 1–13.

Kuhl, P.K., Iverson, P., 1995. Linguistic experience and the, perceptual magnet effect. In: Strange, W. (Ed.), Speech Perception and Linguistic Experience: Issues in Cross-Language Research. pp. 121–154.

Masuda, H., 2016. Misperception patterns of American english consonants by Japanese listeners in reverberant and noisy environments. Speech Commun. 79, 74–87.

Nábělek, A.K., Donahue, A.M., 1984. Perception of consonants in reverberation by native and non-native listeners. J. Acoust. Soc. Am. 75 (2), 632–634.

Nábělek, A., Letowski, T., Tucker, F., 1989. Reverberant overlap- and self-masking in consonant identification. J. Acoust. Soc. Am. 86 (4), 1259–1265.

Nabelek, A.K., Pickett, J.M., 1974. Monaural and binaural speech perception through hearing aids under noise and reverberation with normal and hearing-impaired listeners. J. Speech Hear. Res. 17 (4), 724–739.

Nábělek, A.K., Robinson, P.K., 1982. Monaural and binaural speech perception in reverberation for listeners of various ages. J. Acoust. Soc. Am. 71 (5), 1242–1248.

Osawa, E., Arai, T., Hodoshima, N., 2018. Perception of Japanese consonant-vowel syllables in reverberation: Comparing non-native listeners with native listeners. Acoust. Sci. Technol. 39 (6), 369–378.

Osawa, E., Arai, T., Hodoshima, N., 2020. Perception of Japanese length contrasts with reverberation by native and nonnative listeners. Acoust. Sci. Technol. 41 (5), 751–760.

Pisoni, D.B., Lively, S.E., 1995. Variability and invariance in speech perception: A new look at some old problems in perceptual learning. Speech Perception and Linguistic Experience: Issues in Cross-Language Research 433–459.

R Core Team, 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL: https://www.R-project.org/.

Srinivasan, N.K., Zahorik, P., 2013. Prior listening exposure to a reverberant room improves open-set intelligibility of high-variability sentences. J. Acoust. Soc. Am. 133 (1), EL33–EL39.

Takata, Y., Nábělek, A.K., 1990. English consonant recognition in noise and in reverberation by Japanese and American listeners. J. Acoust. Soc. Am. 88 (2), 663–666.

Watkins, A.J., 2005. Perceptual compensation for effects of reverberation in speech identification. J. Acoust. Soc. Am. 118 (1), 249–262.

Zahorik, P., Anderson, P.W., 2013. Amplitude modulation detection by human listeners in reverberant sound fields: Effects of prior listening exposure. In: Proceedings of Meetings on Acoustics ICA2013. Acoustical Society of America, 050139.

Zsiga, E.C., 2012. The Sounds of Language: An Introduction to Phonetics and Phonology. John Wiley & Sons.