Perception of Japanese length contrasts with reverberation by native and nonnative listeners

Eri Osawa1,†, Takayuki Arai1,‡ and Nao Hodoshima2
1Sophia University, 7–1 Kioi-cho, Chiyoda-ku, Tokyo, 102–8554 Japan
2Tokai University, 2–3–23 Takanawa, Minato-ku, Tokyo, 108–8619 Japan
(Received 13 September 2019, Accepted for publication 13 April 2020)

Abstract: The perception of segmental duration is crucial for the distinction of Japanese length contrasts. However, the perceived duration may be changed in a long reverberation, which adds a “tail” to sounds, making them perceived as being longer. In addition, since lengthened sounds overlap the following sounds, the boundaries of phonemes would become blurred. In the current study, we investigated whether the effects of reverberation distort the distinction of Japanese length contrasts for native Japanese and English listeners. Stimuli were nonword pairs (/baba/–/babaa/, /ata/–/atta/, and /ama/–/amma/) varying in duration along the continuum. The logistic function was used to model the perception. In the distinction of vowel length contrast in the word-final position, even native listeners identified the stimulus with the shortest vowel duration as a long vowel word with reverberation. Regarding the perception of the geminate nasal, “geminate” responses increased with reverberation for native listeners, whereas the results for nonnative listeners indicated that “singleton” responses increased with reverberation. It is assumed that the difference could be attributed to the different prototypes of categories of Japanese between native and nonnative listeners. In addition, the results for nonnative listeners might be attributed to the difference in prosody between English and Japanese.

Keywords: Speech perception, Reverberation, Length contrast, Japanese, Nonnative listener

PACS number: 43.71.Hw, 43.66.Dc, 43.55.Hy [doi:10.1250/ast.41.751]

1. INTRODUCTION

1.1. Japanese Length Contrasts

Temporal structure is essential for comprehending utterances in Japanese. Vowel and consonant lengths make a phonemic difference in Japanese. The primary cue for native Japanese listeners to distinguish between short and long phonemes is the segmental duration. The longer the segmental duration is, the more likely native listeners identify it as a long phoneme, i.e., a long vowel and a geminate consonant. Previous studies have revealed that native Japanese listeners distinguish between short and long phonemes categorically depending on relative duration [4–6]. There is no absolute duration value to identify a long phoneme since segmental duration could change depending on speaking rate [4,5]. Therefore, carrier sentences provide native Japanese listeners with a clue for making a distinction. Hirata and Lambacher [7] examined the effect of a carrier sentence on the distinction of the length contrasts. They showed that native Japanese listeners could not distinguish stimuli accurately when stimuli and a carrier sentence were spoken at different speaking rates from each other. The result indicated that native listeners used not only the segmental duration of a target but also the duration of the phonemes in the context the target was in.

Regarding the contextual effects, previous studies have shown that the length of adjacent vowels affects the distinction between singleton and geminate consonants [e.g., 8,9]. Ofuka et al. [8] indicated that native Japanese listeners tended to identify a consonant as a geminate when the preceding vowel was longer, whereas Kingston et al. [9] showed the opposite results. The effect of the length of adjacent vowels is controversial among studies.

Nonnative listeners often have difficulty in distinguishing Japanese length contrasts. Although “long” responses increase as a function of segmental duration as perceived by native listeners, there is a difference in category

*The current paper includes the results for native listeners in Arai [1,2]. The parts of the results for non-native listeners in the current paper were preliminarily published [3].
†e-mail: eri1989.16.11@gmail.com
‡e-mail: arai@sophia.ac.jp
boundary value between native and nonnative listeners. Kato et al. [10] investigated the categorical perception of Japanese vowel length contrast for nonnative listeners (native English listeners). The study [10] revealed that there was a significant difference in the boundary position of fitted logistic curves between native and nonnative listeners and there was also a significant difference in the boundary width between listener groups. The results of the study [10] indicated that nonnative listeners could distinguish between short and long vowels more sharply after perceptual training, although the change was not statistically significant.

Native listeners distinguish between short and long vowels, taking into account the speaking rate, whereas nonnative listeners are disadvantaged in adjusting their perception to speaking rates when they are not trained with materials spoken at various speaking rates [10–12]. Hirata et al. [12] examined the effects of training on the vowel length contrasts at various speaking rates. The results showed that native English listeners could adapt to various speaking rates after training. The results suggested that nonnative listeners tended to identify short or long phonemes without taking speaking rates into account when they are not trained.

### 1.2. Effects of Reverberation

We are exposed to reverberation at all times in our daily lives. In a closed room, the walls of the room reflect the sound from the sound source repeatedly, and the energy of the sound decreases after each reflection. Sound energy decay caused by multiple reflections is called reverberation. Reverberation time (RT) is often used as the indicator of reverberation. In places where RTs are short, since reflections reach a listener almost simultaneously with the direct sound, speech intelligibility increases [13]. On the other hand, in places where RTs are long, since late reflections add “tails” to the direct sound, a listener perceives the sound as being longer than the actual duration. In addition, since the tails overlap the direct sound, speech intelligibility is reduced.

Previous studies have shown that a listener experiences difficulty in identifying speech sounds accurately in reverberation [14–16]. Bolt and MacDonald [17] discussed two effects of reverberation on speech intelligibility. One is overlap-masking, where the energy of a preceding phone overlaps the following phone owing to the energy from reverberation. The other is self-masking, where reverberation smears the sound energy along the time axis within each speech sound. Bolt and MacDonald [17] assumed that self-masking was relatively unimportant because the self-masking effect could be reduced by speaking slowly. Nakata et al. [18] also showed that speech intelligibility increased at a low speaking rate with reverberation.

Nonnative listeners have more difficulty in understanding speech sounds accurately with reverberation than native listeners. Previous studies [19–21] have shown that nonnative listeners misheard phonemes more often than native listeners under the reverberant condition, even though the performance difference was small between native and nonnative listeners under the nonreverberant condition. Nábělek and Donahue [19] showed that nonnative English listeners who spoke English fluently performed poorer in English consonant identification under the reverberant condition using the modified rhyme test (MRT) (see [22] for details of MRT). Stimuli were recorded with RTs of 0.4, 0.8, and 1.2 s. The test environments were selected to resemble typical real-life environments. The results indicated that nonnative listeners had difficulty in identifying phonemes even with moderate reverberation. In addition, Nábělek and Donahue [19] suggested that the phoneme system in the native languages of nonnative listeners affected the error patterns when there was reverberation. The results showed that Japanese and Chinese confused English liquid consonants with glide consonants more frequently than Polish listeners. Takata and Nábělek [20] also observed the same tendency as the results of Nábělek and Donahue [19].

Masuda [21] also examined the intelligibility of 23 English consonants under various reverberant conditions for native English and native Japanese listeners. Although the results confirmed the main effect of listening conditions on the overall identification rate, there was no effect of language group. The interaction between the listening condition and the native language was not significant. However, the results confirmed the main effect of language group on the identification rate of the subset of consonants /f, d, ɹ, l, r, s, ʃ, th, ð, v, z/ that Japanese listeners often had difficulty in identifying. The results of the subset of consonants also indicated that native Japanese listeners might have a disadvantage in identifying English consonants to which there are no Japanese counterparts in the presence of reverberation.

### 1.3. Aims of the Current Paper

Although many studies have been conducted on the perception of the length contrasts [e.g., 4, 6], because they were mostly conducted under quiet conditions, the results may not have reflected real-life environments where reverberation may be present. Since native Japanese listeners distinguish length contrasts on the basis of segmental duration, their distinction may be affected by the self-masking and overlap-masking effects of reverberation. Segmental duration would be lengthened by the self-masking effect, and the overlap-masking effect would make the boundaries between phonemes unclear with reverberation. The effects may cause a listener to have
difficulty in perceiving the duration of the target vowel or consonant accurately with reverberation.

The current study focuses on the effects of reverberation on the distinction of Japanese length contrasts, and the perceptions of native Japanese listeners and nonnative listeners are compared. Nonnative listeners, even advanced learners, cannot distinguish the length contrasts as sharply as native Japanese listeners [e.g., 10], and reverberation affects the perception of nonnative listeners more than that of native listeners [e.g., 19]. It should be noted that in the current study, native English listeners, whose native language does not have length contrasts, were recruited for the experiments. Therefore, since reverberation might distort the distinction of nonnative listeners more than that of native listeners, the performance difference between listener groups might be larger under the reverberant condition than under the nonreverberant condition.

To compare the categorical perception of native listeners with that of nonnative listeners, we used a logistic function to model listeners’ perception as used in previous studies [6, 10, 23]. In a plot of the logistic function, the slope of a fitted curve indicates the steepness of distinction. The category boundary is estimated as the 50% crossover point of the response rate on the logistic function. The proportion of responses to the extreme stimuli (i.e., the longest and shortest stimuli) is regarded as the distinction of typical short and long phonemes.

2. EXPERIMENT

2.1. Stimuli

The stimuli were pairs of nonsense Japanese words differing in vowel length or consonant constriction length. These words varied in the duration of certain vowels or certain consonants along the durational continuum. The word pairs used in the experiment are as follows.

/baba/–/babaa/
/ata/–/atta/
/ama/–/amma/

For the vowel length contrast, we selected the word pair whose length contrast was in the word-final position since there is always a pitch change between the first and second mora in the Tokyo dialect of Japanese, and the pitch change could be a perceptual cue for distinction. The targets for examining the distinction of consonant length contrasts were the voiceless plosive (/t/) and nasal (/m/) embedded in /a____a/ context to standardize the overlap-masking effect on each target word. The accent pattern of /babaa/ was high-to-low and the accent pattern of /baba/ was high-to-low-to-low.

To create durational continua of /babaa/, the duration of the second vowel of /babaa/ was reduced stepwise. The durational continua of consonant’s contrasts were created by reducing the durations of closure and nasal murmur of /atta/ and /amma/, respectively. The original closure duration of /t/ was shortened in 12 ms steps, and the durations of the vowel and nasal murmur were reduced by one-glottal-cycle steps. There was a pitch fall in the shorter stimuli after reducing the durations of the vowel and nasal murmur. Two of the authors judged the stimuli with the pitch fall to be acceptable since two mora words with an accent pattern of high-to-low were expected to have a pitch drop between the first and second mora. Ten synthetic stimuli were generated along a durational continuum of /ata/–/atta/, and 11 synthetic stimuli were generated along a durational continuum of /babaa/ and /amma/. The durations of the vowel and consonant of the extreme stimuli on the continuum (i.e., the shortest and longest stimuli) are shown in Table 1.

The target words were presented in the carrier sentence, “Korekara kikoete kurunowa ______ desu” (You will hear _____ in the experiments. All stimuli shared the same token of the carrier sentence to standardize the overlap-masking effect on each target word. When the target words were recorded, all of them were spoken in a sentence consisting of the same words as in the carrier sentence. The target words were then extracted and embedded in the carrier sentence, which was recorded separately for shared use as the standard carrier sentence. We selected a shared carrier sentence that was pronounced clearly and whose intensity fluctuation was small. The duration of the carrier phrase preceding the target word was 1,563.1 ms, and the duration of the carrier phrase following the target word was 496.0 ms. Since the fluctuations of intensity in the long vowel word and the geminate consonant word were small, the original intensity of the tokens was used as stimuli.

An adult male speaker of Tokyo standard Japanese in his early 20s spoke the target words and the carrier sentence. The recording was conducted in a sound-proof room. The tokens were digitally recorded at a sampling rate of 44.1 kHz (16 bits) with a linear pulse code modulation (PCM) recorder, Sony PCM-D50.

2.2. Participants

Two groups participated in the experiments. The nonnative listener group consisted of ten native English
listeners whose mean age was 22.8 years ($SD = 2.3$ years). Three were male and seven were female. There were eight Americans, one British, and one Australian. Their mean age of arrival in Japan was 22.2 years ($SD = 2.4$ years). They were all students who learned Japanese at universities in Japan. In the current study, nonnative listeners whose lengths of residence (LORs) in Japan were similar were recruited for the experiments. It is often a problem for learners that nonnative speech sounds in real-life environments differ from those they listen to in a classroom. Masuda [21] regarded LOR as an index of the amount of language experience that helps when listening to nonnative speakers. 

The mean LOR of participants in Japan was 7.2 months ($SD = 3.2$ months, range $= 2–11$ months). 

The native group, the control group, consisted of eleven native Japanese listeners whose mean age was 23.5 years ($SD = 3.1$ years). Three were male and eight were female. No participant reported any hearing or speaking disorder.

2.3. Experimental Conditions

The experiments under both the reverberant and non-reverberant conditions were conducted in a sound-proof room in a laboratory at Sophia University. All stimuli were stored in a PC (Intel® Core™2 Quad CPU, 2.83 GHz), and they were presented via an audio interface (RME Fireface 800) and a digital mixer (Yamaha DME24N). In the current study, a digital reverberator (Roland RSS-303) was used to add reverberation to the stimuli. We selected a reverberant situation from the programs in the reverberator, referring to the RT of an airport where both native and nonnative listeners often have difficulty in understanding announcements. According to Yokoyama et al. [24], the RT of a lobby in Haneda airport was about 2.5 s at 500 Hz. Therefore, in the current study, a reverberant condition where RT was 2.6 s, which was the closest to the RT of a lobby in Haneda airport, was adopted. The nonreverberant condition was achieved by bypassing the reverberator. The RT of the nonreverberant condition was 0.1 s.

The RTs of the experiments were measured using a swept sine signal. The swept sine signal was played through the loudspeakers used in the experiments (Genelec 8020A), and the RTs were measured via a precision sound-level meter (RION NA-27). The RTs were measured at the position of the listeners’ head. The RTs under the two conditions were derived from the average of the RTs at 500, 1,000, and 2,000 Hz.

Stimuli with reverberation were presented from four loudspeakers (Genelec 8020A) surrounding a participant under the reverberant condition. The loudspeakers were placed along a circle with a radius of 1.3 m at equal angles of 45 degrees for left/right and front/rear. The loudspeaker height was 1.2 m. In the current study, the four loudspeakers surrounding a participant were used for simulating reverberant environments where sounds came from various directions. On the other hand, two loudspeakers in front of a participant were used under the nonreverberant condition for simulating the situation that a listener perceived only direct sound from a talker. Since the same stimulus was presented from each of the two loudspeakers at the front, the participants perceived the stimulus to originate from the center of the two loudspeakers.

2.4. Procedure

The perceptual experiments were conducted using Praat software [25]. The participants were asked to identify whether the presented word included a short or long vowel and whether the presented word included a single or geminate consonant. Then, they were asked to give their response by pressing the labeled button on the PC screen. The target words in Japanese katakana letters were presented on the PC screen in the experiments for both native and nonnative listeners. Since all nonnative listeners were learners of Japanese, they could read the katakana letters. Each stimulus was presented to participants three times in random order.

The experiment under the reverberant condition was carried out first since the aim of the experiment was to investigate how listeners identified a word they had not heard before. If the experiment under the nonreverberant condition was conducted first, the listeners might find acoustic cues other than duration for identifying the words, and they could use them under the reverberant condition as well.

In total, 192 trials $\{11$ steps $\times 2$ (vowel and nasal) + 10 steps (stop) $\} \times 3$ repetitions $\times 2$ reverberant conditions $\}$ were conducted for each participant. In addition, the participants were also asked to report on their confidence regarding their answers on a scale of 1 to 5, with 5 being “with confidence” and 1 being “without confidence.”

2.5. Modeling

To compare the perceptual performance between the reverberant and nonreverberant conditions, and between native and nonnative listeners, the performance was required to be modeled mathematically. According to McMurray and Spivey [23], the identification function for the categorical perception is similar to the logistic function. Thus, in the current work, we used the logistic function to model listeners’ responses. The equation for modeling is as follows.

$$y = \frac{1}{1 + \exp(-a - bx)}$$

$y$ indicates the proportion of “long” responses (i.e., long
vowel or geminate consonant) and \( x \) is the number of stimulus steps. \( a \) and \( b \) are maximum likelihood estimates, and \( b \) is the rate of change. In the current study, the perception of each participant was modeled using the logistic function.

In the current study, \( b \) (i.e., slope), the category boundary (CB), and the proportion of “long” responses to the extreme stimuli were compared between conditions and between listener groups. The slope was regarded as the sharpness of the distinction between short and long. The proportion of “long” responses to the longest stimulus was regarded as the distinction of a typical geminate or long vowel word. Also, the proportion of “long” responses to the shortest stimulus was regarded as the distinction of a typical single consonant or short vowel word. Regarding the CBs, the \( x \)-intercept was regarded as the CB where the “long” response rate was 50%. The CBs were obtained by substituting 0.5 for \( y \) in the equation for modeling. Therefore, CBs were estimated using the equation for \(-a/b\).

3. RESULT

3.1. Distinction of Vowel Length Contrast

Figure 1 shows the distinction between /baba/ and /babaa/ for native listeners, and Fig. 2 shows the distinction for nonnative listeners under the reverberant (solid line) and nonreverberant (dashed line) conditions. Both Figs. 1 and 2 were drawn using the average of the results of modeling for each listener. Figures 1 and 2 indicate the proportion of “long vowel” responses as a function of vowel duration. Table 2 shows the average of slopes, CBs, and the proportion of “long vowel” responses to the extreme stimuli (i.e., the shortest and longest stimuli) for native and nonnative listeners under each condition.

A $2 \times 2$ ANOVA for repeated measures was performed on the slope. The results confirmed the main effect of the native language of listeners to be statistically significant [$F(1,19) = 21.41, p < 0.001$]. Also, the results confirmed the main effect of the reverberant condition to be statistically significant [$F(1,19) = 32.815, p < 0.0001$]. The results indicated that the interaction effect was not statistically significant and that the slopes for both native and nonnative listeners were significantly shallower under the reverberant condition.

A $2 \times 2$ ANOVA for repeated measures was also performed on the CB. The data that did not cross the point of 50% was excluded from the analysis. The results confirmed the interaction effect to be statistically significant [$F(1,17) = 11.69, p < 0.01$]. The analysis of the simple main effect revealed that the CBs for native listeners were significantly different between the reverberant and nonreverberant conditions [$F(1,20) = 16.25, p = 0.001$]. However, the CBs for nonnative listeners were not significantly different between the conditions [$F(1,14) = 0.092, p = 0.766$]. The result showed that the CB under the reverberant condition for native listeners was significantly shorter than that under the nonreverberant condition. In addition, the CB for native listeners was shorter than that for nonnative listeners under the reverberant condition.
The results indicated that the CBs were significantly different between native and nonnative listeners under the reverberant condition, whereas there was no difference under the nonreverberant condition. Regarding the CB, reverberation affected the performance of native listeners more than that of nonnative listeners.

The results of a $2 \times 2$ ANOVA for repeated measures of the proportion of “long vowel” responses to the longest stimulus confirmed the interaction effect to be statistically significant $[F(1, 19) = 8.925, p < 0.01]$. The analysis of the simple main effect revealed that the proportion for native listeners was significantly different from that for nonnative listeners under both the nonreverberant $[F(1, 19) = 5.54, p < 0.05]$ and reverberant $[F(1, 19) = 14.89, p < 0.01]$ conditions.

Regarding the proportion of “long vowel” responses to the shortest stimulus, the results confirmed the main effect of the reverberant condition to be significant $[F(1, 19) = 17.478, p < 0.001]$. The results indicated that both native and nonnative listeners gave significantly more “long vowel” responses to the shortest stimulus under the reverberant condition than under the nonreverberant condition. The shortest end of the continuum of vowel duration (i.e., 88.0 ms in Table 1) was regarded as a typical short vowel. However, with reverberation, even native Japanese listeners might have difficulty in identifying the short vowel with an appropriate duration to a short vowel.

### 3.2. Distinction between Single and Geminate Stops

Figure 3 shows the distinction between /ata/ and /atta/ for native listeners, and Fig. 4 shows the distinction for nonnative listeners under the reverberant (solid line) and nonreverberant (dashed line) conditions. Both Figs. 3 and 4 were drawn using the average of the results of modeling for each listener. Figures 3 and 4 indicate the proportion of “geminate stop” responses as a function of closure duration. Table 3 shows the average of slopes, CBs, and the proportion of “long” responses to the extreme stimuli (i.e., the shortest and longest stimuli) for native (NL) and nonnative listeners (NNL).

A $2 \times 2$ ANOVA for repeated measures was performed on the slope, the CB, and the proportion of “geminate” responses to the extreme stimuli. The results of the slope confirmed the main effect of the native language of listeners to be significant $[F(1, 19) = 33.02, p = 0.0001]$. The results of the CBs also confirmed the main effect of the native language of listeners to be significant $[F(1, 19) = 5.244, p = 0.0336]$. The results of a $2 \times 2$ ANOVA for repeated measures of the proportion of “geminate” responses to the extreme stimuli did not confirm any main effects or interaction to be significant. The results indicated that the slopes for nonnative listeners were significantly shallower than those for native listeners, and that the CB for nonnative listeners was longer (i.e., they gave more “singleton stop” responses) than that for native listeners. The performance difference was not large under the reverberant condition.
3.3. Distinction between Single and Geminate Nasals

Figure 5 shows the distinction between /ama/ and /amma/ for native listeners, and Fig. 6 shows the distinction for nonnative listeners under the reverberant (solid line) and nonreverberant (dashed line) conditions. Both Figs. 5 and 6 were drawn using the average of the results of modeling for each listener. Figures 5 and 6 indicate the proportion of “geminate nasal” responses as a function of nasal murmur duration. Table 4 shows the average of slopes, CBs, and the proportion of “geminate” responses to the extreme stimuli (i.e., the shortest and longest stimuli) for native and nonnative listeners under each condition.

A $2 \times 2$ ANOVA for repeated measures was performed on the slope. The results confirmed the main effect of the language of listeners to be statistically significant $[F(1, 19) = 16.96, p < 0.001]$. The results also confirmed the main effect of the reverberant condition to be statistically significant $[F(1, 19) = 25.968, p < 0.0001]$. There was no significant interaction effect. The results indicated that the slopes for both native and nonnative listeners were significantly shallower under the reverberant condition, and that the slopes for nonnative listeners were significantly shallower than those for native listeners. The performance difference did not increase under the reverberant condition.

A $2 \times 2$ ANOVA for repeated measures was also performed on the CB. The results confirmed the interaction effect to be statistically significant $[F(1, 18) = 19.258, p = 0.001]$. The analysis of the simple main effect revealed that the CB under the reverberant condition was significantly different from that under the nonreverberant condition for both native $[F(1, 20) = 7.634, p = 0.012]$ and nonnative $[F(1, 16) = 8.916, p < 0.01]$ listeners. Also, the CB for native listeners was significantly shorter than that for nonnative listeners under the reverberant condition $[F(1, 18) = 35.89, p < 0.0001]$, although there was no significant difference between native and nonnative listeners under the nonreverberant condition $[F(1, 18) = 1.39, p = 0.254]$. The results indicated that the performance difference between native and nonnative listeners became large under the reverberant condition, since native listeners exhibited increased “geminate” responses with reverberation, while nonnative listeners exhibited decreased “geminate” responses with reverberation.

A $2 \times 2$ ANOVA for repeated measures was performed on the proportion of “geminate nasal” responses to the longest stimulus. The results confirmed the main effect of the native language of listeners to be significant $[F(1, 19) = 9.119, p < 0.01]$. Regarding the proportion of “geminate nasal” responses to the shortest stimulus, the results confirmed the main effect of the reverberant condition $[F(1, 19) = 6.813, p = 0.0172]$. The results indicated that both types of listener gave more “geminate nasal” responses to the shortest stimulus under the reverberant condition than under the nonreverberant condition.

### Table 4

|                  | NL     | NNL    |
|------------------|--------|--------|
| Slope            | 2.53   | 1.52   |
| CB               | 86.76  | 92.33  |
| Max              | 100.0  | 96.7   |
| Min              | 0.24   | 2.55   |

Figure 5 shows the proportion of “geminate nasal” responses as a function of nasal murmur duration for native listeners under the reverberant (solid line) and nonreverberant (dashed line) conditions. Lines represent fitted curves.

Figure 6 shows the proportion of “geminate nasal” responses as a function of nasal murmur duration for nonnative listeners in the reverberant (solid line) and nonreverberant (dashed line) conditions. Lines represent fitted curves.
dition. In addition to the distinction of vowel length contrast, nasal murmur of 51.1 ms (see Table 1), which was regarded as the typical singleton nasal under the nonreverberant condition, could be confused with geminate nasal under the reverberant condition.

4. DISCUSSION

4.1. Distinction for Native Listeners in Reverberation

The results showed that even native Japanese listeners had difficulty in distinguishing between /baba/ and /babaa/, and between /ama/ and /amma/ with reverberation. The slopes of the distinction of vowel and nasal length contrasts were significantly shallower under the reverberant condition. This indicated that even native listeners could not distinguish the length contrasts under the reverberant condition as sharply as under the non-reverberant condition. Since the duration of a vowel in the word-final position tends to be ambiguous even under the nonreverberant condition, native listeners seemed to have difficulty in determining the offset of the vowel in reverberation. The distinction of the shortest stimulus also showed that native listeners could not determine the offset, and then identified it as a long vowel word. Similarly, native listeners seemed to have difficulty in accurately determining the nasal murmur period accurately in reverberation since the boundaries between the nasal murmur and adjacent vowels might become blurred in a long reverberation. On the other hand, native listeners could identify singleton and geminate stops under the reverberation and nonreverberant conditions despite the preceding vowel overlapping the onset of the closure period and making the boundary unclear. Since the burst of /t/ could be determined even when it was overlapped by the preceding vowel, the length from the onset of a word to the burst of /t/ could be a perceptual cue for its distinction. The results suggested that the offset of the vowel or constriction period will be a crucial cue for distinguishing the length contrasts.

The results showed that CBs became shorter under the reverberant condition in all distinctions, although it was not significant in the distinction between /ata/ and /atta/. The results indicated that native listeners gave more “long” responses to stimuli with reverberation. The “tail” added by reverberation might make native listeners perceive the duration of vowel and nasal as being longer than the actual duration. They seemed to distinguish between short and long phonemes depending on mostly the duration that they actually heard with reverberation. In addition, there is a possibility that native listeners gave “long” responses to stimuli they wavered between short and long, since they might have a bias toward identifying stimuli as long or geminate words because they knew that the experiment was on the perception of Japanese length contrasts.

Some studies have revealed that listeners could adapt their perception to reverberant environments using carrier phrases or sentences preceding the target [26,27]. The results of the current study did not show this adaptation to reverberation, although the duration of the carrier phrase preceding the targets (i.e., 1,563.1 ms) was longer than that in the previous study [26], and it was assumed to be sufficient to adapt their perception to the reverberation. In the current study, we used a long reverberation close to that of an airport. There is the possibility of a limit of the environment to which listeners could adapt their perception.

4.2. Distinction for Nonnative Listeners in Reverberation

Similarly to native listeners, nonnative listeners also could not distinguish sharply between /baba/ and /babaa/, and between /ama/ and /amma/ under the reverberant condition. On the other hand, they could distinguish between /ata/ and /atta/ under the reverberant and nonreverberant conditions. Geminate stops have been reported to be easier to identify than geminate nasal, and nonnative listeners acquired geminate stops earlier than geminate nasal [28,29]. Regarding coda nasal, Muroi [30] showed that nonnative listeners had difficulty in determining coda nasal even if their average LOR in Japan was more than two years. The difference between stop and nasal consonants under the effect of reverberation might be attributed to the disadvantage in distinguishing the nasal length contrast.

Similarly, the slope under the reverberant condition was 0.37, and Fig. 2 shows the flatter line for nonnative listeners under the reverberant condition. The perception for nonnative listeners could be regarded as a linear distinction rather than a categorical perception. In a future study, the discrimination test should be conducted to examine whether nonnative listeners can distinguish the stimuli across the CB. Oguma [31] showed that for nonnative listeners, the vowel length contrasts in the word-final position were more difficult to distinguish than those in the word-initial and word-medial positions. It seems to become too difficult for nonnative listeners to distinguish them categorically with reverberation.

There was a difference between native and nonnative listeners in dealing with stimuli having ambiguous durations with reverberation. In contrast to native listeners, for nonnative listeners, the CB became longer with reverberation. The results indicated that nonnative listeners gave “short” responses to stimuli that were difficult to distinguish. Japanese is usually said to have a mora-timed rhythm [32]. Native Japanese listeners perceive a short vowel as one mora and a long vowel as two morae.
Regarding geminate consonants, the lengthened part of a consonant is counted as one mora. On the other hand, English has a stress-timed rhythm where the intervals between stressed syllables are isochronous. Native English listeners perceive a short vowel as one syllable and a long vowel also as one syllable. Regarding geminate consonants, the lengthened part of a consonant cannot constitute one syllable. It has been reported that the prosodic difference caused nonnative listeners to perceive stimuli that native listeners identify as “long” as “short” [29]. The results indicated that the tendency caused by the difference in phonemic systems might become more marked with reverberation, as shown by previous studies [19–21]. Phonological rules of native language might be used when we compensate for the degraded sounds.

In addition, nonnative listeners gave more “short” responses with reverberation in spite of the target duration being elongated by reverberation. Their distinction with reverberation might include the expectation of elongation in the stimuli. This expectation might be affected by developing mental categories for nonnative listeners. In the native language magnet (NLM) theory developed by Kuhl and Iverson [33], each category has a prototype that is regarded as a good example of a category and pulls in stimuli surrounding the prototype (i.e., magnet effect). The NLM theory claims that a prototype depends on language experience. Pisoni and Lively [34] claimed that a category consists of many exemplars, and a prototype could be regarded as an average of the exemplars. Since we always hear speech sounds with noise and reverberation in our daily lives, it is suggested that we build the prototype of each phoneme category from many degraded sounds in the process of first-language acquisition.

On the other hand, nonnative listeners might go through the opposite process to native listeners when building prototypes. In the process of learning a nonnative language, students first listen to a good example (i.e., the prototype) of nonnative phonemes in the pronunciation of a teacher or teaching materials. They might initially acquire prototypes with little exposure to various pronunciations. Pisoni and Lively [34] claimed that exposure to various sounds uttered by various talkers and in diverse environments is required to learn unfamiliar sounds. Adequate exposure to various pronunciations helps to create robust prototypes. It could be considered that a prototype could not pull in sounds surrounding the prototype without sufficient exposure in the process of building the prototype since listeners have not learned which differences they can disregard between the prototype and candidates. Therefore, the prototypes of nonnative listeners can not pull in sounds surrounding the prototypes.

The results showed that CBs for nonnative listeners tended to become longer under the reverberant condition. The results indicated that they gave more “short vowel” or “singleton consonant” responses under the reverberant condition. Since English does not have a long vowel or a geminate consonant as a phoneme, native English listeners should build a category of long vowels and geminate consonants in the process of acquiring Japanese. There was a possibility that stimuli with reverberation would be pulled in by the prototype of a short vowel and a singleton consonant, which exist even in English, since the prototype of a short vowel and a singleton geminate might be more robust than the prototype of a long vowel and a geminate consonant.

5. CONCLUSION

In the current study, the categorical perception of Japanese length contrasts for native and nonnative listeners under the reverberant condition was investigated. Stimuli were durational continua of vowel, nasal murmur, and closure. Experiments were conducted under the reverberant (RT = 2.6 s) and nonreverberant conditions. Because a “tail” is added to stimuli by reverberation, even native listeners identified the shortest stimulus as “long” in the distinction of vowel length contrast in the word-final position. The distinction of consonant length contrast seemed to be easier than that of vowel length contrast since the consonant length contrast was in the word-medial position. Similarly, nonnative listeners had the most difficulty with vowel length contrast. Their perception could not be regarded as categorical perception. In addition, there was a difference in the shift of the CB between the listener groups. The results showed that “long” responses increased for native listeners, whereas “short” responses increased for nonnative listeners with the reverberation in the distinction of vowel and nasal length contrasts. It is assumed that the difference could be attributed to the different prototypes of categories of Japanese between native and nonnative listeners. Since the prototypes of categories of Japanese long phonemes for nonnative listeners might not be as robust as those for native listeners, they could not pull in stimuli that had been degraded by reverberation.

REFERENCES

[1] T. Arai, E. Osawa, N. Hodoshima and T. Igeta, “Native Japanese speakers’ perceptual performance declines in reverberation,” Proc. Autumn Meet. Acoust. Soc. Jpn., pp. 293–296 (2016).
[2] T. Arai, E. Osawa, N. Hodoshima and T. Igeta, “Reverberation degrades native listeners’ perception of Japanese monosyllables/special morae,” Acoust. Sci. & Tech., 39, 252–255 (2018).
[3] E. Osawa, T. Arai, N. Hodoshima and T. Igeta, “Non-native speakers perform poorer than native speakers in Japanese speech recognition in reverberation,” J. Acoust. Soc. Am., 140, 3333 (2016).
[4] H. Fujisaki, K. Nakamura and T. Imoto, “Auditory perception of duration of speech and non-speech stimuli,” in *Auditory Analysis and Perception of Speech*, G. Frant and M. A. A. Tatham, Eds. (Academic Press, London, 1975), pp. 197–219.

[5] H. Fujisaki and M. Sugito, “The physical properties of speech sound,” Iwanami Course on Japanese 5: Phonology (Iwanami Shoten, Tokyo, 1977), pp. 63–98 (in Japanese).

[6] T. Uchida, “Categorical perception of relatively steady-static speech duration in Japanese moraic phonemes,” *J. Phon. Soc. Jpn.*, 2(3), pp. 71–86 (1998).

[7] Y. Hirata and S. G. Lambacher, “Role of word-external contexts in native speakers’ identification of vowel length in Japanese,” *Phonetica*, 61, 177–200 (2004).

[8] E. Ofuka, Y. Mori and S. Kiritani, “Perception of a Japanese geminate stop: The effect of the duration of the preceding/following vowel,” *J. Phon. Soc. Jpn.*, 9(2), pp. 59–65 (2005).

[9] J. Kingston, S. Kawahara, D. Chambless, D. Mash and E. Brenner-Alsop, “Contextual effects on the perception of duration,” *J. Phon.*, 37, 297–320 (2009).

[10] H. Kato, K. Tajima, A. Rothwell, R. Akahane-Yamada and K. Munhall, “Perception of phonemic length contrasts in Japanese with or without a carrier sentence by native and non-native listeners,” *Proc. ICA*, pp. I-609–612 (2004).

[11] T. Toda, “The categorical perception of Japanese geminate consonants and long vowel by learners of Japanese,” *The Study of Literature and Language; The Language Issue*, 33, 65–82 (1998).

[12] Y. Hirata, E. Whitehurst and E. Cullings, “Training native English speakers to identify Japanese vowel length contrast with sentences at varied speaking rates,” *J. Acoust. Soc. Am.*, 121, 3837–3845 (2007).

[13] J. S. Bradley, H. Sato and M. Picard, “On the importance of early reflections for speech in rooms,” *J. Acoust. Soc. Am.*, 113, 3233–3244 (2003).

[14] V. O. Knudsen, “The hearing of speech in auditoriums,” *J. Acoust. Soc. Am.*, 1, 56–82 (1929).

[15] A. K. Nábělek, T. R. Letowski and F. M. Tucker, “Reverberant overlap- and self-masking in consonant identification,” *J. Acoust. Soc. Am.*, 86, 1259–1265 (1989).

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

[17] R. H. Bolt and A. D. MacDonald, “Theory of speech masking by reverberation,” *J. Acoust. Soc. Am.*, 21, 577–580 (1949).

[18] Y. Nakata, Y. Murakami, N. Hodoshima, N. Hayashi, Y. Miyachi, T. Arai and K. Kurisu, “The effects of speech-rate slowing for improving speech intelligibility in reverberant environments,” *IEICE Tech. Rep.*, pp. 21–24 (2006).

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

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

[21] H. Masuda, “Misperception patterns of American English consonants by Japanese listeners in reverberant and noisy environments,” *Speech Commun.*, 79, 74–87 (2016).

[22] E. J. Kreul, J. C. Nixon, K. D. Kryter, D. W. Bell, J. S. Lang and E. D. Schubert, “A proposed clinical test of speech discrimination,” *J. Speech Hear. Res.*, 11, 536–552 (1968).

[23] B. McMurray and M. Spivey, “The categorical perception of consonants: The interaction of learning and processing,” *Proc. Chicago Linguist. Soc.*, 34(2), pp. 205–220 (2000).

[24] S. Yokoyama, H. Mukai and H. Tachibana, “Examples of field measurement of acoustic conditions in public spaces,” *J. INCE Jpn.*, 23, 228–231 (1999).

[25] P. Boersma and D. Weenink, “Praat, a system for doing phonetics by computer,” *Glot Int.*, 5(9/10), 341–345 (2016).

[26] A. J. Watkins, “Perceptual compensation for effects of reverberation in speech identification,” *J. Acoust. Soc. Am.*, 118, 249–262 (2005).

[27] T. Kounura and S. Furukawa, “Effect of reverberation and its presentation context on material perception based on impact sounds,” *Proc. Spring Meet. Acoust. Soc. Jpn.*, pp. 1499–1502 (2017).

[28] K. Enomoto, “Interlanguage phonology: The perceptual development of durational contrasts by English-speaking learners of Japanese,” *Edinburgh Work. Pap. Linguist.*, 3, 25–35 (1992).

[29] T. Toda, “Learners’ categorical perception of geminates and long vowels,” *The Study of Literature and Language; The Language Issue*, 33, 65–82 (1998).

[30] K. Muroi, “Problems in perception and production of Japanese special morae for English learners,” *Sophia Linguist.*, No. 38, pp. 41–60 (1995).

[31] R. Oguma, “The perception of Japanese length contrasts by English learners,” *Education of Japanese in the World*, No. 10, pp. 43–55 (2000).

[32] T. J. Vance, *The Sound of Japanese* (Cambridge University Press, New York, 2008).

[33] P. K. Kuhl and P. Iverson, “Linguistic experience and the ‘perceptual magnet effect,’” in *Speech Perception and Linguistic Experience: Issues in Cross-language Research* (York Press, 1995), pp. 121–154.

[34] D. B. Pisoni and S. E. Lively, “Variability and invariance in speech perception: A new look at some old problems in perceptual learning,” in *Speech Perception and Linguistic Experience: Issues in Cross-language Research* (York Press, 1995), pp. 433–459.