Misperception of Japanese words with devoiced vowels and/or geminate consonants by young and elderly listeners

Eri Iwagami1,*, Takayuki Arai1,†, Keiichi Yasu2 and Kei Kobayashi3

1Sophia University, 7–1, Kioi-cho, Chiyoda-ku, Tokyo, 102–8544 Japan
2Tsukuba University of Technology, 4–3–15, Amakubo, Tsukuba, Ibaraki, 305–8520 Japan
3The University of Auckland, 26 Symonds St, Auckland, 1010 New Zealand

(Received 10 April 2017, Accepted for publication 25 October 2017)

Abstract: In this study, two perception experiments were conducted to investigate the misperception of Japanese words with devoiced vowels and/or geminate consonants by young and elderly listeners. In Experiment 1, eight young normal-hearing listeners participated under a white-noise condition, and eight elderly listeners participated in Experiment 2. Two types of word sets which consist of combinations of vowels (V = /i, u/) and voiceless consonants (C = /k, t, s/) were used as stimuli. The first word set involved two- or three-mora words and the second word set had 14 minimal pairs of CVC(:) V, where (:) stands for with or without a geminate consonant. The results of both experiments showed that misperception was great for words with devoiced vowels and even greater for words with geminate consonants. In particularly, the misperception of consonants including high frequency components such as /shi/ or /shu/ was observed for elderly listeners.

Keywords: Misperception, Elderly listeners, Japanese, Devoiced vowels, Geminate consonants

PACS number: 43.71.Es, 43.71.Ky, 43.71.Lz [doi:10.1250/ast.39.109]

1. INTRODUCTION

Although C (consonant) + V (vowel) is the basic structure of syllables in Japanese, combinations or elongations of consonants occur owing to devoiced vowels or geminate consonants. In the Tokyo dialect, speakers often devoice high vowels (/i/, /u/) when they are between voiceless consonants [1]. Devoicing of vowels frequently occurs in spontaneous Japanese and, as a result, several types of consonantal sequences similar to consonant clusters in English have been observed [2]. In any case, when devoicing of vowels occurs, the vocalic portion of the entire duration, or the vocalic rate, of an utterance decreases. Interestingly, Japanese listeners have reported hearing vowels between consonants when they listen to words which have consonant clusters [3]. Geminate consonants can also be a part of an utterance in Japanese. When there is a geminate consonant, the vocalic rate of an utterance also decreases. Devoicing of vowels and geminate consonants can also co-occur in an utterance [4].

In the case of the C1V1C2:V2 structure, where the colon denotes gemination of C2, the devoicing rate of V1 is the highest for the combination of preceding fricative consonants (C1) and following stop consonants (C2). Meanwhile, the devoicing rate of V1 is the lowest for the combination of preceding fricative consonants (C1) and fricative consonants (C2) [4]. On the other hand, Fais et al. reported that vowel devoicing occurs in infant-directed speech in Japanese as well as in adult speech, although they hypothesized that the devoicing rate decreases when mothers speak to their babies for language acquisition purposes [5]. However, it has also been reported that the rate of devoicing vowels is lower when teachers speak to hearing-impaired children, although vowel devoicing also occurs in adult speech [6].

It was also tested whether the vocalic portions of a sentence contribute more to intelligibility than the consonantal portions for both young and elderly listeners. Kewley-Port et al. investigated the contribution of information from vowels and consonants on young normal-hearing and elderly hearing-impaired listeners who were native speakers of English [7]. Vowel sentences where consonants were replaced by speech-shaped noises and
consonant sentences where vowels were replaced by the same noises were presented to both elderly hearing-impaired and young normal-hearing listeners. From the results, it was reported that vowel sentences were more intelligible than consonants for both groups, especially for elderly hearing-impaired listeners.

For elderly hearing-impaired listeners, some studies have found difficulties with perceiving consonants in Japanese. Identification tests using continua of /t s/ and /k, t, ts/ were conducted to investigate the relationship between the perception of fricative and affricate sounds and auditory deficits [8]. The participants participated in the experiment. Their ages ranged from 20 to 23 years old with a mean of 21.3. All participants were native speakers of Japanese with normal hearing.

2. EXPERIMENT 1

In this experiment, we investigated whether the misperception rate by young normal-hearing listeners increases under a white-noise condition to simulate the perception of elderly listeners.

### Table 1

|        | C₁   | V₁   | C₂   | V₂   | C₃  | V₃   | Carrier sentence |
|--------|------|------|------|------|-----|------|------------------|
| two morae | /k, tʃ, ts, jʃ/ | /i, u/ | /k, tʃ, ts, jʃ/ | any V | any C or Ø | any V or Ø | to iimashita     |
| three morae | /k, tʃ, ts, jʃ/ | /i, u/ | /k, tʃ, ts, jʃ/ | /i, u/ | any C or Ø | any V or Ø | to iimashita     |

### Table 2

|        | C₁   | V₁   | C₂   | V₂   | Carrier sentence |
|--------|------|------|------|------|------------------|
| Type I | /k, tʃ, ts, jʃ/ | /i, u/ | /k, tʃ, ts, jʃ/ | any V | to iimashita     |
| Type II | any C or Ø | any V or Ø | /k, tʃ, ts, jʃ/ | /i, u/ | to iimashita     |

2.1. Methods

2.1.1. Participants

Eight young normal-hearing listeners (M:F = 3:5) participated in the experiment. Their ages ranged from 20 to 23 years old with a mean of 21.3. All participants were native speakers of Japanese with normal hearing.

2.1.2. Stimuli

We used words (target words) whose word familiarity was three or more on a seven-point scale [10,11]. Each word consisted of combinations of voiceless consonants (/k, tʃ, ts, jʃ/) and two close vowels (/i, u/) based on the phonotactic rules in Japanese. That is, /k/ and /ʃ/ were combined with /i/ and /u/, resulting in /ki/, /ku/, /ʃi/, and /ʃu/. /ʃi/ and /ts/ were combined with /i/ and /u/, respectively, resulting in /ʃi/ and /ʃu/. There were two sessions in this experiment. Table 1 shows the combinations of the vowels and consonants of the words used in Session 1 (shaded cells in this table mean that any consonant or vowel (or none (Ø)) can be used). Table 2 shows the combinations of the consonants and vowels of the words used in Session 2.

The words of Session 1 consisted of 84 Japanese words which have the C₁V₁C₂V₂ (two morae) or C₁V₁C₂V₂C₃V₃ (three morae) structure such as ‘shika’ (deer) or ‘kitsune’ (fox). In the two-mora words, V₁ was in the devoicing context in which the vowel can be devoiced. In some three-mora words such as ‘shika’ (sea lion), V₁ and V₂ were in the devoicing context, respectively, for future work to analyse whether the correct rate changes depending on the position of the devoiced vowel. Moreover, all the vowels...
were in the devoicing context in some two- and three-mora words such as ‘shikishi’ (message card) or ‘chichi’ (father). The target word was inserted into the carrier sentence ‘watashi wa ____ to imashita. (I said ____).’ and each target word was recorded by a female standard Japanese speaker. All 84 words were recorded twice. The speaker first produced all of them with the devoiced vowels in the devoicing context forcibly uttered; after that, the speaker produced them with the non-devoiced vowels forcibly uttered, even when the vowels were in the devoicing context.

Session 2 had two types of words. One type of word in Session 2 (Type I) consisted of seven Japanese minimal pairs (a total of 14 words) which had C₁V₁C₂V₂ (with non-geminate consonant) and C₁V₁C₂V₂ (with geminate consonant) structures, such as ‘kita’ (came) or ‘kitta’ (cut). In this case, V₁ was in the devoicing context. The other types of words in Session 2 (Type II) also consisted of other seven minimal pairs such as ‘nishi’ (west) and ‘nisshi’ (diary) (a total of 14 words), where V₂ was in the devoicing context because of the voiceless consonant /t/ at the beginning of the carrier sentence after V₂. V₁ and V₂ were in the devoicing context in Type I and Type II, respectively, again for future work to analyse whether the misperception rate changes with the position of the devoiced vowels. All of the 28 words were recorded twice. The speaker first produced all of them with devoiced vowels in the devoicing context forcibly uttered; after that, the speaker produced them with the non-devoiced vowels forcibly uttered, even when the vowels were in the devoicing context.

2.1.3. Procedure

In this experiment, all participants were divided into two SNR groups (Group 1: SNR = 0 dB, Group 2: SNR = 3 dB) of four people. Table 3 shows the relation between the groups of participants and the stimuli which they perceived. The 84 words of Session 1 were divided into four word sets: W1a, W1b, W1c and W1d. Each word set contained 21 words. The suffixes ‘a,’ ‘b,’ ‘c,’ and ‘d’ indicated in Table 3 were for the first two participants in Session 1. For the other two participants, the suffixes, ‘a’ and ‘c’ were exchanged with ‘b’ and ‘d,’ respectively. This was done to counterbalance for the combination of noise vs. devoicing conditions.

The four word sets were presented in the order indicated in parentheses in Table 3 to test the effect of the order of devoiced vs non-devoiced stimuli. For example, W1a and W1c (42 words) were first presented to the first and second participants under the noise condition, respectively, and W1b and W1d (42 words) were presented under clean condition. For the third and fourth participants, W1b and W1d were first presented under the noise condition, respectively, and W1a and W1c were first presented under the clean condition. The words in W1a, W1b, W1c, and W1d were presented in random order in each block.

In Session 2, the uniform word set (W2) was presented to all of the participants under both the noise and clean conditions in the order indicated in parentheses in Table 3. W2 was presented in random order.

224 stimuli in the carrier sentences were presented randomly (168 (Session 1) + 28 (Session 2) × 2 (conditions)) to all participants of both groups. This experiment took 1–1.5 hour, and participants took a break before each session.

They were instructed to write down the words they heard on an answer sheet in hiragana. They were asked to answer without thinking too much, and to write down only the parts of the words that they heard when they could not hear or understand the entire utterance clearly for later analysis of the error pattern. We calculated the correct rate for the case that V₁ in two-mora words, V₂ in three-mora words of Word set 1 or 2, and each preceding consonant

| Table 4 A-weighted sound pressure levels of stimuli and white-noise for each SNR condition. |
|-----------------|-----------------|-----------------|
| Stimuli (from PC 1) | White-noise (from PC 2) |
| SNR 0 dB | 60 dBA | 60 dBA |
| SNR +3 dB | 60 dBA | 57 dBA |

![Fig. 1 Experimental setup.](image-url)
(target mora) were correct. The experiment took place in a sound-proof room at a university in Tokyo. Figure 1 shows the experimental setup.

The stimuli were presented over headphones (Sennheiser HDA200) (Yamaha MW10) at an A-weighted sound pressure level of 60 dB, as measured on a sound level meter and with an artificial ear (Bruel & Kjær Type 4153). White-noise was presented over the headphones passing through a mixer at an A-weighted sound pressure levels of 57 dB (SNR = 3 dB) and 60 dB (SNR = 0 dB), as measured on the sound level meter and with the artificial ear.

Table 4 shows A-weighted sound pressure levels of stimuli and white-noise for each SNR condition.

2.2. Results

Figure 2 shows the correct rate of Word set 1 with white-noise (SNR = 0 dB) and without white-noise. The correct rate with devoiced vowels was lower than with non-devoiced vowels when SNR was 0 dB. The data were analysed using two-factor analysis of variance (ANOVA) with repeated measures in IBM SPSS Statistics (SPSS). There was a significant difference between devoiced and non-devoiced vowels [F(1,3) = 20.555, p = 0.02]. Moreover, a multiple comparison test was conducted using Bonferroni’s method. There was a significant difference between devoiced and non-devoiced vowels when SNR was 0 dB (p = 0.019).

Figure 3 shows the correct rate of Word set 1 with white noise (SNR = 3 dB) and without noise. The data were analysed using two-factor ANOVA with repeated measures. There was a significant difference between clean signals and signals with SNR = 3 dB [F(1,3) = 129.046, p = 0.001]. This figure shows that the correct rate of devoiced vowels with SNR = 3 dB was lower than that with non-devoiced vowels. However, there were no significant differences between devoiced and non-devoiced vowels.

Figure 4 shows the correct rate of Word set 2 with white-noise (SNR = 0 dB) and without white-noise. The data were analysed using two-factor ANOVA with repeated measures. There was a significant difference between SNR = 0 dB and clean conditions [F(1,3) = 16.684, p = 0.027]. These figures show that the rate with devoiced vowels is lower than that with non-devoiced vowels. The correct rate for devoiced vowels with a geminate is especially low. However, there was no significant difference between devoiced and non-devoiced conditions.

Figure 5 shows the correct rate of Word set 2 with white-noise (SNR = 3 dB) and without white-noise. The
data were analysed using two-factor ANOVA with repeated measures. There was a significant difference between SNR = 3 dB and clean conditions \(F(1, 3) = 13.71, \ p = 0.034\). In this figure, the rate with devoiced vowels is lower than that with non-devoiced vowels. Moreover, the rate with devoiced vowels with a geminate is the lowest. However, there was no significant difference between devoiced and non-devoiced conditions.

Table 5 shows the two error patterns in Word sets 1 and 2 which were frequently observed under the white-noise condition in this experiment. The total number of errors in Word set 1 was 127, among which the listeners misperceived /f(u)/ as /f(i)/ and /f(i)/ as /f(t)/, 3 and 49 times, respectively (parentheses show that the vowels are devoiced). In Word set 2, the total number of the error was 14, among while the listeners misperceived /f(u)/ as /f(i)/ and a geminate as a non-geminate, 3 and 4 times, respectively.

2.3. Discussion

The correct rate of vowels under the white-noise condition was decreased by devoicing, although young normal-hearing listeners could almost always perceive utterances correctly under the clean condition. This suggests that the loss of the energy of the vocalic portions had an effect on the increased misperception rate.

Experiment 1 shows there was a tendency to misperceive consonants, especially those having high frequency components, under the white-noise condition. Consonants with high frequency components have the same aperiodic waveforms as white-noise. Thus, it is considered that the misperception increased because the white-noise easily masked consonants such as /f/ and /f(t)/.

From the results of Word set 2 in Experiment 1, it was observed that the misperception rate with the combination of devoiced vowels and geminate consonants was the highest. In terms of temporal resolution, it has been reported that elderly listeners have a longer just noticeable difference (JND) than young normal-hearing listeners [12].

Kawata et al. defined the temporal resolution of hearing as normal when the JND of the gap detection is less than 30 ms [9]. From the results of an identification experiment with different fricative durations, they reported that it is more difficult for elderly listeners with degraded temporal resolution to identify fricative duration correctly than it is for elderly listeners with only high-frequency deficits. In addition, there was a strong tendency of misperception for elderly listeners who had both degraded temporal resolution and high-frequency deficits.

The temporal resolution of the participants in Experiment 1 of this study was not degraded because we conducted this experiment on young normal-hearing listeners and we did not simulate the degradation of temporal resolution by simply adding white-noise.

Therefore, even masking the signals by adding white-noise makes it difficult to perceive consonants. From this, it is assumed that the correct rate with the combination of devoiced vowels and geminate consonants will be lower, especially for elderly listeners whose high-frequency resolution has deteriorated.

3. EXPERIMENT 2

In this experiment, we investigated whether the misperception rate with devoiced vowels in Japanese geminate consonants and non-geminate consonants by elderly listeners increases compared to the rate with non-devoiced vowels in geminate and non-geminate consonants.

3.1. Methods

3.1.1. Participants

Seventeen Elderly listeners (M : F = 9 : 8) participated in the experiment. Their ages ranged from 62 to 82 years with a mean of 72.6. All participants were native speakers of Japanese.

3.1.2. Pre-experimental tests

We conducted an audiometric test and a gap detection test before we carried out the main session in Experiment 2. In the audiometric test, seven pure tones with frequencies between 250 and 8,000 Hz were presented intermittently to all of the participants. The participants were asked to continue pushing a button while they were able to perceive the sound. Figure 6 shows the audiograms of the 17 elderly listeners. Their mean absolute threshold was a hearing level of 21.9 dB at six frequency points between 250 and 4,000 Hz. For the participants in Experiment 1, their mean absolute threshold was a hearing level of 4 dB at the same six frequency points. Figure 7 shows the audiograms of three young normal-hearing listeners in Experiment 1 as a subset of the young normal-hearing listeners.

In the gap detection test, all participants were presented with two types of white-noise, a continuous sound and a continuous sound with a gap in the middle [9]. The participants were asked to choose the sound with a gap in
the middle by touching a button on a PC panel. There was a 1 s break between the two stimuli. The participants were first presented with the target sounds with a 100 ms gap, and then the length of the gap was shortened by using the two-up one-down method. The sounds were presented over headphones (Sennheiser HDA200) at an A-weighted sound pressure level of 70 dB as measured on a sound level meter (ONO SOKKI LA-5111) connected to an artificial ear (Brüel & Kjær Type 4153). The mean of the JND of this gap detection test was 4 ms (excluding one participant whose mean was 159 ms).

3.1.3. Stimuli
The same stimuli as those used in Experiment 1 without white-noise were used in this experiment.

3.1.4. Procedure
In this experiment, the procedure was basically the same as that in Experiment 1. The noise condition was, however, not employed, resulting in 196 stimuli being presented (without white-noise) to all participants (168 (Session 1) + 28 (Session 2)). Table 6 shows the word sets which they perceived. In the results of Experiment 1, we did not find a difference between the correct rates originating from the order effect of devoiced vs non-devoiced stimuli. Therefore, Word sets 1 and 2 were presented in the order of the devoiced stimuli followed by the non-devoiced stimuli in Experiment 2. The numbers indicated in parentheses in Table 6 were the order of presentation of the word sets. This experiment took 1.5–2 hour, and participants took a break before each session.

The stimuli were presented over the headphones without passing through the mixer in Fig. 1 at an A-weighted sound pressure level of about 70 dB, as measured on the sound level meter with the artificial ear.

They were instructed to write down the words they heard on an answer sheet in hiragana. They were asked to answer without thinking too much, and to write down only the parts of the words that they heard when they could not hear or understand the entire utterance clearly for later analysis of the error pattern. We calculated the correct rate in the same way as in Experiment 1, and we also calculated the correct rate for the case that every mora was correct in this experiment.

3.2. Results
Figures 8 and 9 show the correct rates of Word sets 1 and 2, for the case that every mora was correct, respectively.

Figure 8 shows the correct rate of Word set 1. An independent t-test was conducted on the devoiced and non-devoiced conditions in SPSS. There was a significant difference between the devoiced and non-devoiced conditions \( t = 3.719, df = 16, p = 0.002 \).

Figure 9 shows the correct rate of Word set 2. In this figure, the rate with devoiced vowels is lower than that with non-devoiced vowels. The data were analysed using two-factor ANOVA with repeated measures. There was a significant difference between the stimuli \( F(3,1.143) = 4.789, p < 0.038 \).

A multiple comparison test was conducted using Bonferroni’s method. There were significant differences between the combinations of devoiced + geminate and non-devoiced + geminate \( (p < 0.001) \), devoiced + geminate and non-geminate + non-geminate \( (p = 0.011) \), and devoiced + non-geminate and non-devoiced + non-geminate \( (p = 0.009) \).

Figures 10 and 11 show the correct rates of Word sets 1 and 2 calculated from the points given when the target mora was correct, respectively.
Figure 10 shows the correct rate of Word set 1. An independent \(t\)-test was conducted on the devoiced and non-devoiced conditions in SPSS. There was a significant difference between the devoiced and non-devoiced conditions \(t = -3.396, df = 16, p = 0.004\).

Figure 11 shows the correct rate of Word set 2. In this figure, the rate with devoiced vowels is lower than that with non-devoiced vowels. The data were analysed using two-factor ANOVA with repeated measures. There was a significant difference between the stimuli \(F(3, 1.206) = 4.706, p = 0.037\).

A multiple comparison test was conducted using Bonferroni’s method. There were significant differences between the combinations of devoiced + geminate and non-devoiced + geminate \((p = 0.004)\), devoiced + geminate and non-devoiced + non-geminate \((p = 0.020)\), and devoiced + non-geminate and non-devoiced + non-geminate \((p = 0.001)\).

Table 7 shows two error patterns in Word sets 1 and 2 which were frequently observed in Experiment 2. The total number of errors in Word set 1 was 166, among while the listeners misperceived /\(u/\) as /\(i/\) and /\(i/\) as /\(j/\), 39 and 9 times, respectively. In Word set 2, the total number of errors was 66, among while the listeners misperceived /\(u/\) as /\(i/\) and a geminate as a non-geminate, 15 and 10 times, respectively.

** Table 7 ** Error patterns in Experiment 2.

| Consonants which listeners misperceived | Error patterns | Number of error/ total number of error |
|----------------------------------------|---------------|--------------------------------------|
| Word set 1                             |               |                                      |
| /\(u/\)                                | /\(i/\)       | 39/166                               |
| /\(i/\)                                | /\(j/\)       | 9/166                                |
| Word set 2                             |               |                                      |
| /\(u/\)                                | /\(i/\)       | 15/66                                |
| Geminate                               | Non-geminate  | 10/66                                |

** Fig. 9 ** Correct rate of devoiced and non-devoiced vowels by elderly listeners in Word set 2 with geminate and non-geminate consonants, non-d: non-devoiced, d: devoiced, non-g: non-geminate, g: geminate.

** Fig. 10 ** Correct rate of devoiced and non-devoiced vowels by elderly listeners in Word set 1 calculated from the points given when the target mora was correct.

** Fig. 11 ** Correct rate of devoiced and non-devoiced vowels by elderly listeners in Word set 2 in geminate and non-geminate consonants (non-d: non-devoiced, d: devoiced, non-g: non-geminate, g: geminate) calculated from the points given when the target mora was correct.

** Fig. 8 ** Correct rate of devoiced and non-devoiced vowels by elderly listeners in Word set 1.
3.3. Discussion

There was a tendency for elderly listeners to have a lower correct rate for devoiced vowels, although young normal-hearing listeners could almost always perceive utterances correctly in Experiment 1 under a clean condition. This result is consistent with the results of Experiment 1 under a white-noise condition (simulated hearing impairment scenario).

In addition, there was a tendency for the participants to misperceive consonants with high-frequency components such as /ʃ(u)/ as /ʃ(i)/ or /ʃ(i)/ as /ʃ(ɪ)/ observable in the results of Word sets 1 and 2 in Experiment 2. We have to use consonant information as perceptual cues of devoiced vowels, even if their energy is lacking. However, it is considered to be difficult for the participants with high-frequency deficits to perceive these consonants mainly consisting of high frequency components from consonantal cues. Therefore, we found that the misperception rate increased owing to devoicing of the vowels, suggesting that the lack of energy from the vowels and the use of consonants as perceptual cues increase the misperception rate for elderly listeners. It was possible for the young normal-hearing listeners to identify /ʃu/ and /ʃi/ with devoiced vowels under a clean condition in Experiment 1. However, there was a strong tendency for elderly listeners with high-frequency deficits due to hearing impairment to misperceive /ʃu/ as /ʃi/.

Matsui conducted a perception experiment on /ʃu/ and /ʃi/ for young normal-hearing listeners [13]. From the results of the experiment, it was reported that the formant transition of the end of the consonants is important to in perceiving /ʃu/. We analysed the devoiced vowels of /ʃu/ and /ʃi/ among the stimuli used in Experiment 2 following the method of Praat [14]. Figures 12 and 13 show spectra of the devoiced vowels of /ʃu/ and /ʃi/, respectively. From this analysis, we found that there was a difference in the spectra between /ʃu/ and /ʃi/ there are significantly more high-frequency components in /ʃi/ than in /ʃu/. On the other hand, we could not find a difference between the formant transitions of /ʃu/ and /ʃi/. Matsui used stimuli with /ʃu/ in the final position of the word, whereas we used stimuli with /ʃu/ in the word-initial and word-medial positions. Therefore, a formant transition might not be found in the stimuli of Experiment 2. Thus, we suspect that the spectral shape may cause elderly listeners misperceive /ʃu/ as /ʃi/. Hence, it is necessary to further investigate why elderly listeners to misperceive /ʃu/ as /ʃi/ as future work.

The results of Word set 2 in Experiment 2 reveal a tendency for the correct rate of devoiced vowels with geminate and non-geminate consonants to decrease, in particular, there were two types of error patterns: misperception of devoiced vowels and misperception of geminate vs non-geminate consonants.

In the first error pattern, there was the same tendency as that identified in the results of Word set 1. In the second error pattern, it was found that the participants misperceived fricative, affricate, or stop geminates as non-geminate consonants based on the error patterns of Type II. In fricative geminate consonants, there was also a tendency for elderly listeners having degraded hearing with high-frequency deficits to misperceive geminate consonants as non-geminate consonants based on the error patterns of Type II. Hence, it is considered that the elderly listeners’ high-frequency deficits cause the misperception of geminate consonants as seen in Kawata’s study [9] and discussed in Sect. 2.3.

In addition, it is considered that it was more difficult to perceive geminate consonants, especially those with devoiced vowels, in which the frication was elongated. The mean JND of the gap detection of the elderly listeners in the current experiment was less than 30 ms (except one participant: the data of this participant was included in the result of Experiment 2 because the misperception rate was not noticeable different from the others; we considered that there is a possibility that the participant did not understand the procedure of the gap detection test), therefore they were considered not to have degraded temporal resolution.
However, from the results in Kawata’s study, we considered that the misperception rate with devoiced vowels will be higher for elderly listeners whose temporal resolution was degraded.

From the error pattern of Type in Experiment 2, a tendency for the participants to misperceive affricate and stop geminate consonants (such as ‘tsutta’ (fished)) as non-geminate consonants was also found. For the perception of geminate consonants by young normal-hearing listeners, it has been reported that the formant transition affects the perception of geminate consonants [15]. Moreover, the formant transition also affects the perception of geminate consonants for elderly listeners [16,17]. It is possible that there was a formant transition in the devoiced vowels of the words, even when the vowels were devoiced. However, we considered that it would have been difficult for the elderly listeners to use the formant transition in devoiced vowels as the cue of the perception, owing to the lack of energy from the vowels. Hence, we suggest that misperception of geminates by elderly listeners is also caused by the formant transition when vowels are devoiced.

4. CONCLUSION

In this study, we conducted two perception experiments to investigate whether the misperception of Japanese words with devoiced vowels and/or geminate consonants increases for young normal-hearing listeners under a white-noise condition and for elderly listeners.

From the results of Experiment 1 for young normal-hearing listeners, we reported that correct rates of two or three morae of devoiced vowels decreased according to visual inspection; however, there was not a large difference. In particular, the perception correct rate was the lowest with the combination of devoiced vowels and geminate consonants. The tendency of the results of Experiment 1 was also found in Experiment 2.

From the results of both Experiments 1 and 2, we confirmed that the misperception rate with devoiced vowels, especially the combination of devoiced vowels and geminate consonants, increased for young normal-hearing listeners under white-noise and elderly listeners compared with the clean-signal condition for young listeners. We found that the increase in the misperception rate was due to devoicing of the vowels, especially when these vowels co-occur with geminate consonants, suggesting that the lack of energy in the vowels has an effect on participants correctly perceiving the tokens. Hence, we suggested that the energy of the vowels is important in the perception of Japanese words for elderly listeners.

ACKNOWLEDGEMENT

We would like to thank the participants for their cooperation in the experiments. Experiment 1 was reported at the 5th Joint Meeting: Acoustical Society of America and Acoustical Society of Japan [17] and the 2017 Spring Meeting of the Acoustical Society of Japan. Experiment 2 was reported at the 2017 Autumn Meeting of the Acoustical Society of Japan.

REFERENCES

[1] T. J. Vance, An Introduction to Japanese Phonology (SUNY Press, New York, 1987).
[2] T. Arai and S. Greenberg, “The temporal properties of spoken Japanese are similar to those of English,” Proc. Eurospeech ’97, Vol. 2, Rhodes, pp. 1011–1014 (1997).
[3] E. Dupoux, K. Kakahi, Y. Hirose and C. Pallier, “Epenthetic vowels in Japanese: A perceptual illusion?,” J. Exp. Psychol.: Hum. Percept. Perform., 25, 1568–1578 (1999).
[4] K. Maekawa and H. Kikuchi, “Corpus-based analysis of vowel devoicing in spontaneous Japanese: An interim report,” in Voicing in Japanese, J. van de Weijt, K. Nanjo and T. Nishihara, Eds. (Mouton de Gruyter, 2005), pp. 205–228.
[5] L. Fais, S. Kajikawa, S. Amano and J. Werker, “Now you hear it, now you don’t: Vowel devoicing in Japanese infant-directed speech,” J. Child Lang., 37, 319–340 (2010).
[6] S. Imaiizumi, A. Hayashi and T. Deguchi, “Listener adaptive characteristics of vowel devoicing in Japanese dialogue,” J. Acoust. Soc. Am., 98, 768–778 (1998).
[7] D. Kewley-Port, T. Z. Burkle and J. H. Lee, “Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners,” J. Acoust. Soc. Am., 122, 2365–2375 (2007).
[8] K. Yasu, T. Arai, K. Kobayashi and M. Shindo, “Identification of voiceless fricatives/affricates by elderly listeners: Effect of degradation of auditory properties,” J. Acoust. Soc. Jpn. (J), 68, 501–512 (2012) (in Japanese).
[9] H. Kawata, T. Arai, K. Yasu, K. Kobayashi and M. Shindo, “Perception of geminate fricative consonants in Japanese by elderly people: Effects of degradation of auditory properties,” J. Acoust. Soc. Jpn. (J), 71, 653–660 (2015) (in Japanese).
[10] S. Amano and T. Kondo, Nihongo no Goitokusei [Lexical Properties of Japanese], NTT database (Sanseido, Tokyo, 2008) (in Japanese).
[11] NHK, Nihongo Hatsuon Akusento Jiten (NHK Publishing, Tokyo, 1985) (in Japanese).
[12] A. Strouse, D. H. Ashmead, R. N. Ohde and D. W. Grantham, “Temporal processing in the aging auditory system,” J. Acoust. Soc. Am., 104, 2385–2399 (1998).
[13] M. F. Matsui, “The interaction between pole frequency transition and VOT on the perception of fricatives,” J. Fac. Lett. Kobe Shoin Women’s Univ., 2, 19–34 (2013) (in Japanese).
[14] P. Boersma and D. Weenink, Praat: Doing phonetics by computer [Computer program]. Version 6.0.22 (retrieved 15 November 2016).
[15] E. Yanagisawa and T. Arai, “Effects of formant transition and intensity damping of preceding vowel off-glide on perception of Japanese geminate consonant Sokuon,” J. Acoust. Soc. Jpn. (J), 71, 505–515 (2015) (in Japanese).
[16] E. Iwagami, T. Arai, K. Yasu, E. Yanagisawa and K. Kobayashi, “Perception of Japanese geminate consonant by elderly and young listeners,” Proc. Autumn Meet. Acoust. Soc. Jpn., pp. 605–608 (2016) (in Japanese).
[17] E. Iwagami, T. Arai, K. Yasu, E. Yanagisawa and K. Kobayashi, “Perception of geminate consonants and devoiced vowels in Japanese,” J. Acoust. Soc. Am., 140, 3439 (2016).
Eri Iwagami received her M.A. degree from Faculty of Humanities, Hosei University, Tokyo, Japan, in 2015. She is currently a doctoral student in the Faculty of Science and Technology at Sophia University, Tokyo, Japan.

Takayuki Arai received the B.E., M.E. and Ph.D. degrees in electrical engineering from Sophia Univ., Tokyo, Japan, in 1989, 1991 and 1994, respectively. In 1992–1993 and 1995–1996, he was with Oregon Graduate Institute of Science and Technology (Portland, OR, USA). In 1997–1998, he was with International Computer Science Institute (Berkeley, CA, USA). He is currently Professor of the Department of Information and Communication Sciences, Sophia Univ. In 2003–2004, he is a visiting scientist at Massachusetts Institute of Technology (Cambridge, MA, USA). His research interests include signal processing, acoustics, speech and hearing sciences, spoken language processing, and acoustic phonetics.

Kei Yasu received the Ph.D. degree from Faculty of Science and Technology, Sophia University, Tokyo, Japan, in 2013. He is a research associate at Department of Industrial Information, Faculty of Industrial Technology, Tsukuba University of Technology, Japan. He is interested in neuroscience and behavioral aspect of hearing impairment, stuttering, speech perception, and real-time speech processing.

Kei Kobayashi is a research fellow/honorary academic at Acoustics Research Centre in The University of Auckland, New Zealand. He received his Ph.D. in Electronics Engineering from Sophia University, Japan, in 2007, after 10 years working on consumer electronics in Sony corporation. He has developed various auditory software and system and intellectual properties at Audiology section (2007–2016) and the Acoustics Research Centre (2017–current). His current research interest is communication acoustics for hearing instruments, that includes digital signal processing, acoustics, psychoacoustics and speech sciences. He is a profound hearing loss and hearing aids user.