Acoust. Sci. & Tech. **31**, 5 (2010)

PAPER

Processing of consonant clusters by Japanese native speakers: Influence of English learning backgrounds[1]

Hinako Masuda* and Takayuki Arai†

Department of Science and Technology, Sophia University, 7–1 Kioi-cho, Chiyoda-ku, Tokyo, 102–8554 Japan

(*Received 10 July 2009, Accepted for publication 22 March 2010*)

Abstract: Previous research has revealed that Japanese native speakers are highly likely to both perceive and produce epenthetic vowels between consonants. The goal of the present study is to investigate the influence of English learning backgrounds in the perception and production of consonant clusters by Japanese native speakers. In Experiment 1, a forced-choice AXB task to identify VC(u)CV is assigned to 17 highly fluent Japanese-English bilinguals and 22 Japanese monolinguals. Results show that monolinguals made significantly more errors than bilinguals. In Experiment 2, the influence of English proficiency on the production of consonant clusters, and the effect of consonant voicing on vowel epenthesis are investigated. The epenthetic vowels are acoustically analyzed and categorized into three degrees: full, partial and no epenthesis. The voicing combinations of the consonant clusters are C[+voice]-C[+voice], C[−voice]-C[+voice], and C[−voice]-C[−voice]. Results show that monolinguals inserted more epenthetic vowels than bilinguals, and that the influence of consonant voicing was stronger in monolinguals than bilinguals. Furthermore, monolinguals' epenthetic vowels between C[−voice]-C[+voice] and C[−voice]-C[−voice] tended to become devoiced than bilinguals. This result suggests a stronger L1 influence on monolinguals. The results of the two experiments thus suggest that the English proficiency influences the perception and production of consonant clusters.

Keywords: L2 perception, L2 production, Vowel epenthesis, Consonant cluster, Proficiency

PACS number: 43.71.Hw, 43.70.Kv [doi:10.1250/ast.31.320]

1. INTRODUCTION

Numerous studies of foreign language acquisition have revealed that foreign language learning is influenced by the learners' native language [1,2]. A well-known example is the difficulty of Japanese natives to discriminate English /r/ and /l/ [3–5]. Such difficulty is caused by the phonemic differences of Japanese and English. While the Japanese phoneme /r/ is pronounced as an apico-alveolar tap [ɾ] [6], English liquid phonemes /r/ and /l/ are pronounced as alveolar [ɹ] and lateral [l] approximants although many variations exist among various dialects of English [7]. The two liquid phonemes of English are thus perceptually assimilated into one Japanese liquid sound [8].

Difficulties are observed not only in the phonemic differences, but also in the differences in syllable structures. For example, Japanese are likely to produce the English word "strike" as /sutoraiku/ with a vowel inserted after a consonant in order to avoid syllables to end with a consonant [9]. Generally, [u] is inserted after a consonant except after [t, d] and [ʧ, ʤ] in which case the vowels inserted are [o] and [i], respectively. Insertion of [u] is preferred because it is the shortest vowel in duration to be pronounced in Japanese, and it is also most likely to become devoiced along with [i] [9]. Tajima *et al.* [10] analyzed the epenthetic vowels in English words with word-final two-consonant clusters and word-final single consonant by Japanese native speakers and English native speakers. Their results showed that for Japanese native speakers, epenthetic vowels increased in slow rate speech than normal rate speech, and that epenthesis occurred more within voiced

[1]This paper is a summary of our previous works; Masuda and Arai "Perception and production of consonant clusters in Japanese-English bilingual and Japanese monolingual speakers," *Proc. Interspeech*, pp. 1988–1991 (2008), Masuda and Arai "Vowel epenthesis between stop consonants by Japanese native speakers with different English learning backgrounds," *Proc. Autumn Meet. Acoust. Soc. Jpn.*, pp. 323–326 (2008), and Masuda and Arai "Effect of consonant voicing on the vowel epenthesis of Japanese learners of English," *Proc. Spring Meet. Acoust. Soc. Jpn.*, pp. 397–400 (2009). Additional analyses are also introduced.

*e-mail: h-masuda@sophia.ac.jp
†e-mail: arai@sophia.ac.jp

consonants than voiceless consonants. Furthermore, Fujimoto and Funatsu [11] compared vowel epenthesis in Japanese native speakers and German native speakers, and found that Japanese tended to produce epenthetic vowels between consonants while German did not.

Vowel epenthesis by Japanese natives is seen not only in production but also in perception. Since Japanese syllables almost always end with a vowel with exceptions of nasals and geminates [9], consonant clusters are illegal in Japanese phonotactics. Dupoux *et al.* [12] conducted four identification and discrimination experiments to investigate Japanese and French listeners' differences in the ability to perceive whether an epenthetic vowel is present or absent between consonants. The results of their experiments revealed that Japanese listeners perceived "illusory vowels" between consonants even when there was no vowel inserted, and that Japanese listeners were unable to discriminate VCCV and VCVCV. They therefore concluded that while French are able to identify and discriminate epenthetic vowels, Japanese cannot, and that the abilities are based on the phonotactics of their native language.

Although both perception and production of vowel epenthesis by Japanese native speakers have been researched in the past, the effect of English proficiency has not been fully investigated. Beginner to intermediate level and advanced-level learners of English participated in the experiments, which will enable us to explore the degree of native language influence on the perception and production of foreign sounds. The present paper is a report of two experiments conducted by the authors, in which we focused on the English proficiencies of Japanese native speakers in the perception and production of consonant clusters [13–15].

The aim of Experiment 1 is to investigate whether advanced English learners' perceptual ability to discriminate VCCV and VCuCV is higher than that of beginner to intermediate level learners. If the exposure to English has a positive effect on the identification of a foreign phonotactic structure, advanced learners should achieve higher percentages of correct answers. The influence of consonant voicing on the perception of consonant clusters is also investigated. In Experiment 2, the productions of consonant clusters are acoustically analyzed to investigate the existence of epenthetic vowels. The two aims in Experiment 2 are to categorize the production of epenthetic vowels into three degrees (full, partial, and no epenthesis), and to investigate the effect of consonant voicing on vowel epenthesis. The consonant clusters in both Experiment 1 and 2 are thus divided into three groups of consonant voicing combinations: C[+voice]-C[+voice], C[−voice]-C[+voice], and C[−voice]-C[−voice]. The same participants from Experiment 1 participated in Experiment 2.

2. EXPERIMENT 1: PERCEPTION TEST

2.1. Participants

2.1.1. Group A: Advanced-level learners of English

Seventeen advanced-level learners of English (5 males, 12 females) took part in the experiment. All advanced learners have the experience of living in an English-speaking country for more than 2 years (mean 5.7 years) and had received education in English during their stay abroad. Ages varied from 19 to 25 years (mean 23.4 years), and their ages of acquisition ranged from 0 to 17 years of age. Eleven participants had learned a foreign language other than English: two had learned German, three had learned French, and six had learned Spanish.

2.1.2. Group B: Beginner to intermediate-level learners of English

Twenty-two beginner to intermediate-level learners of English (14 males, 8 females) took part in the experiment. Although none of the participants in Group B have experienced staying overseas for more than one month, all have received typical English education at Japanese junior high schools from about age 12 throughout university. Ages varied from 18 to 25 years (mean 20.8 years). Twelve participants had learned a foreign language other than English: three had learned German, two had learned French, three had learned Spanish, three had learned Chinese, and one had learned Italian. All but two participants had less than one year of L3 learning experience (mean 8 months): two participants had learned French for a total of 7 years, and Chinese for 1 year and 8 months.

2.2. Stimuli

The speaker of the stimuli used in the perception experiment is a Japanese-French female bilingual. She is also an experienced phonetician. The stimuli are all pseudo-words that are meaningless in both English and Japanese. The words were produced in the carrier phrase "Dites (pseudo word) deux fois" [Say (pseudo-word) two times].

Eighteen pairs of pseudo-words with $V_1C_1(u)C_2V_2$ sequence were recorded (see Table 1). All the pseudo-words used in Experiments 1 and 2 are from Dupoux *et al.* [12]. All consonant clusters are positioned in the middle of the word. The consonant clusters are categorized in terms of voicing [+/−voice]: C[+voice]-C[+voice], C[−voice]-C[+voice], and C[−voice]-C[−voice]. The 36 words (18 VCCV words + 18 VCuCV words) were produced three times; thus a total of 108 utterances were recorded. The recording of the stimuli took place in a sound-proof room. The sounds were recorded with a digital sound recorder (Marantz Professional Solid-state Recorder PMD671) and a microphone (SONY ECM-959DT) at a sampling fre-

Table 1 List of 36 pseudo-words (Dupoux *et al.* [12]).

VCCV	VCuCV
abmo, ebzo, abge, agmi, akmo, ashmi, ebza, egdo, ekshi, eshmo, ibdo, igna, ikma, ishto, obni, ogza, okna, oshta (18 words)	abumo, ebuzo, abuge, agumi, akumo, ashumi, ebuza, egudo, ekushi, eshumo, ibudo, iguna, ikuma, ishuto, obuni, oguza, okuna, oshuta (18 words)

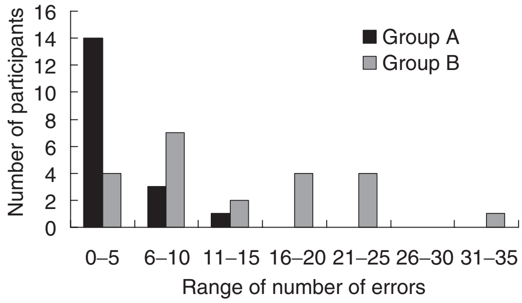

Fig. 1 Histogram of the range of number of errors in Groups A and B.

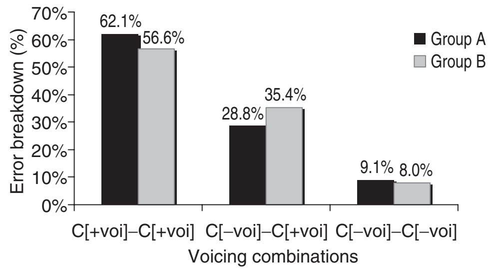

Fig. 2 Percentages of errors in each voicing combination by Groups A and B.

quency of 48 kHz. The target words were later saved as individual files. The speaker was instructed to read aloud the target words written on paper. The words were instructed to be read as English pseudo-words; therefore /ebuzo/, for example, was asked to be pronounced as [ebuzo] instead of [ebjuzo]. Instructions to the speaker were conducted in Japanese.

2.3. Procedure

An AXB forced-choice task was assigned to the 39 participants. All A, X, and B were three different utterances spoken by the Japanese-French bilingual speaker, for example, "ebzo(A), ebuzo-1(X), and ebuzo-2(B)" (ebuzo-1 and ebuzo-2 were two different tokens of "ebuzo"), in which case the correct answer would be "B" since X and B both have the VCuCV sequence. The order of the appearance of VCCV and VCuCV in X are randomized. The inter-stimulus interval (ISI) between A, X, and B are 300 milliseconds.

Two pairs of pseudo-words were used for practice trials (/ebzo-ebuzo/ and /abmo-abumo/ pairs), and the remaining 16 pairs were used for the main trials. Participants were instructed to listen to the sounds through headphones and to judge whether the first utterance (A) or the third utterance (B) was more similar to the second utterance (X). A personal computer and headphones (SONY MDR-CD900ST) were used. Participants were given 8 practice trials from the two pairs before proceeding to the main 192 trials. The 8 practice trials were not used for this experiment. Participants did not receive any feedback in neither the practice trials nor the main trials. The 192 trials were divided into 3 blocks, and the participants could take breaks every 64 trials, if needed.

2.4. Results and Discussion

The distribution of the number of errors differed greatly between the two groups (see Fig. 1). The average number of errors made by Group A was 3.8 out of 192 trials (approximately 2%), whereas those made by Group B were 13.1 (approximately 6.8%). The number of errors made by Group A ranged from 0 to 12, and the number of errors made by Group B ranged from 2 to 35. The Mann-Whitney test found a significant difference between the two groups' average number of errors ($p < 0.05$). The difference in the number of errors made by the two listener groups is caused presumably due to English proficiency. The results of this experiment suggest that Group A have an advantage of discriminating VCCV and VCuCV compared to Group B.

The total numbers of errors made by participants in Groups A and B are 66 and 288, respectively. The breakdown of those errors classified by the consonant voicing combinations is shown in Fig. 2. Within the 2% and 6% of errors made by Group A and B, approximately 62% and 56%, respectively, were C[+voice]-C[+voice]. Although the number of errors in the two listener groups differed significantly, the breakdown of the errors revealed that the consonant voicing on the perception of consonant clusters influenced the two listener groups in a similar way (no significant difference, $p = 0.5$). As the graph shows, the percentages of errors increased with the emergence of voiced consonants in both Groups A and B.

There were two participants in Group B who had more than one year of L3 learning experience; however, their numbers of errors were seven and nine which are among the highest distribution (see Fig. 1). Thus, the length of L3 learning did not have an influence on the results of the present experiment.

Table 2 List of target words produced by participants that are categorized in terms of consonant voicing.

	Target words		
C[+voice]-C[+voice]	abge	egdo	ibdo
C[−voice]-C[+voice]	akmo	ashmi	okna
C[−voice]-C[−voice]	ekshi	ishto	oshta

Table 3 Criteria for the measurement of the degree of vowel epenthesis.

Full	Periodic waveform, voice bar, glottal pulse
Partial	Pseudo-periodic waveform and voice bar, devoiced short vowel
None	No waveform nor formants

However, the average numbers of errors differed between participants in Group B who had learned an open-syllable language or a closed-syllable language as their L3. Participants in Group B who had learned an open-syllable language (Spanish or Italian) made 23 errors on average (approximately 12%) while those who had learned a closed-syllable language (German, French, or Chinese), like English, made 11 errors on average (approximately 6%). Such results indicate that learning a closed-syllable language, either as L2 or L3, may have an influence on the perception of consonant clusters.

3. EXPERIMENT 2: PRODUCTION TEST

3.1. Stimuli

The nine words are a subset of the stimuli in Experiment 1. There are three voicing combinations, and three words per combination are selected (see Table 2). C[+voice]-C[−voice] combination was not in Dupoux *et al.*'s [12] list.

3.2. Recording of Material

The nine pseudo-words, each produced by the same 17 advanced-level learners and 22 beginner to intermediate-level learners as in Experiment 1, were recorded with a digital sound recorder (Marantz Professional Solid-state Recorder PMD671) and a microphone (SONY ECM-959DT) at a sampling frequency of 48 kHz. A total of 351 utterances were analyzed (39 participants × 9 pseudo-words). The recordings took place in a sound-proof room. The participants were instructed to read aloud the words written on paper. They were told that the words are meaningless, and that there was no right or wrong in how they pronounced each word. Instructions to participants were conducted in Japanese.

3.3. Acoustical Analysis

The epenthetic vowels between consonants were categorized by the degree of epenthesis: full epenthesis, partial epenthesis, and no epenthesis. Partial epentheses are divided into voiced and devoiced vowels. The criteria for each category are shown in Table 3, and examples of full, partial, none, and devoiced epenthetic vowels are shown in Figs. 3–6, respectively. The criteria for the measurement of a devoiced vowel are pseudo-periodic waveform and no

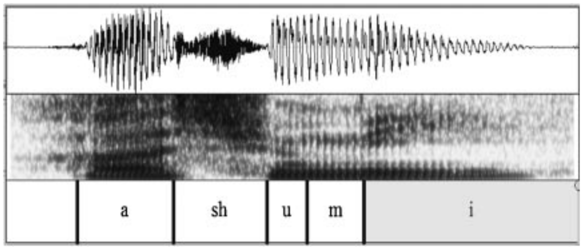

Fig. 3 Example of full epenthesis in "ashmi."

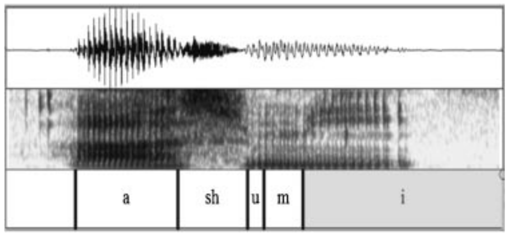

Fig. 4 Example of partial epenthesis in "ashmi."

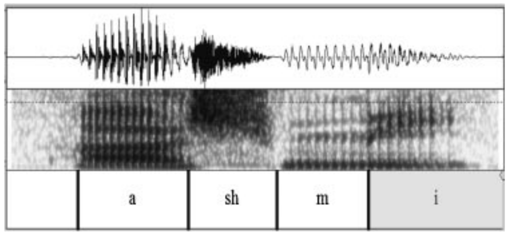

Fig. 5 Example of no epenthesis in "ashmi."

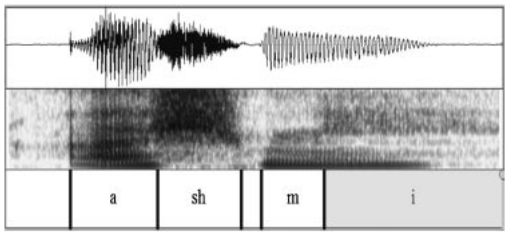

Fig. 6 Example of devoiced, partial epenthesis in "ashmi."

voice bar. Devoiced vowels could be distinguished from a neighboring voiceless consonant due to faint but noticeable traces of vowel-like formants. No epenthesis has no waveform, although no waveform with durational gap was classified as partial epenthesis. Acoustical analyses were performed using Praat [16]. Analyses were mainly conducted by a single phonetically-trained analyst, and the results were discussed with another phonetically-trained analyst when there was hesitation in the judgment. After the 351 utterances (39 participants × 9 words) were categorized into full, partial, and no epenthesis, 50 utterances were randomly chosen for quantitative analysis. The length of full vowels were between approximately 60–90 milliseconds which account for approximately 14–22% within a word, and partial vowels were between approximately 15–45 milliseconds which account for approximately 3–5% within a word. Durational gaps of approximately 45 milliseconds are regarded as partial epenthesis, while durational gaps of approximately 25–30 milliseconds or less are regarded as no epenthesis. There was no overlap among the categories in the 50 samples that were analyzed; therefore, the remaining samples are also considered to have the same acoustic characteristics as the 50 samples. 20 samples of devoiced vowels were also randomly chosen which were double checked by the first analyst and also by a different analyst. Consistent results were obtained between the two analysts; therefore, the remaining samples are also expected to have been through a valid analysis.

3.4. Results and Discussion

3.4.1. Degree of vowel epenthesis

The degree of vowel epenthesis produced by Groups A and B differed significantly (chi-square test, $p < 0.001$). Table 4 shows the numbers of utterances in each level of epenthesis. The highest percentage of the degree of epenthesis was full epenthesis for Group B and partial epenthesis for Group A (shaded in gray).

In previous research [10,11], participants whose native language is a closed-syllable language (English native speakers in Tajima *et al.*, and German native speakers in Fujimoto & Funatsu) were unlikely to insert any vowels between consonants. Although the results in Table 4 cannot be compared directly with their studies due to differences in participants, experimental methods, and target words, Group A in the present study inserted less vowels compared to Group B, and Group A inserted more vowels compared to English and German native speakers in Tajima *et al.* and Fujimoto & Funatsu. Such results suggest that even advanced learners of English do not perform similar to English and German native speakers.

Table 4 Number of utterances in each epenthesis degree.

	Group A ($N = 153$)	Group B ($N = 198$)
Full	2 (1.3%)	108 (54.5%)
Partial	106 (69.3%)	71 (35.9%)
None	45 (29.4%)	19 (9.6%)

(chi-square test, $p < 0.001$)

The aim of Experiment 2 was to assess the presence and absence of vowel epenthesis in Japanese native speakers, and to examine the influence of English proficiency on the degree of vowel epenthesis. The results of the previous studies and Experiment 2 in the present study suggest that the English proficiency does have an influence on how much vowel one inserts between consonants: generally speaking, participants who are beginner to intermediate-level in English proficiency insert full epenthesis, advanced learners of English insert partial epenthesis, and English native speakers insert no epenthesis.

3.4.2. Effect of consonant voicing

The numbers of utterances for each degree of vowel epenthesis in the three voicing combinations of the consonants are shown in Tables 5–7. In C[+voice]-C[+voice] and C[−voice]-C[+voice], Group B inserted full vowels, whereas Group A inserted no vowels the most in C[+voice]-C[+voice] and partial vowels the most in

Table 5 Numbers of utterances for each degree of vowel epenthesis in C[+voice]-C[+voice] by Groups A and B.

	Group A ($N = 51$)	Group B ($N = 66$)
Full	1 (1.9%)	51 (77.3%)
Partial	11 (21.6%)	6 (9.1%)
None	39 (76.5%)	9 (13.6%)

(chi-square test, $p < 0.001$)

Table 6 Numbers of utterances for each degree of vowel epenthesis in C[−voice]-C[+voice] by Groups A and B.

	Group A ($N = 51$)	Group B ($N = 66$)
Full	0 (0.0%)	48 (72.7%)
Partial	49 (96.1%)	12 (18.2%)
None	2 (3.9%)	6 (9.1%)

(chi-square test, $p < 0.001$)

Table 7 Numbers of utterances for each degree of vowel epenthesis in C[−voice]-C[−voice] by Groups A and B.

	Group A ($N = 51$)	Group B ($N = 66$)
Full	1 (1.9%)	9 (13.6%)
Partial	46 (90.2%)	53 (80.3%)
None	4 (7.8%)	4 (6.1%)

(chi-square test, $p = 0.03$)

C[−voice]-C[+voice]. In C[−voice]-C[−voice], both groups of speakers inserted partial vowel the most. However, while approximately 90% of Group A inserted partial vowels and only 2% inserted full vowels, 80% of Group B inserted partial vowels and 10% inserted full vowels. Such results suggest that advanced learners either insert a lower degree of vowel epenthesis than beginner to intermediate-level learners and/or beginner to intermediate-level learners are likely to insert more full vowels than advanced learners.

4. GENERAL DISCUSSION

The results of the two experiments in the present study show that both perceptual and productive abilities of consonant clusters differ between Japanese native speakers with advanced and beginner to intermediate-level English proficiency. In Experiment 1, the perceptual differences of VCCV-VCuCV by Groups A and B were investigated by assigning 192 AXB trials. Although both groups' percentages of correct responses achieved higher than 90%, the numbers of errors differed significantly between the two groups. This result shows that English proficiency has an effect on the discrimination ability of VCCV-VCuCV. However, a similar tendency in errors was observed in the two groups regarding consonant voicing combinations.

The closest experiments to Experiment 1 in the present study are that of Dupoux *et al.* They carried out two ABX discrimination tasks where Japanese and French natives were asked whether X was more similar to A or B, and considered the effect of speaker change (same-talker or different-talker). Their results showed that 1) French natives could discriminate VCCV and VCuCV by approximately 96% while Japanese natives could discriminate by 86% in same-talker stimuli (where X was acoustically identical to either A or B), and 2) French natives could discriminate VCCV and VCuCV by approximately 95% while Japanese natives could discriminate by 81% in different-talker stimuli (where A, B, and X were spoken by different talkers). Given that the different-talker experiment in Dupoux *et al.* is closer to Experiment 1 in the present study in that X is acoustically different to A or B, the difference in the correct rate of the two experimental results is approximately 10%. The difference between the results of Experiment 1 and Dupoux *et al.*'s experiment may be due to the difference in the stimuli; Experiment 1 in the present study were naturally produced consonant clusters, while Dupoux *et al.* had used consonant clusters that were digitally edited.

In Experiment 2, the voicing combination that had the highest degree (full epenthesis) and percentage of vowel epenthesis was C[+voice]-C[+voice] at 77.3% for Group B (see Table 5), although the results of C[−voice]-C[+voice] resembles that of C[+voice]-C[+voice] as well at 72.7%. On the contrary, C[+voice]-C[+voice] had the highest percentage of no epenthesis for Group A. In both speaker groups, vowel epenthesis occurred more frequently in C[−voice]-C[+voice] than in C[−voice]-C[−voice]: Group A produced a higher percentage of partial vowel in C[−voice]-C[+voice] than in C[−voice]-C[−voice], and Group B produced full epenthesis the most in C[−voice]-C[+voice] while partial epenthesis was the highest in C[−voice]-C[−voice]. Thus the results indicate that voicing of consonants strongly effect vowel epenthesis in beginner to intermediate-level learners, but not as strong in advanced learners. Although the present experiment focused on the voicing of consonant clusters alone, there were changes in the manner of articulation as well. It is of great importance to point out that the influence of manner of articulation of the consonant clusters may also have played a role in vowel epenthesis. The results of the present experiment cannot provide strong evidence; therefore, further investigation is needed to clarify other factors that may be influencing the degree of vowel epenthesis.

Table 8 Number of utterances with devoiced partial epenthesis.

	Group A ($N = 106$)	Group B ($N = 71$)
C[+voice]-C[+voice]	2 (18%)	1 (17%)
C[−voice]-C[+voice]	47 (96%)	9 (75%)
C[−voice]-C[−voice]	46 (100%)	53 (100%)

The numbers of utterances with devoiced partial epenthesis are shown in Table 8. According to Vance [6], vowel devoicing in Japanese is most common in /i/ and /u/, especially in between voiceless consonants, or after a voiceless consonant followed by a pause. Analyses of spontaneous speech by Japanese speakers in Arai *et al.* [17] also revealed that vowel devoicing is common not only in high vowels surrounded by voiceless consonants, but also in non-typical environments such as high vowels preceded by a voiceless consonant and followed by a voiced consonant. The results in Table 8 correspond well with the two studies above: partial vowel insertions in C[−voice]-C[+voice] and C[−voice]-C[−voice] were highly likely to become devoiced in both speaker groups, although the number of utterances of devoiced partial vowels were low in C[+voice]-C[+voice]. The most common degree of epenthesis in C[−voice]-C[−voice] was partial in both speakers, and all partial insertions were devoiced. The inserted vowel was all /u/ in the present study, except for one utterance [ekishi] by a participant in Group B. The results of vowel devoicing show that both advanced learners and beginner to intermediate-level learners' realization of epenthetic vowels were similarly influenced by the phenomenon of vowel devoicing in Japanese.

Table 9 Number of utterances with devoiced and spirantized preceding consonant in C[+voice]-C[+voice].

	Group A ($N = 51$)	Group B ($N = 66$)
Devoiced	23 (45%)	8 (12%)
Spirantized	7 (13%)	22 (33%)

Variations in the articulation of preceding consonants were also observed. According to Jones [18], voiced stop consonants in English such as /b, g/ may be weakened, and sometimes be pronounced as devoiced consonants [b̥, g̊] in initial and final positions, or when preceded by a voiceless consonant. On the other hand, voiced stop consonants in Japanese such as /b, g/ may be weakened and spirantized, and sometimes be pronounced as fricatives [β, ɣ] [17,19]. The preceding consonants of the voiced consonant cluster pairs (/b/ and /g/ in /abge, egdo/, respectively) were positioned at syllable codas, and thus tended to become devoiced for Group A presumably due to influence of English consonant lenition, and spirantized for Group B presumably due to influence of Japanese consonant lenition (see Table 9). Such results indicate that the realizations of the voiced preceding consonants for advanced learners are influenced more strongly by English, and beginner to intermediate-level learners are influenced more strongly by Japanese.

The results of the two experiments conducted by the present study suggest that the perception and production abilities of Japanese natives with advanced and beginner to intermediate-level English proficiency are different, and that advanced learners outperformed beginner to intermediate-level learners in both perception and production. This is probably caused by the greater L1 interference for beginner to intermediate-level learners. However, beginner to intermediate-level learners who had experienced learning a closed-syllable language other than English made less errors in perception compared to those who had experienced learning only English as a closed-syllable language. Thus, more exposure to closed-syllable languages may result in higher VCCV-VCuCV discrimination ability for Japanese native speakers with beginner to intermediate-level English proficiency. While learning more than two closed-syllable languages seems to be advantageous in the perceptual domain, it did not have a positive influence on the production domain, at least in the present study. The results of the present study may hence suggest that perception is acquired before production in second language learning.

5. CONCLUSION

The present study investigated the abilities of advanced learners and beginner to intermediate-level learners of English to identify VC(u)CV in a forced-choice AXB task, and the degree of vowel epenthesis in consonant cluster production. The results of the perception experiment found a significant difference between advanced learners' and beginner to intermediate-level learners' numbers of errors. Such result indicates that Japanese native speakers' ability to perceive consonant clusters varies with one's English learning background. However, the effect of consonant voicing on error occurrence was similar in both speaker groups.

We focused mainly on two factors in the analyses of consonant cluster production: the degree of vowel epenthesis and the influence of voicing of consonants. The epenthetic vowels were divided into three categories: full, partial and no epenthesis. The analyses on the degree of vowel epenthesis were performed on three voicing combinations: C[+voice]-C[+voice], C[−voice]-C[+voice], and C[−voice]-C[−voice]. Analyses revealed a significant difference in the participant groups' degree of epenthesis in all voicing combinations. The difference between Groups A and B was especially distinct in C[+voice]-C[+voice]. Also, the percentages of devoiced partial epenthesis and the realization of the preceding consonant differed between the two speaker groups.

The findings from the two experiments performed in the present study suggest that there are differences in both perception and production of consonant clusters by Japanese native speakers with different English learning backgrounds. However, further investigation is needed to clarify the underlying L1 influence on the speakers' performances.

ACKNOWLEDGMENT

This study work was partially supported by Sophia University Open Research Center from MEXT. The authors would like to thank Shigeko Shinohara and Kensaku Yoshida for their insightful comments.

REFERENCES

[1] C. Pallier, A. Christophe and J. Mehler, "Language-specific listening," *Trends Cogn. Sci.*, **1**, 129–132 (1997).
[2] N. Sebastian-Galles, "Cross-language speech perception," in *The Handbook of Speech Perception*, D. B. Pisoni and R. E. Remez, Eds. (Blackwell Publishing Ltd., Oxford, 2005), pp. 546–566.
[3] H. Goto, "Auditory perception by normal Japanese adults of the sounds "L" and "R"," *Neuropsychologia*, **9**, 317–323 (1971).
[4] K. Miyawaki, W. Strange, R. Verbrugge, A. M. Liberman, J. J. Jenkins and O. Fujimura, "An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English," *Percept. Psychophys.*, **18**, 331–340 (1975).
[5] N. Takagi, "The limits of training Japanese listeners to identify English /r/ and /l/: Eight case studies," *J. Acoust. Soc. Am.*,

111, 2887–2896 (2002).

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

[7] P. Ladefoged, *Vowels and Consonants, An Introduction to the Sounds of Languages*, 2nd ed. (Blackwell Publishing Ltd., Oxford, 2005).

[8] R. Komaki, R. Akahane-Yamada and Y. Choi, "Effects of native language on the perception of American English /r/ and /l/: Cross-language comparison between Korean and Japanese," *Tech. Rep. IEICE*, SP99-45, pp. 39–46 (1999).

[9] H. Kubozono, *Nihongo no Onsei* (Iwanami, Tokyo, 1999) (in Japanese).

[10] K. Tajima, D. Erickson and K. Nagao, "Phonetic analysis of vowel epenthesis in native Japanese speakers' production of English words," *Proc. 14th General Meeting of the PSJ*, pp. 195–200 (2000).

[11] M. Fujimoto and S. Funatsu, "Vowel epenthesis in consonant clusters by Japanese speakers," *IEICE Tech. Rep.*, SP2007-204, pp. 105–109 (2008) (in Japanese).

[12] E. Dupoux, K. Kakehi, Y. Hirose, C. Pallier and J. Mehler, "Epenthetic vowels in Japanese: A perceptual illusion?," *J. Exp. Psychol.: Hum. Percept. Perform.*, **25**, 1568–1578 (1999).

[13] H. Masuda and T. Arai, "Perception and production of consonant clusters in Japanese-English bilingual and Japanese monolingual speakers," *Proc. Interspeech*, pp. 1988–1991 (2008).

[14] H. Masuda and T. Arai, "Vowel epenthesis between stop consonants by Japanese native speakers with different English learning backgrounds," *Proc. Autumn Meet. Acoust. Soc. Jpn.*, pp. 323–326 (2008) (in Japanese).

[15] H. Masuda and T. Arai, "Effect of consonant voicing on the vowel epenthesis of Japanese learners of English," *Proc. Spring Meet. Acoust. Soc. Jpn.*, pp. 397–400 (2009).

[16] P. Boersma and D. Weenink, "Praat: doing phonetics by computer (Version 5.0.09) [Computer program]." Retrieved April 3, 2008, from http://www.praat.org/

[17] T. Arai, N. Warner and S. Greenberg, "Analysis of spontaneous Japanese in a multi-language telephone-speech corpus," *Acoust. Sci. & Tech.*, **28**, 46–48 (2007).

[18] D. Jones, *An Outline of English Phonetics*, 9th ed. (Cambridge University Press, Cambridge, 1989).

[19] S. Kawakami, *Nihongo Onsei Gaisetsu* (Ohfu, Tokyo, 1977) (in Japanese).

Hinako Masuda received the B.A. and M.A. degrees in linguistics from Sophia Univ., Tokyo, Japan, in 2006 and 2008, respectively. She is currently working on her PhD at the Graduate School of Science and Technology, Sophia Univ. Her research interests include second language perception and production, language acquisition, and bilingualism.

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 was a visiting scientist at Massachusetts Institute of Technology. His research interests include acoustics, speech communication (speech and hearing sciences), spoken language processing, and signal processing.