

Consonant cluster production in Japanese learners of English

Yoshiho Shibuya¹, Donna Erickson²

¹ Department of General Education, Kanazawa Medical University, Japan

² Department of Linguistics, Showa Music University, Japan

yshibuya@Kanazawa-med.ac.jp, ericksondonna2000@gmail.com

Abstract

Japanese speakers often face difficulty in producing complex syllable onsets in English and insert an extra vowel. We examined whether the vowel inserted by Japanese speakers was epenthetic (phonological) or excrescent (phonetic).

The acoustic data suggested that an L1 phonological process was involved in vowel insertion by Japanese speakers with lower-level English competency, because the inserted vowels were similar to vowels in Japanese. More advanced speakers' results, on the other hand, suggested that phonetics may be involved. The articulatory data from this pilot study with one speaker supported the findings of the acoustic data, suggesting that both phonetics and phonology affect Japanese speaker's vowel insertion in a complex way. This paper is based on work previously reported in [1].

Index Terms: L2 acquisition, acoustic and articulatory phonetics, phonology/phonetics interface

1. Introduction

Japanese speakers (JS) have difficulty producing English consonant clusters (CC) and often use some 'repair' strategy in producing CC. We approach this question very simply: Was this repair strategy a phonological or phonetic one? If phonological, the inserted vowel would create a permissible Japanese CV syllable type. On the other hand, if speakers produce an intrusive V-like (schwa) transition but not a full CV syllable, the repair would be classified as a phonetic process. Since JS do not have much practice in the articulatory coordination of CC, they may produce their consonant gestures with insufficient temporal overlap, compared to native speakers (NS) of English, and therefore phonetically produce a short vowel-like transition between the consonants. Although many studies have been conducted about the acquisition of CC, little acoustic or articulatory examination of phonetic data has been conducted. Without such examination, however, it is difficult to know how or to what extent L1 phonology, or simply phonetics is involved in the process. Two experiments were conducted in the present study and the results showed that there seems to be influence of both phonotactics and gestural timing patterns from Japanese. The points argued here are that for JS's production of English CC's, phonology and phonetics are both involved and that they interact in a very subtle way. We aimed to provide some phonetic evidence to contribute to the understanding of the issue of the phonology/phonetics interface and to L2 acquisition. Our intent is to explore this interface further in order to develop teaching materials for helping JS to produce English syllable types without inserting vowels between consonants.

2. Experiments

Acoustic and articulatory experiments were conducted to examine the following questions:

(1) How do native JS realize English CC? Do they delete consonants or insert vowels?

(2) If they do produce intrusive vowels, is the vowel inserted phonetically (excrescence) or phonologically (epenthesis)?

Following [2], the evidence for L1 phonological insertion is listed in (a) below, and for phonetic, in (b) below.

- a) Evidence for L1 phonological processes
 - ◆ similar to underlying Japanese Vs (either /o/ or /u/ depending on the preceding C)
 - ◆ respect Japanese phonotactics (/o/ after /t, d/ and /u/ after /s, b/)
 - ◆ be consistent in quality and duration
 - ◆ participate in phonological processes: may be stressed or accented
- b) Evidence for phonetic processes
 - ◆ more schwa-like, and not the same as a Japanese V
 - ◆ do not respect Japanese phonotactics
 - ◆ not consistent in quality or duration
 - ◆ never stressed or accented (do not participate in phonological processes)

2.1. Acoustic experiment

The objective was to test the following hypotheses: (1) JS produce a vowel or a vowel-like transition in their production of consonant clusters; they do not delete consonants. (2) If JS do insert vowels phonologically, /o/ is predicted to appear after /t, d/ while /u/ is predicted after /s, b/, following Japanese loanword phonology. It was therefore predicted that the formant frequencies will be different, i.e., F1 will be higher and F2 will be lower, for the vowel that appears after /t, d/, compared with those after /s, b/. (3) If the intrusive vowel is accented, it is a phonological vowel, since JS tend to rely on pitch difference and not durational difference for stressing English vowels. Thus, relatively high F0 will indicate epenthesis (a phonological process).

2.1.1. Subjects and materials

Three groups of subjects took part in this experiment: (1) five female Japanese students learning English at a junior college (J1-J5); (2) three NS of North American English (N1-N3); (3) one Japanese female who had lived abroad for several years (J6). JS's English levels were assessed by experienced language teachers and only J6 was thought to be an advanced speaker of English. J1-J5 were judged as having intermediate-level proficiency in English. In this paper, only groups (1) and (3) will be reported on.

Two kinds of English pseudowords, one with CC- and CCC-clusters as onsets and the other with CV-syllables, were placed in English sentences, along with the English instruction. The English pseudowords were minimal pairs such as /trok/-/truk/, /snok/-/snuk/ along with /sunuku/ and /toroku/, etc, as shown in Table 1 below. The latter types of words were also used as Japanese pseudowords and placed in Japanese sentences. In this paper, only the English pseudowords, produced by the 6 JS speakers are reported.

In order to ascertain the acoustic characteristics of the vowels that appear between the consonant clusters, /t, d/ vs. /s, b/ consonants in the first position of the CC(C) structures were compared. For subjects J1-J5, utterances were recorded in a quiet room using a good quality tape recorder and either a head-worn microphone or a hand-held microphone. They repeated the sentences five times, so for JS, the tokens collected were (20 English sentences + 10 Japanese sentences) x 5 participants x 5 repetitions (a total of 750 tokens), and for NS of English, 20 English sentences x 3 participants x 5 repetitions (a total of 300 tokens). Only the English pseudowords, produced by JS speakers are reported here. For complete results, the reader is referred to [1].

Table 1. English pseudowords

CC-type	with a CC(C)	with CV(CV)	<V>
/t/ + sonorant	trok truk	toroku toruku	/o/
/d/ + sonorant	drok druk	doroku doruku	/o/
/s/+ sonorant	snok snuk	sunoku sunuku	/u/
/b/+ sonorant	brok bruk	buroku buruku	/u/
/st/+ sonorant	strok struk	sutoroku sutoruku	/u/after /s/ /o/ after /t/

2.1.2 Acoustic measurements

Two different sets of data were used for acoustic measurements. The first set of data was obtained from recordings of five Japanese subjects; the second set was taken from the acoustic data of the EMA articulatory experiment (J6). For the six JS, measurements were made for the duration (of the interconsonantal interval), F0, F1, and F2 (if any of these were present). Formants were computed by LPC (linear predictive coding) and F0, by ESPS (Entropic Speech Processing System). When there was audible sound, visible F1 and F2, a vowel-like complex (periodic) pattern in the waveform and an increase in amplitude, it was considered that there was a vowel or a vowel-like transition.

2.2. Articulatory experiment

In order to compare acoustic data with articulatory data, an Electromagnetic Articulograph (EMA) experiment was done. Specifically, we were interested in the production of vowel insertion by the advanced JS speaker (J6). Also, we wished to examine the articulatory nature of the devoiced vowel, for which acoustic measurements could not be made.

2.2.1 Subjects and materials

The subject was a female speaker of Japanese, identified as J6 in the acoustic experiment. Target sequences were the same as the ones used in the acoustic experiment.

2.2.2 Articulatory recordings and measurements

The articulatory recordings were made with the 2D EMA system at NTT Speech Communication Science Research Laboratories, Atsugi, Japan. The speaker produced ten repetitions of randomized English and Japanese carrier sentences with the target words, five of which were analyzed for the current experiment. The tokens were (20 English sentences + 10 Japanese sentences) x 1 participant x 5 repetitions = 150 tokens. EMA recordings were made within a window frame of 20 sec, with a break in recordings of about 3 seconds between frames. In order to record articulatory motions, receiver coils were attached to the lower incisor (mandible), upper lip, lower lip, and T1 (tongue tip), T2 (tongue blade) and T3 (tongue dorsum). T1, T2, and T3 were attached along the longitudinal sulcus of the speaker's tongue. The positions of the coils were measured at a sampling frequency of 250 Hz. The positions of the transmitter coils determine the coordinate system [3] with the origin positioned slightly in front of and below the chin.

Articulatory measurements were made for the x-y pellet positions for the upper and lower lip (UL, LL), for the mandible (J), and the tongue (T1, T2, T3) at the time of maximum jaw opening for the utterance, using a MATLAB-based analysis program, courtesy J. Dang, ATR Laboratories, Kyoto, Japan. For each of the utterances, articulatory (as well as acoustic) measurements were made at the time of maximum jaw opening for the vowels.

3. Discussion

3.1. Results of acoustic experiment

The focus of the pilot study reported here is on the acoustic and articulatory results for the single speaker J6 (section 3.2). However a summary of the acoustic results of the 5 JS speakers reported in detail in [1] is given below.

- All of JS' pronunciations of consonant clusters were different from those of NS of English. None of the Japanese subjects were 'native-like' (i.e., English-like) in their production of consonant clusters. Every subject produced a vowel or a vowel-like transition between C1 and C2 in CC.
- Speakers J1 and J2 seemed to be inserting vowels phonologically – the vowels inserted after C1 in CC-clusters were epenthetic, not excrescent, vowels, because the vowel that appeared after /t, d/ for these speakers showed similarity to the /o/ in the CV(CV) syllable type words.
- F0 values of the intrusive vowels for J1 (and J2, to some extent) were significantly high, suggesting that pitch accent was placed on the intrusive vowel. She may be following the Japanese antepenultimate accent rule. Since this is a phonological rule, the vowel in her production of consonant clusters must be an epenthetic vowel.
- Data from the other subjects were rather inconclusive, but the quality of the vowel that appeared in J4's production of consonant clusters did not show similarities to the vowels in the CVCV English pseudowords. Our interpretation is that she produced a vowel-like transition for ease of pronunciation. It is known that it is difficult for JS to make CC clusters, since, they "have little practice in the articulatory coordination" [4,5] of consonant clusters.

Overall findings suggest that J1 and J2, seemed to be inserting epenthetic vowels while J3, J4, and J5 were

producing excrescent vowels. The speaker's English speaking level seems to be playing a role in their production of consonant clusters as well, since J1 and J2 were assessed as lower-intermediate level. Further study on more subjects from different pronunciation levels is necessary to confirm the above findings.

3.2. Results of articulatory experiment

3.2.1. Articulation of /o/ and u/ by JS6

Table 2 below shows the mean value for each measurement point (the places where pellets were attached) for 5 repetitions of the /u/ after /s/ in /sunuku/ and of the /o/ after /t/ in /toroku/. T-tests were done to assess significant differences in articulation between /o/ and /u/. The tables below show only the values that were significant ($p < 0.01$). The y-values are displayed in columns 2-5, and show vertical articulator position, with low values indicating low articulator position, and the x-values in columns 6-9, showing horizontal articulatory position, with low values indicating more fronted articulatory position.

Looking at Table 2, we can say that generally articulation of /u/ involved higher and fronted articulatory positions, while that of /o/, lower and backer articulatory positions. A raised, more fronted tongue for /u/ compared to /o/ was also reported by [6], for Japanese vowels. This is to be expected given that for /u/ compared to /o/, F1 is lower (high jaw position) and F2 is higher (front tongue position). The acoustic data from J6 show this same pattern of F1 and F2 (see [1], also, Table 4.).

Table 2. Significant mean articulatory values(mm) of /u/ & /o/ in /sunuku/ & /toroku/, respectively (JS6)

V	uly	tly	t3y	jy	t1x	t2x	t3x	jx
u	15.4	14.6	15.8	12.9	7.9	9.4	10.5	7.5
o	15.3	14.3	15.6	12.7	8.7	10.1	11.1	7.6

3.2.2. Articulation of inserted vowels by JS6

Table 3 shows the significant mean values for 5 repetitions of the inserted vowels after /s/ in /sunuk/ and after the /t/ in /tork/. The only significant value ($p < 0.01$) is for jaw-y: the vowel after /s/ has higher jaw position than that after /t/. This suggests that possibly the vowel after /s/ was similar to the /u/ vowel (with higher jaw position) and the vowel after /t/ was similar to the /o/ vowel (with lower jaw position), as indicated in Table 2 above. There were no significant differences in tongue or lip position, but trends were in the direction of a higher, fronted, less round, vowel after /s/ than after /t/, as reported in [1].

Table 3. Significant mean articulatory values(mm) of inserted vowels after /s/ in /sunuk/ and after /t/ in /tork/ (JS6)

location of inserted V	jy
s+inserted vowel	12.89
t+inserted vowel	12.79

3.2.3. Acoustic measurements for J6

Table 4 show the acoustic measurements for the vowels in the /su/ and /to/ CVCV syllables, and Table 5, shows those for the inserted vowels, after /s/ and /t/ for the CC-onset words. The items in bold indicate significant differences ($p > 0.01$).

Table 4. Acoustic measurements for vowels in /su/ & /to/ CVCV syllables (JS6)

V	dur (ms)	F0 (Hz)	F1(Hz)	F2 (Hz)
u	0.70	241	414	1719
o	0.82	223	417	1195

Table 5. Acoustic measurements for inserted vowels after /s/ and /t/ in CC onset syllables

V	dur (ms)	F0 (Hz)	F1(Hz)	F2 (Hz)
after /s/	0.35	242	378	1741
after /t/	0.63	215	432	1400

3.2.4. Summary of JS6 articulatory and acoustic results

The acoustic and articulatory results found for J6 showed that, although the vowel she was intruding after /t/ and /s/ revealed some similarities to their underlying counterparts according to Japanese phonotactics, they were not really the same. It is possible that J6 was actually trying to avoid producing a clear /o/ or clear /u/ following Japanese phonotactics. Being a fairly advanced speaker, J6 did not stress the intrusive vowel (unlike J1), suggesting that J6 was aware that she was speaking in English and knew that there should not be a vowel inserted in the consonant clusters in English. The duration of J6's intrusive vowels was generally shorter than the underlying vowels, which also seems to suggest that J6 was trying to not insert a full vowel. Note, however, that there was clearly a vowel, as both acoustic and articulatory data demonstrated.

Were those excrescent vowels then? Although it is hard to conclude from the limited sample, the following argument is proposed: there is phonological vowel insertion in J6's production of consonant clusters – i.e., J6's intrusive vowels were epenthetic vowels. The reasons for this argument are that (1) at least after /t/, where /o/ was expected, the duration was .63 ms, which is a substantial length (given that a full vowel was .82 ms) and (2) the formants of the intrusive vowels were generally clear, which also suggests a real vowel.

However, in addition, we suggest there is some influence of phonetic difficulty for speaker J6. That is, J6 seemed to be making a gesture (jaw movement) to reduce the complexity of the syllable onset. We suggest that a phonological vowel insertion occurs, but because the speaker knows that the vowel is not supposed to be present, the vowel is subject to phonetic shortening and reduction. Thus, the inserted vowel after /t/ is significantly different from the vowel after /s/ in terms of vowel quality and duration.

3.2.5. Devoiced vowels by JS6

Since Japanese speakers generally devoice high vowels which appear between voiceless consonants, no acoustic measurements can be done. Consequently, The acoustic

characteristics of these word pairs could not be compared: /strok/ vs. /sutoroku/ and /struk/ vs. /sutoruku/. (see [1]).

The results of the articulatory data showed no significant differences between the subject's devoiced /u/ and the devoiced intrusive vowel after /s/. This suggests that the subject was producing devoiced vowels and intrusive vowels in a similar way. Moreover, that the intrusive vowel after /s/ was /u/-like suggests a phonological process was involved; however, devoicing might also be a type of phonetic overlap due to glottal opening of the voiceless consonant /s/.

4. Discussion

The results of the acoustic experiment were the following: JS were found to always insert a vowel (or a vowel-like transition) after the initial consonant in consonant clusters. For three out of six Japanese subjects, the intrusive vowels that appeared were context-dependent on at least some acoustic measures and differences that were found were generally in the direction of Japanese phonotactics. These differences suggest that these subjects were indeed inserting epenthetic vowels. Furthermore, the F0 patterns in two of the subjects revealed that they seemed to be stressing the intrusive vowel, another piece of evidence for the phonological nature of the vowel insertion. The other three subjects did not show evidence of epenthesis. It has been argued that these subjects were not inserting vowels phonologically but were exhibiting decreased overlap between consonant gestures, resulting in a vowel-like transition between the consonants.

The results of the pilot articulatory experiment were generally consistent with our findings in the acoustic experiment. Although the intrusive vowels were not exactly produced as their underlying counterparts, the movements of articulators such as lips, jaws, and tongue generally tended to show patterns for /o/-like vowels after /t, d/ and /u/-like vowels after /s, b/, similar to what was seen in the acoustic data. Based on these results, the following can be proposed: a phonological vowel insertion clearly occurs, but because the speaker knows that the vowel is not supposed to be present, the vowel is subject to phonetic shortening and reduction. These results raise the difficult question as to what is phonological, and what is phonetic. More speakers are needed to explore this further.

5. Conclusions

Results found in this study strongly suggested that there is inter-subject variability as well as the influence and interaction of phonetics and phonology in L2 pronunciation. For some speakers, the process involved in their intrusive vowels could be identified as phonological, showing characteristics of epenthetic vowels such as longer duration, clearer formants, and similar tendencies to the vowels that would have appeared following Japanese loanword phonology. The finding of epenthesis supports the acoustical, articulatory, and psychological reality of syllables and syllable-structure constraints for these JS. Since Japanese, characterized as a CV language, is much more restricted than English in its syllable inventory, it has been argued that L1 transfer is the most prominent source of syllable structure errors that Japanese learners of English make ([7, 8] among others). The findings of the current research seem to be consistent with this.

In OT terms (following [7]), the learners (with their L1 rankings) may be learning the nonnative rankings of the L2 and rerank the constraints through exposure to L2 forms. If

learners' varying levels of proficiency can be captured in terms of the degree to which their rankings get closer to those of L2, it is assumed that less advanced speakers are considered to have not reranked the constraints yet. On the other hand, some of the more advanced speakers have already reranked the constraints for syllable structures and moved from L1 (Japanese) ranking to L2 (English) ranking. Note that the fact that such speakers still intrude a vowel-like transition seems to suggest that gestural coordination of L2 still needs to be learned.

Finally, it should be emphasized that two types of processes—phonology and phonetics—were involved and it is necessary to tease them apart carefully. A significant finding of this investigation was that through examination of the acoustic and articulatory data a strong influence of L1 in the abstract knowledge of “syllables” could be seen.

Questions we have for future research are: what can be done to help JS speakers change their Japanese syllable articulatory strategy to that of an English syllable articulatory strategy; how do learners' productions change as exposure to L2 moves them from nonnative to more native-like rankings. We are currently investigating some ideas about this, based on the articulatory model proposed by [9] that discusses articulatory organization of syllables (e.g., [10,11]).

6. Acknowledgements

We express our sincere gratitude to Elizabeth C. Zsiga for her support and guidance in every respect. We would also like to thank Takemi Mochida at NTT for making EMA data available and the Japanese speaker who participated in the experiment. This work was supported in part by the Japanese Ministry of Education, Science, Sport, & Culture, Grant-in-Aid for Scientific Research (C), (2010-2013).

7. References

- [1] Shibuya, Y. “Production and perception of consonant clusters in the L2 phonology of Japanese learners of English”. Ph.D. Dissertation, Georgetown University, 2005
- [2] Levin, J. “Between epenthetic and excrescent vowels”, Proc of West Coast Conference on Formal Linguistics 6:187-201, 1987.
- [3] Kaburagi, T., & Honda M. “Electromagnetic articulograph based on a nonparametric representation of the magnetic field”, *JASA* 104(4):March 2002:1414-1421, 2002.
- [4] Tajima, K., Erickson, D., & Nagao, K. “Factors affecting native Japanese speaker's production of intrusive (epenthetic) vowels in English words”, Proc. of the 2000 International Conference on Spoken Language Processing, Beijing, China, 2000a.
- [5] Tajima, K., Erickson, D., & Nagao, K. “Phonetic analysis of vowel epenthesis in native Japanese speaker's production of English words”, Proceedings of the 14th General Meeting of the Phonetic Society of Japan, Chiba, Japan, 2000b.
- [6] Hashi, M., Westbury, J. R., & Honda, K. “Vowel posture normalization”, *JASA*, 104 (4):2426-2437, 1998.
- [7] Hancin-Bhatt, B., & Bhatt, R. “Optimal L2 syllables: Interactions of transfer and developmental effects”, *Studies in Second Language Acquisition*, 19:331-378, 1997.
- [8] Saunders, N. J. “Morphophonemic variation in clusters in Japanese English”, *Language Learning* 37(2): 247-272, 1987.
- [9] Fujimura, O. “The C/D Model and Prosodic Control of Articulatory Behavior”, *Phonetica* 57, 128-138, 2000.
- [10] Fujimura, O. & Williams, J.C. “Syllable concatenators in Japanese, Spanish, and English”. In Fujimura, O., Joseph, B., and Palek, B. (eds.). *Item Order: Proceedings of LP '98*, pp. 461-498. Prague: Charles University Press, 1999.
- [11] Fujimura, O. & Williams, J.C. “Syllable-based representation and its phonetic implementation by the C/D model”. *Kyushu U.*, 9/2009