Not vowel epenthesis: Mandarin and Japanese ESL learners' production of English consonant clusters

Akitsugu Nogita & Yanan Fan
University of Victoria
akitsugu@uvic.ca, echofyn@gmail.com

The present study is based on Funatsu *et al.*'s (2008) experimental study about Japanese ESL learners' perception and production of vowel insertion. To further investigate the process of second language vowel insertion, the present study employed a reading task and a repetition task. A syllabification task was also conducted after each task. Both Japanese and Mandarin ESL participants were involved to explore the effect of language experience. The results showed that both Mandarin and Japanese ESL learners with a relatively short length of residence in Canada only occasionally inserted a short vowel in English consonant clusters when they immediately repeated sound stimuli. The between-group difference was that Mandarin speakers with correct phonological representations often attempted to produce consonant clusters without vowel epenthesis but occasionally failed in gestural coordination which resulted in a schwa-like vowel, while Japanese speakers mostly had incorrectly stored English consonant clusters with extra vowel phonemes in their interlanguage mental lexicon but they phonetically deleted or weakened such vowels when they imitated native English speakers' production.
Keywords: vowel epenthesis; vowel intrusion; vowel weakening; vowel deletion; interlanguage phonological representation

1 Introduction

It is well attested that second language (L2) learners may insert an extra vowel into a consonant cluster that is illegal in their first language (L1). For example, native speakers of English tend to insert [ə] in a Polish consonant cluster which is illegal in English, like /zb/ produced as [zəb] (Davidson & Stone, 2003). This paper examines whether Mandarin Chinese and Standard Japanese speaking English-as-a-Second-Language (ESL) learners insert a vowel in English consonant clusters with different motivations. If there are different processes in vowel insertion in L2, the learning and teaching of L2 consonant clusters will be improved.

This present study consists of both production tasks and syllabification tasks with Mandarin and Japanese ESL learners. The production tasks were designed based on Funatsu, Imaisumi, Fujimoto, Hashizume, and Kurisu's (2008) reading task and repetition task to investigate L2 learners' performance in English consonant clusters that are illegal in their L1, which we will discuss in

§2.2. The syllabification tasks are to examine L2 learners' knowledge of English syllables and whether they add extra vowels in their syllabification process. We found that in the production tasks, Japanese speakers inserted a vowel in an English consonant cluster more often than Mandarin speakers did. Both Mandarin and Japanese speakers tend to insert a vowel more frequently when they read English words aloud than when they repeated native speakers' production without written cues. In the syllabification tasks, regardless of the language background, the participants with explicit knowledge of English syllabification notably performed better than those without explicit knowledge. Our interpretation is that there are three different processes for vowel insertion: 1) those who had known syllabification for a long time (all the Chinese participants) correctly understood where to or not to pronounce a vowel but occasionally failed to pronounce consonant clusters because of difficulty in articulation, 2) those who had recently learned syllabification (a few of the Japanese participants) were still interfered by L1 phonotactics and epenthesized vowels, 3) those who did not know syllabification (the rest of the Japanese participants) did not know where not to pronounce a vowel.

2 Background and Hypothesis

2.1 Studies of Dupoux et al. and Funatsu et al.

Dupoux, Hirose, Kakehi, Mehler, and Pallier (1999) found that native Japanese speakers hear "illusory" [u][1] in a consonant cluster when they hear the VCCV structure where the CC is an illegal sequence in Japanese, such as [ebzo] perceived as [ebuzo]. Dehaene-Lambertz, Dupoux, and Gout (2000) approached the same topic by using a brain-based measure. When native French speakers heard the stimulus change from "igumo" to "igmo," they showed strong responses, while native Japanese speakers showed significantly weaker responses when hearing the same change. Dupoux et al. and Dehaene-Lambertz et al. concluded that Japanese speakers did not pay much attention to the difference between CuC and CC if the CC is a phonotactically illegal consonant cluster in Japanese. However, Funatsu, Imaisumi, Fujimoto, Hashizume, and Kurisu (2008) criticized that Dupoux et al. synthetically deleted [u] of the original sound [ebuzo] to make [ebzo] but listeners might have heard a coarticulation effect from the deleted vowel to the neighboring consonants. Funatsu et al. conducted their study consisting of a reading task and a repetition task. In the reading task, novice Japanese English-as-a-foreign-language (EFL) learners read aloud written stimuli that were both real English words and nonsense words. In the repetition task, the participants immediately repeated sound stimuli that were the same words as the written stimuli pronounced by a native English speaker. Funatsu et al.'s study mostly focused on word-initial /Cɹ/ sequences. In the reading task, the speakers

[1] Dupoux et al. (1999) used the symbol [u] for the Japanese high back vowel instead of [ɯ] to be consistent with French. Since roundness is not a distinctive feature in Japanese, either [u] or [ɯ] can be used. We use [ɯ] in this paper according to the custom.

inserted a full vowel (most of the time [o]) after /t/ or /d/, and [ɯ] elsewhere; note that these patterns are the same as vowel epenthesis in loanword adaptation (see §2.1 for more detail). However, in the repetition task, Japanese speakers correctly pronounced the target consonant clusters or inserted only a very short vowel. Thus, Funatsu et al. argued that Japanese speakers can correctly perceive consonant clusters without "illusory" vowels, and occasionally inserted short vowels were due to their gestural mistiming of coordinating two consonants rather than lexical epenthetic vowels.

However, we take a different view. In the repetition task in Funatsu et al.'s study, the short inserted vowels were still [o] before [t] and [d] and [ɯ] elsewhere, which was the same as the reading task and the loanword adaptation patterns. If these were really gestural mistiming, these would have been more like transitional vowel occurring in a gap between two consonants. If these were transitional vowels, there is not phonetic explanation why those were [o] before [t] and [d] and [ɯ] elsewhere. We interpret that in the Japanese EFL learners' production, underlyingly /o/ and /ɯ/ were still present but phonetically shortened or weakened when the participants were influenced by the sound stimuli. In other words, we suspect that as Funatsu et al. mentioned, Japanese speakers can correctly hear English consonant clusters at the phonetic level, and as Dupoux et al. (1999) mentioned, in Japanese speakers' phonological representation, there is still [o] before [t] and [d] and [ɯ] elsewhere in illegal consonant clusters, regardless of whether or not [o] or [ɯ] is phonetically present. Besides, we are also interested in other L1 speakers' behavior in the same tasks. For example, Mandarin speakers used both deletion and insertion in the production of English consonant clusters (see §2.3). As well, L2 learners' modification strategies vary in different tasks.

Our research questions are whether vowel insertion by Japanese learners of English is really gestural mistiming as Funatsu et al. discussed, and whether Mandarin speakers behave differently from Japanese speakers.

2.2 Mandarin Chinese and Standard Japanese syllables

The maximal syllable structure in both Mandarin Chinese and Standard Japanese is generally CGVX (C=consonant, G=glide, V=vowel, X=nasal or vowel) (Lin, 2001). The only complex (or branching) onset is a consonant-glide sequence: e.g. /mjæn/ "Noodle" in Mandarin and /kjó:to/ "Kyôto" in Japanese. These consonant clusters have a great sonority distance; a glide is the most sonorous consonant, followed by liquids, nasals, and obstruents, and the greater the sonority distance, the less marked the cluster is (Eckman & Iverson, 1993). This suggests that Mandarin and Japanese allow complex onsets but only unmarked ones. Note that there are some other interpretations for their complex onsets; in Japanese, a minimal pair like /makɯ́/ "drop curtain" vs. /mjakɯ́/ "pulse" is sometimes interpreted as the plain-palatal distinction and /mj/ can be transcribed as /m^{j}/ like a single consonant as in Akamatsu (2000). In Mandarin, the glide can be included in a nucleus rather than an onset: e.g. /m-jæn/ instead of /mj-æn/. As

for codas, for both languages, a word-final coda can only contain nasals. A nucleus can be either a single vowel or a diphthong in Mandarin. Japanese has an additional condition that a nucleus can be a phonologically distinctive long vowel (Lin, Y., 2007).

In loanword adaption in Mandarin and Japanese, when phonotactically illegal consonant clusters are in the borrowed words, vowel epenthesis can be seen. Miao (2005) summarized the process of loanword adaption of complex onsets and codas in Mandarin. In the corpus, vowel epenthesis is the most common strategy. In the onset position, the epenthesis only occurs between the two consonants. In the coda position, the epenthesis may happen between the two consonants like CVCVC, or after the second consonant like CVCCV, or after the first consonant as well as after the second one like CVCVCV. Loanword adaptation in Japanese also prefers epenthesis, as in the complex onset "play" adapted to /pɯ.ɾé:/, the complex coda "old" to /ó:.ɾɯ.do/, and the syllable boundary "badminton" to /ba.do.míN.toN/. In Japanese the default epenthetic vowel in loanword adaptation is /ɯ/, but after /t/ and /d/, /o/ is the most common (Funatsu et al., 2008). From these processes, we infer that both Mandarin and Japanese speakers would prefer a CV structure. Thus, epenthesis might happen in the production of consonant clusters.

However, phonotactics in Mandarin and Japanese is more complicated. In more detail, Japanese allows three-mora syllables, which do not fit the aforementioned CGVX template (Vance, 2008). For example, the second syllable [homp] in [ni.homp.pói] (/ni.hoNQ.pói [2] /) "Japanesy" has a complex (or branching) coda (Vance, 2008, p.131), where as Chinese does not allow complex codas. On the other hand, Mandarin allows more types of consonant clusters in syllable boundaries. When the coda of the first syllable is a nasal (/n/ or /ŋ/), the onset of the second syllable can be a plosive (e.g. *benbu* 34[3], /pʰɤn.pʰu/, "headquarters"), a fricative (e.g. *fenshou* 13, /fɤn.ʂo/, "break up"), an affricate (e.g. *fenqi* 12, /fɤn.tɕʰi/, "difference"), a liquid (e.g. *benlai* 32, /pʰɤn.lai/, "original"), and a nasal (e.g. *benneng* 32, /pʰɤn.nɤŋ/, "instinct"). In contrast, in Japanese consonant clusters across syllables are allowed as long as the two consonants are a nasal followed by a homorganic consonant, or two consonants that agree in place (Inozuka & Inozuka, 2009), as in [ba.do.mín.toN] (/ba.do.míN.toN/) "badminton," [dén.tɕi] (/déN.tɕi/) "battery," and [tem.móŋ.ga.kɯ] (/[teN.móN.ga.kɯ/) "astronomy." Therefore, Mandarin speakers may have more experience in consonant clusters across syllable boundaries than Japanese speakers do.

Then again, at a more phonetic level, Japanese speakers may have more experience in consonant clusters with obstruents due to vowel devoicing. As a common generalization, the Japanese short high vowels /i/ and /ɯ/ are regularly devoiced between voiceless consonants and in a pre-pausal position. So-called devoiced vowels are often not true vowels but entirely disappear (Vance, 2008).

[2] /N/ is a placeless nasal, and /Q/ is the first half of a geminate consonant (Vance, 2008). Both are moraic.

[3] These numbers indicates phonemic tones in each syllable.

One between voiceless stops is realized as a fricative, as in /kitai/ → [kɨ̥tai] → [kçtai] "expectation," and it may entirely disappear when involving a fricative, as in /sɯtáiɾɯ/ → [sɯ̥táiɾɯ] → [stáiɾɯ] "style" (Vance, 2008, pp. 207-208). Ôno (2007) mentioned that contrary to Ôsaka or Kyôto dialects, Eastern dialects, including Standard Japanese, prefer to use closed syllables both phonologically with /N/ and /Q/ (see footnote 2) and phonetically with devoiced vowels as in the sentence-final polite copula /désɯ/ → [désɯ̥] → [dés]. Ôno speculates that the preference of closed-syllables in Eastern dialects may be due to a substratum of Japanese, such as the Ainu language, although there is much debate regarding the origin of Japanese. Interestingly, the Japanese singers named /ásɯka/ and /gákɯto/ with devoiced [ɯ̥] spell their own names as "Aska" and "GACKT" respectively instead of the regular spellings "Asuka" and "Gakuto." This may imply that these singers not only adopt fancy foreign-looking spellings but interpret [sɯ̥] and [kɯ̥] as the coda of the first syllable. Although Japanese is generally treated as a language with single onsets and codas, we argue that it allows branching onsets and codas on the basis of the preceding evidence. It also allows (pseudo) obstruent clusters at the phonetic level. It would not be very surprising if Japanese ESL learners can pronounce English consonant clusters without much difficulty.

Another difference between Mandarin and Japanese is that based on the traditional classification, Mandarin is a syllable-timed language in which a syllable is the smallest unit (Lin, 2008), whereas Japanese is a mora-timed language in which a mora is the smallest unit (Vance, 2008). Ueyama (2003) found that Japanese ESL learners who had been in the United States for more than five years still incorrectly syllabified English words more than 50% of the time and tended to divide a word into morae. Even advanced Japanese ESL learners may not be able to naturally acquire English syllabification without formal instruction. Ueyama speculated that a possible reason is that the English writing system does not show syllabification. In contrast, in Chinese orthography, each morpheme-syllable is represented by a character (Sun, 2006). Thus, Mandarin speakers have enough experience to identify syllable boundaries. More specifically, in Modern Chinese, morphological compounding leads to the dominance of disyllabic words. Unlike Old Chinese, two, sometimes three free words are allowed to form a compound word (Sun, 2006, p50). The preference of disyllabic words is also reflected by the prefix and suffix in Modern Chinese. When the free morphemes like "*lao*," "*xiao*," and "*zi*" are used as a prefix or suffix, the new word still has the same meaning with the original monosyllabic word. For example, "*hu*" and "*laohu*" have the same meaning "tiger" and "*bei*" and "*beizi*" share the same meaning "cup". All this subconscious knowledge may help them acquire the syllabic system, especially disyllabic words in English.

2.3 Vowel insertion in second language acquisition

Previous research on the acquisition of consonant clusters in L2 has been well documented. Most of these studies investigated this issue in terms of error types

and accuracy rate (Yoo, 2004; Hansen, 2001; Carlisle, 1998; Eckman and Iverson, 1994). In all the studies above, they found that epenthesis is the major modification strategy used by L2 learners. As well, previous studies accounted for learners' acquisition of the consonant clusters in terms of L1 transfer (Kim & Jung, 1998; Bayley, 1996; Broselow, 1987), markedness (Yoo, 2004; Hansen, 2001), sonority (Carlisle, 2006; Davidson, Jusczyk and Smolensky, 2004), and frequency (Davidson, 2006; Pitt, 1998). In a paper about Mandarin speakers' production of consonant clusters, Hansen (2001) used the interaction of all these linguistic constraints above to explain her results. In her study, she found two patterns of epenthesis by Mandarin speakers, CVCVC & CVCCV, which she explained as the interaction between the linguistic environment and the universal preferences of syllable structure. Interestingly, phonological vowel epenthesis by L2 learners may not be always phonetically present as a voiced vowel. Goad and Kang (2003) reported that Japanese ESL learners with higher proficiency commonly devoice and aspirate the word-final voiced stop in English (e.g. [fɑnt^{h}] "fond"). Such aspiration is considered as a voiceless vowel, often [ɯ̥] (/ɯ/) as mentioned in §2.1. Urbanczyk (1996) reported a similar phenomenon in Lushootseed; there is syllabic aspiration, which can be interpreted as a voiceless schwa. In fact, Goad and Kang also reported that Japanese ESL learners delete the final vowel in an English vowel final word and aspirate the preceding stop (e.g. [fɑnt^{h}] "(Jane) Fonda"). In other words, [fɑnt^{h}] "fond" and [fɑnt^{h}] "Fonda" are actually [fɑndɯ̥] and [fɑndə̥] respectively. Phonological epenthesis may phonetically occur as aspiration.

Recently, a few of the studies about L2 consonant clusters began to focus on the gestural model. Davidson et al. (2003) used ultrasound to detect insertion caused by gestural mistiming in L2. Some Russian speakers in this study did not necessarily use phonological epenthesis to repair illegal consonant clusters, but rather failed to employ the appropriate gestural coordination for English initial consonant clusters, which is vowel intrusion. Davidson (2006) conducted a study comparing the acoustic values of a lexical schwa and a so called "transitional schwa." She found that English speakers' pronunciation of "a transitional schwa" was significantly different from a lexical schwa. The values of the first formant (F1), second formant (F2) and duration were much lower in "a transitional schwa" regardless of linguistic environment. Zsiga (2003) found that the articulatory timing in L1 could be transferred to L2. She found the presence of unexpected release burst in Russian speakers' English consonant clusters. Zsiga explained that the audible release burst is typical in Russian which is not common in English and the Russian speakers transferred the L1 articulatory timing pattern to English. Gafos (2002) claimed that the pattern of consonantal coordination is language specific. Generally, there are two ways to produce a transition between two consonants. One is "close transition" which means the release of the first gesture and the target of the second gesture occur at the same time. Thus, there is no release of the first consonant. The other is called "open transition" in which the onset of the second gesture aligns to the c-center of the first gesture as shown in Figure 1.

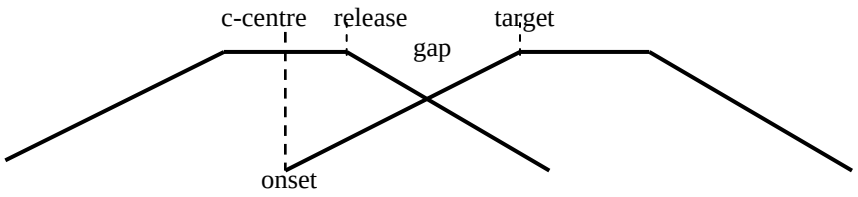

Figure 1. "Open transition" explained by Gafos (2002).

This coordination results in a gap or an open vocal tract between the two consonantal gestures which provides the possibility that the inserted vowel between two consonant is due to gesture mistiming. When L2 learners encounter English consonant clusters, which, according to Gafos, have the "close transition," they may fail to manage the articulatory timing and pull apart two gestures to make an open vocal tract between two consonants, resulting in a short schwa-like vowel or an intrusive vowel.

While epenthesis is phonological and intrusion is phonetic, the third approach is induced errors. We cannot overstate Japanese ESL learners' loanword interference as Goble (2001) found that there is an inordinate amount of loanword interference in Japanese English-as-a-Foreign (EFL) learners. Martin (2004) also states that there is a strong tendency to use *Katakana*-English in English classes in Japan. There is a wide range in the definition of *Katakana*-English, but in this paper, we define *Katakana*-English as almost standardized extension of Japanese loanword adaptation to all the English words: for example "This is a fridge" is almost unanimously pronounced by Japanese ESL/EFL learners as ディス イズ ア フリッジ /dísɯ izɯ a ɸɯɾíQdʒi/ where none of these words is the Japanese vocabulary. When Japanese ESL learners pronounce extra vowels, it may be due to *Katakana*-English stored in their interlanguage mental lexicon regardless of their ability to articulate.

As for the difference between Mandarin and Japanese speakers, Eckman (1981) found more insertion for advanced Mandarin-English speakers, while Weinberger (1987) found equal proportion of insertion and deletion for intermediate level Mandarin-English speakers. Weinberger (1994) explained the difference in terms of recoverability. The advanced learners have more knowledge about the target lexicon than intermediate learners. Thus they will more often realize the ambiguity. To avoid the ambiguity, they will employ more insertion. On the other hand, in Japanese, Smith (2006) mentioned that English words are usually imported to Japanese through written materials rather than spoken materials. There are also a few auditorily imported English loanwords, but there is a notable difference. In auditorily imported words, deletion of consonants is preferred, as in /ɾamɯne_/ "lemonade" where /d/ is deleted and /hé_boN/ "Hepburn" where /p/ is deleted (p. 68), while in orthographically imported words, vowel epenthesis is preferred, as in /aisɯkɯɾíi:mɯ/ "ice-cream." Since Japanese EFL learners commonly use aforementioned *Katakana*-English pronunciation, it is expected that the Japanese speakers with little exposure to

native English speakers' production may prefer vowel insertion especially with written material. As mentioned above, advanced ESL learners are expected to insert devoiced vowels.

2.4 Hypothesis of three different types of vowel insertion

We hypothesize that there are three insertion types: 1) an intrusive vowel, which is not a lexical vowel (or a vowel in the phoneme inventory) but occurs between a consonant cluster when the first consonant is released before the second consonant starts, 2) an epenthetic vowel, which is a lexical vowel that occurs to satisfy lexical syllabification (Hall, 2006; Davidson et al., 2004), and 3) a vowel in misinterpreted L2 phonological representation, which is a lexical vowel that exists in L2 learners' incorrectly memorized mental lexicon. For example, when a Japanese ESL learner pronounces /ɹʌgbi:/, in 1), s/he attempts to produce /gb/ as [gb] without being interfered by L1 phonological constraints but fails to coordinate the two consonants in articulation, which results in a short schwa-like vowel. Or s/he has acquired the L2 phonological structure but has not acquired its phonetic characteristic. In 2), s/he is aware of the L2 /bg/ sequence which is phonotactically illegal in L1, but he/she still has trouble performing it in practice and the L1 syllable structure comes out. In 3), s/he has incorrectly stored "rugby" as /ɹʌgɯbi:/ in his/her interlanguage mental lexicon and pronounced it without knowing that it is incorrect. This study examines whether Mandarin and Japanese ESL learners insert a vowel in English consonant clusters, and if they do, which of the three types is used.

3 Research design

3.1 Participants

We recruited six Mandarin and eight Japanese ESL learners in Canada reportedly at lower-intermediate to intermediate proficiency levels with relatively short length of residence (3 to 13 months) in an English speaking country. Three Mandarin speakers were male and the rest were female in their 20's. Japanese speakers were all from Kantô or Chûbu regions. None of the participants knew other languages besides English and their L1. No hearing problem was reported. Mandarin speakers were labelled as M1, M2, and so forth and Japanese speakers were labelled as J1, J2, and so forth.

3.2 Stimuli

Funatsu et al. mostly focused on /p, b, t, d, k, g/-/ɹ/ sequences. There is a possibility that their Japanese participants pronounced vowels, which might have been phonetically realized as devoiced vowels after voiceless consonants. Dupoux et al. (1999) also used stimuli where devoiced vowels can occur. Our focus in the present study is vowel insertion. Thus, the appearance of voiceless

vowels might affect our results. For this reason, we only chose voiced stops. As well, the target clusters in Funatsu et al.'s study were mostly in the word initial position, which is salient in speakers' perception (Toro, Sebastia´n-Galle´s, & Mattys, 2009). Also, a stop-/ɹ/ cluster is great in sonority, which is relatively unmarked; stops are the least sonorous and /ɹ/ is the most sonorous consonant in English except for glides. In other words, Funatsu et al. used relatively unmarked clusters, which may have enabled novice Japanese EFL learners to perform well, but they may have more difficulty in more marked clusters. Therefore, we use more marked clusters (smaller in sonority distance than stop-/ɹ/), /b, d, g/-/l, m, n, O/ (O=obstruent) (3 × 4 = 12 combinations) in the more marked word-medial position.

The stimuli were both real English words and non-sense words. There were 24 critical items (12 reals and 12 nonsenses) and 16 fillers (8 reals and 8 nonsenses). All the critical items were disyllabic (except for "badminton") with the primary stress on the first syllable and most of the fillers were not disyllabic. The real words were considered familiar to the participants. The nonsense words were made based on English phonotactics. In the nonsense words we avoided /u/ and /oʊ/, which are perceptually similar to Japanese common epenthetic vowels /ɯ/ and /o/ respectively (Nishi, Strange, Akahane-Yamada, Kubo, & Trent-Brown, 2008) in order to avoid potential influence by /oʊ/ or /u/ in the adjacent syllables. These 20 real words were randomized in order, and so were the nonsense words, and the written stimuli and audio stimuli were in different order. The written stimuli were printed on a sheet. The audio stimuli were produced by a phonetically trained female native speaker of Canadian English in an anechoicchamber in the UVic Linguistics Speech Research Lab. The speaker produced each word five times and only the third one, which we assumed to be the most stable and natural, was used as a stimulus. Following are the stimuli. Parenthesized words are fillers.

1. Real words: subject, webmail, webnet, tablet, foodbank, badminton, Sydney, badly, rugby, eggman, magnet, ugly, (avocado, banana, coconut, fruit, grape, ice-cream, strawberry, vegetable)

2. Nonsense words: ebdet, gabmee, gabno, cabla, idgay, cadma, pednay, edlee, agday, egmad, hegneb, agla, (ba, cantukpeg, gamboozee, jeejee, ma, muzz, smecks, sna)

3.3 Procedure

There were four tasks: two production tasks and two syllabification tasks. In the first task, the participants were recorded reading aloud the written real words and nonsense words. The recording was done with the software Audacity set at 44100 Hz and 32-bit float in the UVic Phonetics Lab. In the second task, the participants were asked whether they know what 'syllable' (or "*yinjie*" and "*onsetsu*" in Mandarin and Japanese respectively) was, and they were asked to

separate each written word into syllables. We demonstrated how to divide the Mandarin word "*yinjie*" (syllable) into "yin-jie" by making a pause between syllables to Mandarin participants. To the Japanese speakers, because most Japanese participants did not know syllable, we demonstrated the multi-CV-syllable Japanese word "*wasabi*" divided into "wa-sa-bi," and explained that the monosyllabic Japanese word "*ka*" (mosquito) could not be divided into a smaller unit. Because the number of syllables and moras in both "*wasabi*" and "*ka*" match, the participants would not figure out the difference between syllable and mora. The participants who did not know syllable were instructed to syllabify according to their impression. In the third task, they listened to each stimulus without looking at written cues and immediately mimicked the stimulus. To avoid practice effects, the participants heard each stimulus only once, except for a few cases when the participants could not repeat at all. In the fourth task, the participants listened to each word and divided it into syllables without looking at written cues.

For data analysis, error patterns and duration of inserted vowels were analyzed on the phonetic software PRAAT. Duration was measured from the left-most zero crossing of the first pitch pulse to the right-most zero crossing of the last pitch pulse on the waveform. Both authors individually analyzed the data and when there were disagreements, we discussed to make the final decisions. When we could not make a decision, we consulted with a third phonetician about how the token should be measured.

4 Results

4.1 Overall results

We categorized five levels of coordination of consonant clusters. 1) "Feature change"; two consonants are so close that phonological assimilation occurs: e.g. the nonsense word "cadma" pronounced as "canma" (nasal assimilation) by J5 and J6, "badly" pronounced as "banly" (sonorant assimilation) by J8 all in the repetition task. 2) "Unreleased"; a cluster shows close transition without release burst. 3) "Released"; the first consonant has release burst or aspiration. 4) "Coalescence"; if participants inserted a vowel, the first voiced stop in the cluster ended up intervocalic and spirantized to a voiced fricative, and then the inserted vowel was absorbed by the fricated stop, which turned into a long fricative or even approximant (e.g. "rugby" → [ɹʌgɯbi] → [ɹʌɣːbi]), or the inserted vowel was merged also with the following /l/ (e.g. "ugly" → [ʌgɯli] → [ʌɣɬːi]). 5) "Vowel insertion"; a vowel is inserted and the spectrogram shows formants and periodic pulses. As going from 1) to 5), the two consonants in the cluster becomes phonetically further apart. Besides these five vowels, we made the category "others" that includes misinterpretation of spellings, careless mistakes, and so on. For example, "g" in "rugby" pronounced as [dʒ] in the reading task was considered misinterpretation of the spelling-sound correspondence of <g>. Incidentally, Japanese speakers often pronounced /l/ as a flap [ɾ] or rather [ɹ]-like

sound because /ɾ/ is the only liquid phoneme in Japanese. It does not count as "feature change" or "others." Table 1 shows the overall results. Note that native English speakers typically unreleased the first consonant in a consonant cluster, but "unreleased" in Table 1 does not necessarily mean that the participants' productions were target-like, and other category does not necessarily mean strong foreign accents.

Table 1. The percentage and the number (bracketed) of each error type in each task. Nb. Each task contains 144 tokens (12 clusters × 2 word types (real, nonsense) × 6 participants) in the Mandarin group and 192 tokens (8 participants) in the Japanese group.

	Reading task		Repetition task		Total	
	Mand.	Japan.	Mand.	Japan.	Mand.	Japan.
1. Feature change	5.6% (8)	0% (0)	2.8% (4)	3.1% (6)	4.2% (12)	1.6% (6)
2. Unreleased	**35.4% (51)**	**13.5% (26)**	**41.0% (59)**	**35.9% (69)**	**38.2% (110)**	**24.7% (95)**
3. Released	36.1% (52)	12.0% (23)	48.6% (70)	30.2% (58)	42.4% (122)	21.1% (81)
4. Coalescence	0.7% (1)	4.2% (8)	0% (0)	2.6% (5)	0.3% (1)	3.4% (13)
5. Vowel insertion	20.8% (30)	66.1% (127)	6.9% (10)	28.1% (54)	13.9% (40)	47.1% (181)
Others	1.4% (2)	4.2% (8)	0.7% (1)	0% (0)	1.0% (3)	2.1% (8)

First, vowel insertion occurred much more frequently by Japanese speakers. Coalescence also occurred more often by Japanese speakers. As mentioned in §2.3, beginner Mandarin ESL learners prefer deletion while intermediate learners are more inclined to vowel insertion for the sake of recoverability. In the present study, all the participants were at lower intermediate level. They may not be aware of recoverability as much as advanced learners. Thus, less insertion might be used to repair the nonnative clusters. As well, it is understandable that the Japanese participants preferred vowel insertion especially with written material in the reading task due to *Katakana*-English pronunciation (extension of loanword adaptation to all the English words).

To compare reading and repetition tasks, in both groups, "Vowel insertion" (as well as "Coalescence," which is a type of vowel insertion in this experiment as mentioned above) dramatically decreased from the reading task to the repetition task, while unreleased stops, which are the closest to target-like productions, increased. These results agree with Funatsu et al.'s study, suggesting that both Mandarin and Japanese participants perceived consonant clusters and succeeded to imitate target-like clusters in many cases. Interestingly, a paired T-test shows that Mandarin speakers' unreleased stops were not significantly more frequent in the repetition task ($p>0.1$), while Japanese speakers showed

significant increase ($p<0.001$). This suggests that the sound stimuli may have been more helpful for Japanese speakers than for Mandarin speakers. In addition, released stops were more frequently pronounced in the repetition task in both groups. Again, this increase is significant in the Japanese group ($p<0.005$), while in the Mandarin group it is marginally significant ($p=0.056$). This indicates that both groups changed the vowel insertion strategy to the release or unreleased strategy in the repetition task, but this change is notable in the Japanese group. We will discuss this further in §5. Feature change decreased in the repetition task in the Mandarin group, while feature change occurred only in the repetition task in the Japanese group. In the Mandarin group, deletion-like feature change sometimes occurred: e.g. "idgay" realized as [ɪggeɪ] where place assimilation took place in /d/ resulting in the geminate [gg]. This may be the trace of the aforementioned non-advanced Mandarin ESL learners' preference of the deletion strategy. In the repetition task, the sound stimuli may have helped Mandarin speakers avoid the deletion-like strategy. In contrast, Japanese speakers prefer insertion in the reading task, so that feature assimilation could occur between the two consonants. In the repetition task, the sound stimuli helped Japanese speakers pronounce consonant clusters, which triggered feature assimilation between the consonants.

4.2 Vowel insertion patterns

We also did find a duration difference between the inserted vowels in the two tasks like Funatsu et al.'s study. Table 2 reports the overall frequency of insertion and the mean duration of the inserted vowels in two tasks according to the consonantal contexts. Table 3 to 6 show individual vowel insertion, release (including aspiration), and unreleased productions. We omitted feature change, coalescence, and others due to their small numbers of occurrences.

Table 2. The frequency of insertion and mean duration of inserted vowels. Nb. For each cluster, the Mandarin group has 12 occasions (6 participants × 2 word types) and the Japanese group has 16 occasions (8 participants).

clusters	Reading task Frequency of insertion (Mean duration)		Repetition task Frequency of insertion (Mean duration)	
	Mand.	Japan.	Mand.	Japan.
/b/-Obs	1 (34ms)	12 (39ms)	0	6 (24ms)
/bm/	1 (54ms)	2 (41ms)	0	0
/bn/	3 (26ms)	12 (40ms)	0	4 (12ms)
/bl/	5 (40ms)	11 (44ms)	2 (46ms)	7 (27ms)
/d/-Obs	1 (75ms)	11 (47ms)	0	1 (27ms)
/dm/	1 (18ms)	13 (42ms)	1 (36ms)	6 (29ms)
/dn/	4 (42ms)	12 (62ms)	2 (61ms)	4 (48ms)
/dl/	5 (39ms)	12 (53ms)	3 (37ms)	8 (47ms)
/g/-Obs	0	9 (31ms)	0	0
/gm/	0	6 (30ms)	0	2 (27ms)
/gn/	4 (97ms)	15 (39ms)	0	10 (23ms)
/gl/	5 (59ms)	12 (59ms)	2 (51ms)	6 (40ms)

Table 3. Consonant clusters in the real words by each Mandarin speaker. Nb: 'V Insertion' Indicates the number of vowel insertion; 'Mean dur.' indicates mean duration of the inserted vowels; 'Released' indicates the number of release or aspiration; 'Unreleased' indicates the number of unreleased stops.

Participants	M1		M2		M3		M4		M5		M6	
	Read	Rep	Read	Rep	Read	Rep	Read	Rep	Read	Rep	Read	Rep
V Insertion	5	2	2	2	1	0	4	1	3	1	0	0
Mean dur.	(31)	(45)	(25)	(21)	(24)		(71)	(80)	(40)	(56)		
Released	5	5	8	8	4	3	4	5	3	10	1	0
Unreleased	**2**	**4**	**2**	**2**	**7**	**9**	**3**	**6**	**6**	**1**	**11**	**12**

Table 4. Consonant clusters in the nonsense words by each Mandarin speaker.

Participants	M1		M2		M3		M4		M5		M6	
	Read	Rep	Read	Rep	Read	Rep	Read	Rep	Read	Rep	Read	Rep
V Insertion	3	1	3	1	0	0	4	1	5	1	0	0
Mean dur.	(51)	(43)	(27)	(33)			(88)	(48)	(69)	(65)		
Released	6	8	9	9	3	5	3	7	6	10	0	0
Unreleased	**0**	**2**	**0**	**2**	**7**	**6**	**2**	**3**	**0**	**0**	**11**	**12**

Table 5. Consonant clusters in the real words by each Japanese speaker.

Participants	J1		J2		J3		J4		J5		J6		J7		J8	
	Rea	Rep	Rea	Rep	Rea	Rep	Rea	Rep	Rea	Rep	Rea	Rep	Rea	Rep	Rea	Rep
V Insertion	7	6	7	4	5	0	10	7	9	4	8	6	12	6	2	2
Mean dur.	(42)	(28)	(22)	(20)	(47)		(28)	(32)	(61)	(57)	(36)	(29)	(60)	(40)	(17)	(14)
Released	2	1	4	5	3	4	0	4	2	3	3	2	0	5	3	1
Unreleased	**0**	**2**	**1**	**3**	**4**	**8**	**2**	**1**	**1**	**4**	**0**	**3**	**0**	**1**	**7**	**8**

Table 6. Consonant clusters in the nonsense words by each Japanese speaker.

Participants	J1		J2		J3		J4		J5		J6		J7		J8	
	Rea	Rep	Rea	Rep	Rea	Rep	Rea	Rep	Rea	Rep	Rea	Rep	Rea	Rep	Rea	Rep
V Insertion	6	1	5	1	8	0	11	2	8	3	8	4	12	7	9	1
Mean dur.	(41)	(20)	(36)	(27)	(71)		(33)	(33)	(73)	(31)	(49)	(35)	(49)	(29)	(26)	(35)
Released	3	4	0	4	0	3	0	7	1	4	1	3	0	3	1	5
Unreleased	**2**	**5**	**3**	**7**	**2**	**9**	**0**	**3**	**2**	**4**	**0**	**4**	**0**	**2**	**2**	**5**

In the reading task, insertion happened in almost every type of cluster (except /b/-Obs and /gm/ by Mandarin speakers). Within tasks, insertion by Mandarin speakers occurred less often before obstruents and /m/ in both tasks. After /d/ and before /l/, there seems a bigger chance of insertion. By Japanese speakers, a similar tendency can be seen although it is not as clear as that of Mandarin speakers. Even in the reading task, there were only two occasions of vowel insertion in /bm/ by Japanese speakers. The /bm/ cluster may be the easiest for Japanese speakers. As for vowel duration, vowels by Japanese speakers in the reading task are clearly longer than those in the repetition task. Each participant showed almost the same tendency as displayed in Table 5 and 6. In contrast, Mandarin speakers' vowels in the reading task were not necessarily longer, probably because there were too few occasions to observe the tendencies. In more detail, the inserted vowels by Mandarin speakers in the repetition task were all schwas. This indicates that in the repetition task insertion was likely to be intrusion, while in the reading task, some of the cases may have happened at a phonological level. On the other hand, in the Japanese group, a short schwa-like vowel occurred only twice in J8's reading task, but most of the others were the common Japanese epenthetic vowels [o] after [d] and [ɯ] elsewhere in both reading and repetition tasks, as in Funatsu et al.'s study. J8's schwas may be intrusion but we considered the others to be epenthetic or vowels originating from misinterpreted L2 representation. The syllabification tasks will tell more details.

4.3 Syllabification Tasks

As insertion happened, how speakers syllabify these words with target consonant clusters may reflect the process of insertion. Thus, we asked the speakers to syllabify each word (both target words and fillers) and produce each syllable separately in both reading task with written stimuli and repetition task with sound stimuli.

First, we asked the participants' knowledge of English syllabification. Reportedly, all the Mandarin speakers knew it, while only J1 and J8 in the Japanese group had received formal instruction in Canada, but not in Japan, about a month and half a year prior to the experiment respectively. J2 reported that she might have been taught it in an English phonetics class at her university in Japan but she forgot it. The numbers of errors were summarized in Table 7.

Table 7. The numbers of syllabification errors in each task by each participant (reading task–repetition task). NB. Those with * had never been taught English syllabification.

	M1	M2	M3	M4	M5	M6		
Real words	1–1	5–5	0–3	6–4	3–5	1–3		
Nonsense words	0–0	2–0	2–2	1–0	1–0	1–0		
Overall	1–1	7–5	2–5	7–4	4–5	2–3		
	J1	(*)J2	*J3	*J4	*J5	*J6	*J7	J8
Real words	4–4	17–8	4–7	7–8	8–7	16–16	6–9	0–0
Nonsense words	1–1	17–11	4–3	2–4	6–2	17–17	3–5	0–0
Overall	5–5	34–19	8–10	9–12	14–9	33–33	9–14	0–0

Overall, only one Japanese participant (J8) who had been taught syllable half a year ago did the entire syllabification correctly. Although all the Chinese participants reported that they all knew the concept of syllable, none of them got full marks. The overall correctness of the syllabification task by the Japanese speakers is 65% in the reading task and 68.1% in the repetition task including the extreme case of J8, while the results of the Mandarin speakers were 90.4% in both tasks. This difference may partially come from the fact that Mandarin is a syllable-timed language while Japanese is a mora-timed language, but come more from the fact that all the Mandarin participants had taken formal instructions of English syllabification in China while none of the Japanese participants had in Japan. In fact, the participants with explicit knowledge of syllabification (without * marks) notably made fewer errors than those without it (with *). J8 performed better than J1 probably because J8 had known English syllabification for several months, while J1 had known it only for a month. This suggests that formal instruction can help ESL learners become aware of the concept of syllabification in English. Although it is unwise to generalize based on such a small sample size, recall that Japanese ESL learners with more than five years of residence in the U.S. cannot acquire English syllabification without formal instructions, as mentioned in §2.2. Interestingly, J2 who reportedly learned syllable in an English phonetics course at university in Japan did not correctly syllabify the stimuli. Makino (2008) stated that an English phonetics course at university in Japan is a mere drop in the bucket for Japanese EFL learners who did not take formal instruction of pronunciation at the beginning. The difference among EFL in Japan and China, and ESL can be a future topic. There was no significant difference between the reading and repetition tasks calculated by paired t-test ($p=1$ Mandarin, $p>0.1$ Japanese). However, both speakers made significantly more errors with real words ($p<0.005$ Mandarin, $p<0.05$ Japanese). One of the possible reasons is they were biased by loanword pronunciation.

In the Mandarin group, among the incorrect tokens, the errors can be categorized in Table 8, which shows that all the errors but one occurred for monosyllabic and trisyllabic words. Mandarin speakers seemed to be good at disyllabic words as mentioned in §2.2.

Table 8. Syllabification errors in the reading task and the repetition task by the Mandarin speakers (Totally 240 tokens). Nb. For "vegetable," either "vege-ta-ble" or "ve-ge-ta-ble" was counted correct.

Error types	Examples and frequency
Referring to foot	straw-berry (9); avo-cado (8); vege-table(5); coco-nut (4); bad-minton (3); ugly (1)
Exclude coda	frui-t (1); gra-pe (1); co-co-nu-t (3); mu-zz (3); sme-cks (2)
Separate clusters	f-ruit (2); sme-k-s (2); smeck-s(1); s-na (1)

From the syllabification, we can see that the Mandarin participants did not have problem in syllabifying the target consonant clusters, except M4 misjudged "ugly" as one syllable in the reading task (and J1 made exactly the same error as well). Errors were caused by other positions or other cluster. For the target clusters, phonologically Mandarin speakers accept the consonant clusters and assign them to the correct syllables, which rules out a vowel in misinterpreted L2 phonological representation. The participants had already acquired the phonological structure of the target clusters, but their phonetic abilities did not reach a ready state. About error patterns, participants syllabified the trisyllable words into feet, which are higher units in the prosodic hierarchy. In the monosyllable words, they exclude the codas or pull apart the consonant clusters, suggestion the participants preferred universally unmarked CV syllables. However, such errors also have the possibility to the results of native language transfer, because CV and CVC syllables are the basic syllable structures in Mandarin Chinese. In any case, the participants were on the way to acquire the English syllabification and prosody.

As for Japanese speakers, contrary to our expectation that they may syllabify English words based on morae, the actual error patterns were quite inconsistent. For example, J2 divided "webnet" into "we-b(ɯ)-ne-t(ɯ̥)" referring to a mora while dividing "icecream" into "i-ce-cream." Individual differences also varied: e.g. the nonsense word "gamboozee" syllabified as "ga-m-boo-zee" in reading but correctly in repetition by J2, as "gam-boozee" in repetition but correctly in reading by J4, as "gamboo-zee" in repetition but correctly in reading by J5, and as "ga-m-boo-oo-zee-ee" in both tasks by J6. Even the CV sequences "banana" was incorrect six times, like "ban-nana" in repetition by J2 and "banana" without division by the others. This implies that the Japanese participants did not know what to do in the syllabification task. We interpret that their errors were random or pre-systematic errors. Their correct answers were also likely due to chance, except for J1 and J8. This time, we simply counted the numbers of errors in each word due to their random error patterns, shown in Table 9. The only words correctly syllabified by Japanese speakers were the one-syllable CV nonsense words "ba" and "ma." Unlike Chinese speakers, Japanese speakers made errors with the target clusters.

Table 9. Syllabification errors in the reading task and the repetition task by the Japanese speakers (Totally 320 tokens). Nb. For "vegetable," either "vege-ta-ble" or "ve-ge-ta-ble" was correct.

Word types	Words and frequency
3-or-more-syllable wrd.	badminton (7); avocado (9); banana (6); coconut (10) ; strawberry (13); vegetable (10); cantukpeg (5); gamboozee (5)
2-syllable wrd.	subject (7); webmail (4); webnet (4); tablet (4); foodbank (4); Sydney (6); badly (4); rugby (3); eggman (4); magnet (4); ugly (8); ebdet (5); gabmee (4); gabno (9); cabla (7); idgay (5); cadma (5); pednay (4); edlee (8); agday (4); egmad (3); hegneb (5); agla (5); jeejee (1)
1-syllable wrd.	fruit (4); grape (5); ice-cream (5); muzz (4); smeks (7); s-na (7)

What we are focusing on is not how Japanese speakers syllabified, but how they pronounced each word in the syllabification task. For example, J6 correctly pronounced the consonant cluster in "eggman" without releasing /g/ in the repetition task, while when she syllabified in the repetition task, she pronounced [ɛ-gɯ-mæn] with clear [ɯ] after [g]. Other participants also tended to add clear [ɯ] or [o]. Even if they heard a native English speaker's production without vowel insertion, they still heard "illusory" [ɯ] or [o] as Dupoux et al. (1999) mentioned. This indicates that even though the participants could hear and pronounce target-like consonant clusters at the phonetic level as Funatsu et al. found, they still heard vowels at the phonological level. In other words, [g] and [gɯ], for example, are allophonic variations of /gɯ/ in their interlanguage. Interestingly, when Japanese speakers correctly syllabified (by chance), Japanese speakers often did not insert a vowel, suggesting that not inserting a vowel at the surface representation (SR) is not the biggest problem. The more serious problem is that they misinterpreted English consonant clusters, assuming that there is a vowel in underlying representation (UR) in their interlanguages. Their vowel insertions are neither epenthesis nor intrusion, but their misinterpreted L2 phonological representation.

More interestingly, when syllabifying, J1 and J8 looked careful not to add extra vowels. When they happened to add a vowel, they self-corrected it, which gave us a glimpse of their, especially J8's, correct phonological representations. This rules out misinterpreted L2 representation. However, J1 and J8 sometimes inserted [ɯ] or [o] in the production tasks. To figure out these inserted vowels' true identity, we did a follow-up interview to J1 and J8, who gave two answers. First, their old bad habit, namely *Katakana*-English, sometimes came out when they got careless. Second, J8 sometimes focused too much on syllable division. The first case is considered as epenthesis, and the second case may be intrusion.

An even more interesting case is J2 who had taken a phonetics course in Japan. She frequently divided the stimuli into morae instead of syllables and added a vowel. However, the vowels added in wrong places were often quite short [ɯ̆] or [ŏ] and were often aspiration or devoiced vowels [ɯ̥] or [o̥] after

voiceless consonants: e.g. [sʌ-bɯ̆-dʒɛ-kɯ̥-to̥] "subject." She might have mistaken morae for syllables, but possibly she had some explicit knowledge of where to and not to pronounce a vowel. She might have attempted to pronounce consonants alone and have ended up with a short vowel or a devoiced vowel, as in /sʌ-b-dʒɛ-k-t/ → [sʌ-bɯ̆-dʒɛ-kɯ̥-to̥]. In her case, vowel insertion may be epenthesis, rather than misinterpreted L2 phonological representation. Still, further study will be needed.

Referring to Table 5 and 6, J8 most correctly pronounced consonant clusters. Interestingly, J8's overall accent was not necessarily the best, such as the issue of [l]. J1 and J2 were better in /l/. According to the limited number of samples, the relationship between explicit knowledge of syllabification and production of consonant clusters was observed. As for J1, who just started to learn syllable, she still seems under the development from *Katakana*-English to target-like English syllabification. J3 showed a dramatic difference between the reading task and the repetition task in production as shown in Table 5 and 6. It is probably because J3 was good at imitating sounds although she did not have explicit knowledge of syllable.

As for Mandarin speakers, they did not add a vowel in syllabification unlike Japanese speakers, except M5 occasionally did. M5 also frequently inserted vowels in the production tasks compared to others. Mandarin speakers, especially M2 and M5, frequently devoiced voiced stops in coda and aspirate them in syllabification. M2 and M5 also frequently aspirated voiced stops in the target clusters in the production tasks. In other words, Mandarin speakers correctly understand phonological representations, but either did not know the English phonetic characteristic or do know it but sometimes fail to produce it.

5 Discussion

5.1 Japanese speakers' L2 phonological representation

The most commonly inserted vowels by Japanese speakers were [o] before /d/ and [ɯ] elsewhere, the common epenthetic vowels in Japanese loanwords, which agrees with the results of Funatsu et al. (2008). In the real words, as Funatsu et al. interpreted, the participants likely associated the English words with the corresponding Japanese loanwords. In fact, their stress pattern also occasionally reflects their loanword interference. For example, P4, P5, P6, P7, and P8 put stress on "min" in "badminton," which corresponds to the Japanese loanword /badomíNtoN/ with the phonemic accent nucleus in /mí/. However, even in nonsense words the participants had never seen, they clearly inserted vowels. Nogita (2010) reported that many Japanese learners of English are never taught the English basic symbol-sound correspondence rules in the six years of English curriculum in Japan. Therefore, it is likely that Japanese learners of English have built their own English symbol-sound correspondence rules. Considering the fact that they almost always added an extra vowel, Japanese learners' interpretation of the English orthography is *abugida* or alphasyllabary, rather than alphabet. Each

consonant letter, probably except for <n> that corresponds to the moraic placeless nasal /N/, has a default following vowel which is pronounced every time a following vowel is required according to Japanese phonotactics; for example, the italicized consonant letters in "su*b*je*c*t" are pronounced as /bɯ/, /kɯ/, and /to/ with the default vowels although there are no vowels in the spelling. The followings are some examples of Japanese speakers' interpretation of the English symbol-sound correspondence rules: <b> - /bɯ/, <c> - /kɯ/, <t> - /to/ (or /tɯ/).

As for the repetition task, Japanese speakers less frequently inserted vowels, but they also released the first consonants in clusters, or devoiced the first consonants and aspirated them, as shown in Table 1, 5, and 6. Japanese speakers' aspiration in a consonant cluster is considered a voiceless vowel, often [ɯ] as mentioned in §4.2. This interpretation is consistent with the interpretation that there is a vowel after a consonant in their UR. When the participants released the first consonant, such as [b] in 'subject', the release was actually /ɯ/ in their mind, which was phonetically minimized. This interpretation is also consistent with their UR, rather than interpreting as gestural mistiming. Even if Japanese speakers pronounced target-like clusters, there was still /ɯ/ in their UR that was phonetically minimized to ø. In Japanese ESL learners' interpretation of English words where a vowel exists in the first place, the vowel is allophonically weakened or deleted. A learner misinterprets 'gb' as /gɯb/ but this /gɯ/ is altered to [gɯ], [k^h], [g] (with release), [g̚] and so forth in SR. We consider these vowel alternations as free allophonic variations. Figure 2 to 5 show examples of vowel alternations by the Japanese participants. Note that only F4 with clear vowel is from the reading task and the others are all from the repetition task.

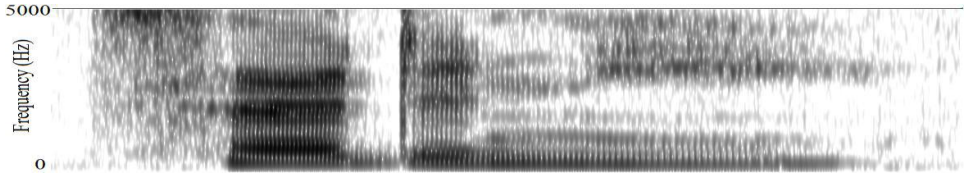

Figure 2. Full vowel: "Sydney" with [ɯ] insertion produced by J3 in the reading task

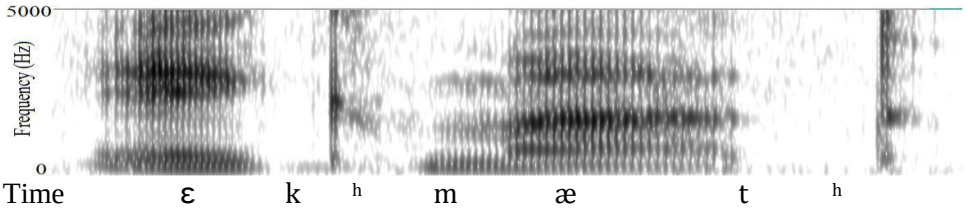

Figure 3. Weakening: "egmad" with aspirated /g/ produced by J4 in the repetition task

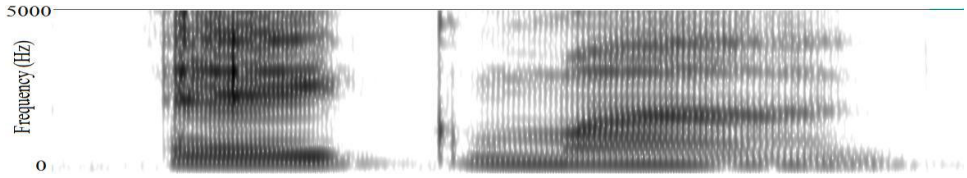

Figure 4. Weakening: "eggman" with release produced by J5 in the repetition task

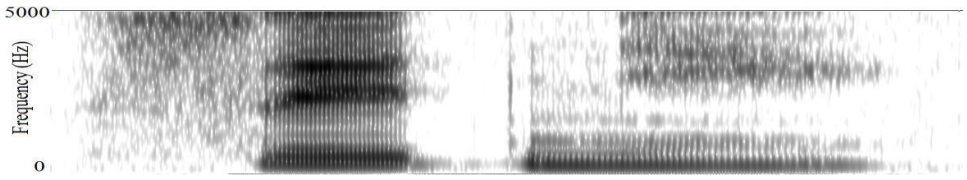

Figure 5. Deletion: "Sydney" with no release produced by J3 in the repetition task

J1 and J8 who had explicit knowledge of English syllabification and possibly J2 may have more similar processes to Mandarin speakers to some extent.

5.2 Mandarin speakers' process of vowel insertion

The Mandarin ESL learners did insert a vowel between two consonants but only in a few cases in the reading task and fewer in the repetition task. Like Japanese speakers, the vowels inserted in the repetition task are shorter than those in the reading task. Meanwhile, the syllabification tasks indicated that Mandarin speakers assigned these clusters correctly. In their processes of target consonant clusters, there was no inserted vowel underlyingly. Moreover, their inserted vowels were mostly schwas rather than lexical vowels, especially in the repetition task. Thus, we inferred that the inserted vowels were due to factors other than native phonological transfer. The Mandarin speakers still may have not fully acquired the articulatory timing patterns of English consonant clusters and they could not correctly manage the coordination of the clusters and break the close transition into open transition, which resulted in an audible sound. If articulatory mistiming or is true in the present study, participants' random insertion may be explained. Incidentally, M6's production of consonant clusters was the best among all the participants. He might have almost acquired target-like clusters in L2. His syllabification task was quite good as well. Figure 6 to 9 show Mandarin speakers short vowel insertion, aspiration, unreleased, and a nasal assimilation, all in the reading tasks.

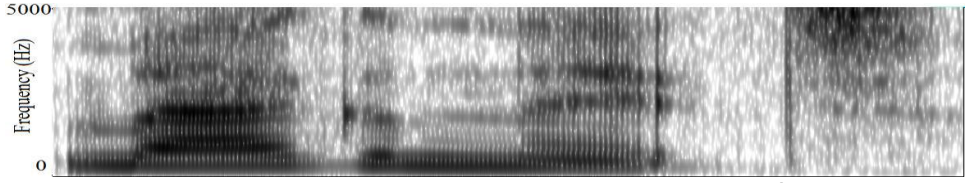

Figure 6. "magnet" with short [ə] insertion produced by M1 in the reading task

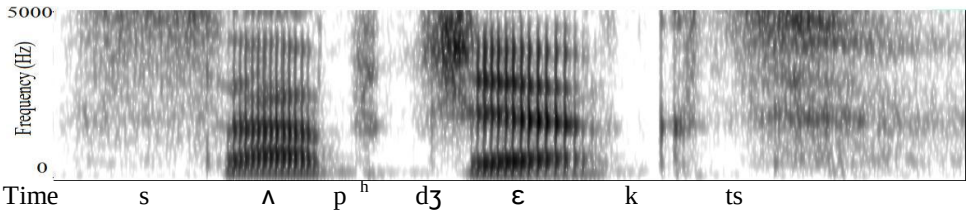

Figure 7. "subject" with devoiced and aspirated /b/ produced by M2 in the reading task

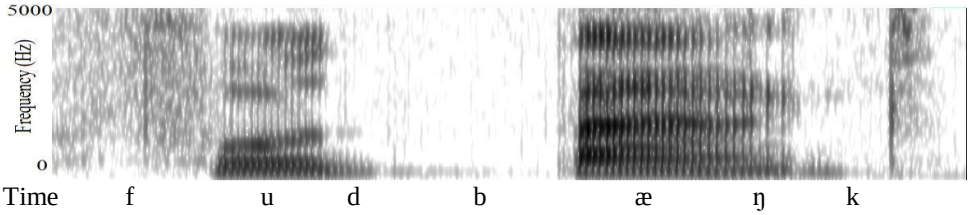

Figure 8. "foodbank" with no release produced by M6 in the reading task

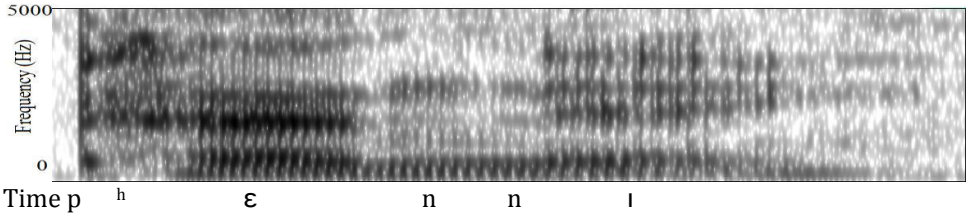

Figure 9. "pednay" with /d/ realised as [n] pronounced by M3 in the reading task

To sum up, Figure 10 and 11 show the three types of vowel insertion. Figure 10 shows (1) intrusion and (2) epenthesis, and Figure 11 shows (3) misinterpreted L2 phonological representation and the variations of SRs. The example is "rugby" that contains an illegal cluster /gb/ in Mandarin and Japanese.

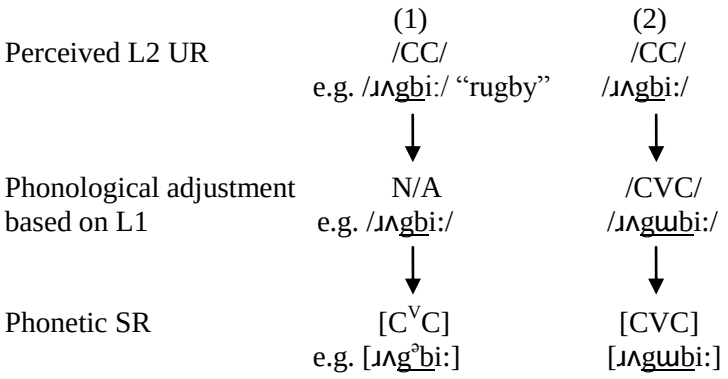

Figure 10. The difference among (1) intrusion, (2) epenthesis.

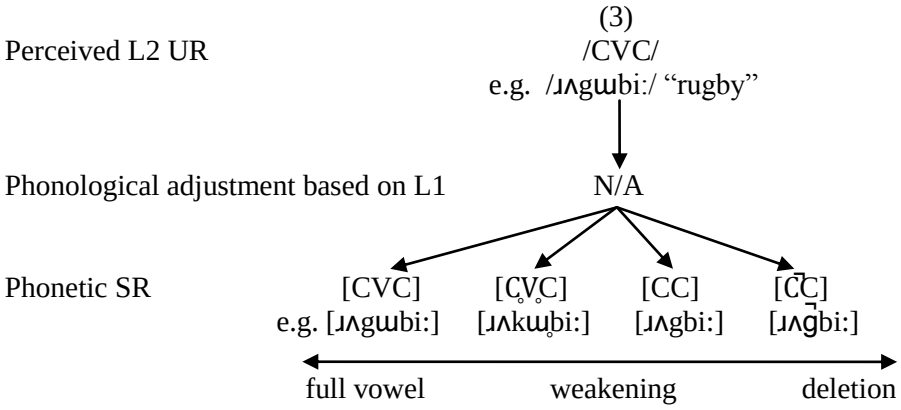

Figure 11. (3) Misinterpreted L2 phonological representation and phonetic manipulation.

5.3 Teaching Implication

There are three processes, intrusion, epenthesis, and misperceived L2 representation. Since interlanguages are in progress, a learner may not exclusively use only one process. Six of the Japanese participants' vowel insertion was mostly misperceived L2 representation. That of the mandarin participants was mostly intrusion, especially in the repetition task. That of J1 and J8 were somewhere between the other Japanese speakers and Mandarin speakers.

For six of the Japanese speakers, English syllabification should be explicitly taught. Their problem is that they do not know when to pronounce consonant clusters, but not necessarily how to pronounce clusters. They should first know the simple rule that if there is a vowel in the spelling, they are supposed to pronounce a vowel (except for silent letters as in "e" in "cake"), and otherwise, they are not supposed to. For the rest of the participants, since they pronounced consonant clusters better in the repetition task, auditory input can

help them acquire the gestural timing of coordination in consonant clusters in the target language.

5.4 Limitation and future study

The major limitation is that it is still not clear whether the participants' inserted vowels with shorter durations were vowel intrusion or epenthesis. In the future research, detailed acoustic measurement should be conducted to detect the gestural mistiming in the production of both Japanese and Mandarin speakers. Also, an ultrasound study is needed to investigate this phenomenon from an articulatory perspective. As well, if Japanese ESL learners' vowel insertion is a knowledge issue, there needs to be an experiment consisting of pre-test, lesson, and post-test in order to examine whether knowledge of syllabification will improve their production of consonant clusters. We also attributed Japanese speakers' more frequent vowel insertion to the difference in language education between Japan and China, but we have to more carefully investigate the educations in these two countries to back up our interpretation. In the methodology, as the English /l/ is notoriously problematic for Japanese ESL learners, all but J1 and J2 almost always pronounced more flap-like or more rhotic liquids. Therefore, we could not observe true C-[l] clusters by these participants. As well, in the present study, the target clusters were across the syllable boundaries, so that tautosyllabic clusters were not examined. We do not know whether we can generalize our interpretations in tautosyllabic clusters.

6 Conclusion

Assuming that vowel insertion in English consonant clusters is always vowel epenthesis due to transfer of L1 phonology is too simplistic. In English consonant clusters that are not allowed in either Mandarin Chinese or Standard Japanese, the Mandarin and Japanese ESL learners in our experiment inserted a vowel with different processes, mainly because of their explicit knowledge of English pronunciation. The Mandarin speakers' vowel insertion was much less frequent than that of the Japanese speakers. Two syllabification tasks confirmed that the Mandarin participants phonologically assigned the two consonants correctly to different syllables, which might indicate that the inserted vowels may not be a problem in UR. Their inserted vowels were often short schwa-like vowels rather than lexical vowels. Therefore, we interpret their productions to be a case of intrusion caused by gestural mistiming. Still, there is detailed measurement needed to judge whether the results indicate epenthesis or intrusion. In contrast, in the case of six of the Japanese participants, a vowel in consonant clusters existed in the first place in their interlanguage phonological interpretation because most Japanese learners of English are exposed to almost standardized *Katakana*-English whose orthography is *abugida* or alphasyllabary; each consonant-letter has a default following vowel which is pronounced every time a following vowel is phonotactically required. Also, none of our subjects had taken

formal instruction of English syllabification during the six years of English curriculum in Japan. However, all the Japanese participants successfully deleted those incorrectly existing vowels at least several times in the repetition task. Such vowel deletion or weakening are free allophonic variations and there is still a vowel in their UR. Therefore, even when they phonetically heard the stimulus [sʌb̚dʒɛk̚t], for example, they syllabified it as /sʌ.bɯ.dʒɛ.kɯ.to/ or the like. This can support Dupoux et al.'s (1999) conclusion that Japanese speakers hear an illusory vowel between an illegal CC sequence. Two of the Japanese participants, who recently learned English syllabification in Canada, behaved more like Mandarin speakers, except their old habit, *Katakana*-English, still came out, which is considered epenthesis. Because Mandarin speakers had known syllables much longer than these two Japanese speakers, Mandarin speakers may have been almost internalized L2 phonological structures and could pronounce consonant clusters without vowel epenthesis or phonological adjustment, while these two Japanese speakers who were new to syllable still had phonological L1 interference. Meanwhile, both Mandarin and Japanese speakers inserted a shorter vowel and less often in the repetition task than in the reading task. This indicates that both speakers can detect and produce the phonetic (not necessarily phonological) difference between CC and CVC, which supports Funatsu et al.'s (2008) conclusion that Japanese speakers can perceive and produce consonant clusters.

Acknowledgements

We are grateful to Dr. Ewa Czaykowska-Higgins for her guidance in this study and expertise in phonology. We are also grateful to all the participants.

References

Akamatsu, T. (2000). *Japanese phonology*. LINCOM Studies in Asian Linguistics 38.

Carlisle, R. (1998). The acquisition of onsets in a markedness relationship: A longitudinal study. *Studies in Second Language Acquisition*, 20, 245-260.

Coulmas, S, (2003). *Writing System*. New York: Cambridge University Press.

Dehaene-Lambertz, G., Dupoux, E. & Gout, A. (2000). Electrophysiological correlates of phonological processing: A cross-linguistic study. *Journal of Cognitive Neuroscience, 12*(4), 635.

Davidson, L., & Stone, M. (2003). Epenthetsis Versus Gestural Mistimimg in Consonant Cluster Production: An Ultrasound Study. In G. Garding & M. Tsujimura (eds.), *WCCFL 22 Proceedings*. Cambridge, MA: Cascadilla Press, 165-178.

Davidson, L., Jusczyk, P., & Smolensky, P. (2004). The initial and final states: Theoretical implications and experimental explorations of richness of the base. In R. Kager, W. Zonneveld, & J. Pater (Ed), *Fixing Priorities:*

Constraints in Phonological Acquisition. Cambridge: Cambridge University Press.

Davidson, L. (2006). Phonology, phonetics, or frequency: Influences on the production of non-native sequences. *Journal of Phonetics*, 34, 104-137.

Dupoux, E., Hirose, Y., Kakehi, K., Mehler, J., & Pallier, C. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology, 25*(6), 1568-1578.

Eckman, F. (1991). On the naturalness of interlanguage phonological rules. *Language Learning, 31*, 195–216.

Eckman, F., & Iverson, G. (1993). Sonority and markedness among onset clusters in the interlanguage of ESL learners. *Second Language Research*, 9, 234–252.

Eckman, F., & Iverson, G. (1994). Pronumciation difficulties in ESL: Coda consonants in English interlanguage. In Yavas, M. (Ed.), First and Second Language Phonology, 251-266. San Diego, CA: Singular.

Funatsu, S., Imaisumi, S., Fujimoto, A., Hashizume, M., & Kurisu, K. (2008). Do Japanese speakers perceive nonexistent vowels in non-native consonant clusters?. *The Journal of the Acoustical Society of America, 123*(5), 3072.

Gafos, A. I. (2002). A grammar of getural coordination. *Natural Language & Linguistic Theory*, 20, 269-337.

Goad, H. & Kang, H. (2003). Word-final syllabification in L2 acquisition with emphasis on Korean learners of English. In Juana M. Liceras, Helmut Zobl & Helen Goodluck (eds.) *Proceedings of the 6th Generative Approaches to Second Language Acquisition Conference*. Somerville, Mass.: Cascadilla Press. 122–129.

Goble, D. (2001). Loanword-induced interference in Japanese students' foreign language acquisition: Developing student awareness through experiential learning . *Journal of Chikushi Jyogakuen Junior College, 36*, 55-77.

Hall, N. (2006). Cross-linguistic patterns of vowel intrusion. *Phonology 23*. The United Kingdom: Cambridge University Press. 387-429.

Hansen, J. G. (2001). Linguistic constraints on the acquisition of English syllable codas by native speakers of Mandarin Chinese. *Applied Linguistics*, 22, 338-365.

Inozuka, E., Inozuka, H., & Machida, K. (Ed.). (2009). *Nihongo onsêgakuno shikumi*. Tôkyô, Japan: Kenkyûsha.

Kim, D. Y., & Jung, Y. (1998). Adult ESL Korean speakers' interlanguage acquisition of consonant clusters. *SECOL Review*, 22, 150-169.

Lin, H. (2008). Interlanguage rhythm in the English production of Mandarin speakers. *Proceedings of the 8th Phonetic Conference of China and the International Symposium on Phonetic Frontiers*. Beijing.

Lin, Y. (2001). *The Sounds of Chinese*. Cambridge University Press: United Kingdom. 115-117.

Makino, T. (2008, September). *English pronunciation teaching and phonetic study*, Presented a Phonetics Mini Symposium (1), Ôsaka, Japan.

Mao, R. P. (2005). Loanword adaptation in Mandarin Chinese: Perceptual, phonological and sociolinguistic factors. Ph D. Dissertation in Stony Brook University. 105-110.

Martin, A., (2004). The 'katakana effect' and teaching English in Japan. *English Today, 20*(1), 50-56.

Nishi, K., Strange, W., Akahane-Yamada, R., Kubo, R., & Trent-Brown. S. A. (2008). Acoustic and perceptual similarity of Japanese and American English vowels. *The Journal of the Acoustical Society of America, 124*(1), 576-588.

Nogita, A. (2010). Do Japanese ESL learners' pronunciation errors come from inability to articulate or misconceptions about target sounds?.*Working Papers of the Linguistics Circle of the University of Victoria, 20*, 82-116.

Ôno, S. (2007). *Nihongono genryûo motomete*. Tôkyô, Japan: Iwanami-shoten.

Pitt, M. A. (1998). Phonological process and perception of phonotactically illegal consonant clusters. *Perception & Psychophysics*, 60 (6), 941-951.

Smith, J. (2006). Loan phonology is not all perception: Evidence from Japanese loan doublets. in Timothy J. Vance & Kimberly A. Jones (eds.). *Japanese/Korean Linguistics 14. Palo Alto: CLSI.*, 63-74.

Sun, Ch. F. 2006. *Chinese: A Linguistic Introduction*. UK: Cambridge University Press.

Urbanczyk, S. (1996). *Patterns of Reduplication in Lushootseed*. Ph.D. dissertation, University of Massachusetts.

Ueyama, M. (2003) Awareness of L2 syllable structure: The case of L2 Japanese and L2 English. *Journal of the Phonetic Society of Japan*, *7*(2), 84-100

Vance, T. J. (2008). *The sound of Japanese*. New York: Cambridge University Press.

Weinberger, S. H. (1987). The influence of linguistic context on syllable structure simplification. In G. Ioup & S. H. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system*, 401-417. Rowley, MA: Newbury House.

Weinberger, S. H. (1994). Functional and phonetic constraints on second language phonology. In M. Yavas (Ed.), *First and second language phonology*, 283-302. San Diego, CA: Singular Publishing Group.

Yoo, H. (2004). A longitudinal study of consonant cluster acquisition. *Studies in Phonetics, Phonology and Morphology* 10.3. 481-503.