Speech Prosody 2016
31 May - 3 June 2016, Boston, USA

The influence of second language experience on Japanese-accented English rhythm

Saya Kawase, Jeesun Kim, Chris Davis

The MARCS Institute, Western Sydney University, Sydney, NSW, Australia
s.kawase@westernsydney.edu.au

Abstract

Compared to stress-timed English, mora-timed Japanese is characterized by a simpler syllabic structure and no vowel reduction. Such differences may explain some aspects of the problems that Japanese talkers have in producing English speech rhythm, i.e., an L1 influence on L2 rhythm production. The present study tested whether this L1 influence on L2 could be moderated by an increase in L2 experience. We examined English sentences spoken by Japanese ('experienced' and 'inexperienced' English learners) and native Australian English talkers. The mean duration and variability of consonant and vowel intervals were calculated using rhythm metrics. The results showed that the mean duration of phoneme intervals was relatively longer in L2 speech, particularly the inexperienced L2, compared to L1 speech. Furthermore, the inexperienced L2 talkers exhibited the least vowel durational variability, with the English talkers having the most; the values of the experienced L2 talkers were intermediate. Differences among the talker groups were well described by the coefficient of variations of vowel and consonant durations, more specifically, durational variability increased as the phoneme duration got shorter. Overall, the results demonstrated that an L1 influence on L2 speech rhythm production decreases as a function of L2 experience.

Index Terms: Speech rhythm, duration, second language speech production

1. Introduction

It is widely known that second language (L2) speech production is influenced by native language (L1) experience, particularly when the two languages differ markedly in speech properties. Such differences can be at the segment level and affect L2 segment production. For example, Japanese talkers tend to assimilate English /l/ and /ɹ/, which do not exist in Japanese, to the Japanese lateral flap /ɾ/ [1, 2]. Languages can also differ at the supra-segmental level. For example, stress and vowel reduction exist in stress-timed languages (e.g., English, Dutch) but not in syllable-timed languages (e.g., Spanish, French) nor in mora-timed languages (e.g., Japanese) [3]. The different rhythm types between L1 and L2 have also been shown to affect how L2 rhythm is produced [e.g., 4-10]. Extensive research demonstrated the influence of L1 segments on L2 ones [e.g., 1, 2, 11], yet relatively fewer studies have examined how native suprasegmental properties affect non-native speech production. The current study therefore aimed to investigate the L1 influence on L2 rhythm production in a systematic manner and also determine the role of L2 experience in L2 rhythm production.

In attempts to quantify speech rhythm, three different types of rhythm metric have been primarily used: interval measures, rate-normalized measure and the Pairwise Variability Index (PVI). Ramus et al [12] proposed interval measures to measure durational variability in vowel and consonant durations, including the standard deviation of the duration of the vocalic interval (ΔV; located between the onset and offset of vowels or vowel clusters) within each sentence; the standard deviation of the duration of the consonant interval (ΔC; located between the onset and offset of consonants or cluster of consonants) within each sentence; and the proportion of vocalic intervals (%V; the sum of vocalic intervals divided by the total duration of sentences).

Whilst the interval measures are still extensively used in many studies, it has been pointed out that these are strongly influenced by speaking rate [13]. Thus, Dellwo [14] proposed rate-normalized interval measures for vowel (VarcoV: coefficient of variation of vowel interval) and consonant intervals (VarcoC: coefficient of variation of consonantal interval), which normalize ΔV and ΔC for speaking rate. Furthermore, Grabe and Low [15] introduced an additional metric, namely the Pairwise Variability Index (PVI) of vocalic and consonantal intervals. This measure considers the variability of all pairs of vocalic and intervocalic intervals in speech, so the ordering of intervals within a sentence can be taken into account. Overall, the above three different types of measures have been developed to capture durational variability of consonant and vowel durations in different rhythm classes with L1 speech, and have also been used to examine L2 speech.

Although some L2 rhythm production studies have shown an influence of L1 rhythm, the results have not always been consistent across studies [e.g., 4-10]. For example, Spanish speakers of British English (syllable-timed speakers of stress-timed language) showed smaller vowel durational variability with VarcoV compared to the native British English speakers [10]. In contrast, an L1 rhythm influence did not occur with different syllable-timed speakers of stress-timed language such as French speakers of British English [9] as well as Mandarin and Cantonese speakers of British English [8]. Even with same L1-L2 rhythm comparisons, inconsistent findings have been reported, e.g., Carter [4] found an influence of L1 on durational variability on successive vowel intervals (nPVI-V) on L2 English among Spanish-English bilinguals, but White and Matty [10] did not find the influence.

Unlike cross-linguistic comparisons with different L1 groups, studies of L2 rhythm production need to carefully consider several factors that L1 studies do not. Firstly, L2 proficiency may play a role in the degree of L1 influence on L2 rhythm production. For example, Grenon and White [5] did

10.21437/SpeechProsody.2016-153

not find an effect of vowel durational variability (i.e., VarcoV) for Japanese learners of English although their VarcoV in L1 Japanese were significantly smaller than L1 English speakers (Japanese doesn't possess stress nor vowel reduction). It could be the case that the selected participants were advanced learners of English, so they may not have shown an L1 influence. To our knowledge, limited number of studies investigated levels of L2 proficiency when examining L2 rhythm [6, 16]. In particular, it remains to be understood how L2 rhythm develops over time. Extensive studies and models of L2 speech learning (e.g., Perceptual Assimilation Model of Second Language Speech Learning (PAM-L2) [17]) have indicated the importance of the first 12 months with regards to their speech learning, thus here we aim to examine the L2 rhythm development by comparing two groups of L2 talkers differing their length of residence (i.e., 6 months vs. 12 months).

In addition, although previous researchers have applied the metrics developed for L1 speech to L2 rhythm studies, some metrics may not successfully capture durational variability in L2 due to differences in overall duration between L1 and L2 speech. It is known that duration of L2 speech tends to be longer than L1 speech [e.g., 18]. Given this, the rhythm metrics that use standard deviations, ΔC and ΔV, may not properly capture the durational variability, particularly when comparing L1 to L2 speech. For instance, previous research using ΔC has found that advanced Chinese speakers of Canadian English exhibited a similar consonant variability compared to the native Canadian English speakers, suggesting that they have acquired their L2 stress-timed rhythm [7]. However, this result may be due to the longer speech duration among the Chinese speakers, rather than similar durational variability between the native and non-native speakers. It is thus important to use rate-normalized interval measures, namely VarcoV and VarcoC, in order to capture durational variability especially when comparing groups who differ in terms of the overall vowel and consonant durations.

Taken together, the present study examines the production of L2 rhythm by Japanese learners of English as a function of L2 experience (experienced vs inexperienced). Given that L2 speech is relatively longer than L1 speech [18], the current study considers overall duration in vowel and consonant intervals in addition to their variability. That is, variability of L2 vowel and consonant intervals would be well captured using rate-normalized metrics (i.e., VarcoV, VarcoC), rather than the standard deviations (i.e., ΔV, ΔC). It is also expected that there are a number of potential L1 (Japanese) influences on L2 (Australian English) durational variability. For instance, although both Australian English and Japanese have phonemic vowel duration, Japanese does not have vowel reduction, leading to Japanese learners being less successful in producing vowel reduction in English [19]. In addition, unlike Japanese, English has stress, so there is a further durational contrast between stressed and unstressed syllables. Thus, it was predicted that for Japanese L2 learners there would be less variability in vowel duration, particularly with the inexperienced Japanese learners of English who are more likely to exhibit an L1 influence compared to the experienced learners.

As for the consonant intervals, we also expect the similar reduced variability among the Japanese production (with a further decrease in the inexperienced Japanese one). Previous research showed that Japanese learners of English have difficulty in producing consonant clusters, resulting in occurrence of epenthetic vowels [e.g., 20-22] particularly with a slower speaking rate [20]. As Japanese syllable structure only allows simple onset, producing non-native complex onset might be difficult due to their native language influence. Another articulatory finding using an electromagnetic articulograph (EMA) also suggests a weaker coarticulation effect between the first and second consonants for Japanese speakers [23]. In turn, Japanese learners of English may produce a relatively fixed duration in consonant production regardless of different syllable structures in English. Thus, it was predicted that there would be reduced variability in consonant duration, particularly with the inexperienced Japanese learners of English who are more likely to exhibit an L1 influence compared to the experienced learners. Finally, we also predict that the reduced variability in both vowel and consonant durations would additionally manifest in less variability in successive vocalic and consonant intervals (i.e., nPVI-V, rPVI-C).

2. Method

2.1. Participants

The talkers were selected from our database consisted of audio and video recordings from 10 native English and 19 Japanese talkers. For this study, we analyzed four Japanese and two Australian English talkers' production (all female). All of them resided in Sydney at the time of recording. The Japanese talkers consisted of two 'inexperienced talkers' of English whose mean length of residence in Australia was relatively short (M_{age} = 29.0; M_{LOR} = 4.5 months) and two 'experienced talkers' who have lived in Sydney for more than a year (M_{age} = 29.0 years; M_{LOR} = 12.5 months). All participants started to learn English as a foreign language in Japan at approximately age 13. The two monolingual Australian English talkers were postgraduate students were also recruited at the University of Western Sydney. All participants reported no history of speech, vision or hearing problems.

2.2. Materials

The materials consisted of a total of 60 IEEE Harvard Sentences produced by each of the two experienced and two inexperienced Japanese and two Australian talkers. The sentences were segmented into vocalic and consonantal intervals (n = 1418) using Praat [24] by following commonly used conventions [10]. All the segmentations were conducted by the first author.

2.3. Acoustic measurements

All of the intervals were extracted using a tailored Matlab script (MATLAB R2013a). Following convention, pauses and disfluencies were excluded. Several irregular pauses and vowel insertion in consonant clusters were observed in the productions of the English sentences by the inexperienced Japanese talkers, but the current analyses did not take these into account. A set of rhythm metrics were used to analyze durational variability of vowel and consonant intervals including ΔC, ΔV, %V, VarcoC, and VarcoV. For the pairwise variability index (PVI), successive intervals were normalized only for the vowel intervals (nPVI-V), but not for the consonant intervals (rPVI-C) by following Grabe and Low [15]. Table 1 summarizes the rhythm metrics used in this study.

We also measured sentence, consonant, and vowel durations for each talker. While previous studies did not take into account the overall speech duration, it is important to consider as duration of L2 speech tends to be longer [18], resulting in a larger value in ΔC and ΔV due to a larger mean value, not because of larger variability per se. Thus, the current study includes additional durational measures including overall sentence, consonant and vowel durations for each talker group.

Metrics	Description
ΔC	Standard deviation of the duration of the consonant interval
ΔV	Standard deviation of the duration of the vocalic interval
%V	Sum of vocalic intervals divided by the total duration sentences
VarcoC	Coefficient of variation of consonantal interval duration
VarcoV	Coefficient of variation of vocalic interval duration
rPVI-C	Pairwise variability index for consonantal intervals. Mean of the differences between successive consonantal intervals
nPVI-V	Normalized pairwise variability index for vocalic intervals. Mean of the differences between successive vocalic intervals divided by sum, multiplied by 100

Table 1: *Summary of rhythm metrics*

3. Results

Separate ANOVAs were run for all the rhythm metrics described in Table 1 (ΔC, ΔV, %V, VarcoC, VarcoV, rPVI-C, nPVI-V) as well as for the mean sentence (Sent), consonant (C) and vowel (V) durations, with the three talker groups (NJ-I, NJ-E, NE) as the between-group factor. First, results of mean duration are presented in 4.1. Results of a series of rhythm metrics are presented in section 4.2. and 4.3.

3.1. Mean duration of Sent, C and V

Figure 1 shows mean durations of consonant and vowel intervals by inexperienced (NJ-I), experienced Japanese (NJ-E) talkers as well as native English (NE) talkers.

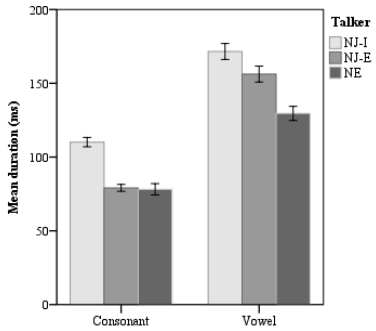

Figure 1: *Mean consonant and vowel duration by inexperienced (NJ-I), experienced Japanese (NJ-E) and native English talkers (NE) talkers. Error bars indicate +/- one standard error.*

The ANOVAs for mean sentence, consonant, and vowel durations showed significant effects of talker group on sentence [$F(2, 57) = 47.951$, $p < .001$], consonant [$F(2, 57) = 32.888$, $p < .001$], and vowel durations [$F(2, 57) = 16.527$, $p < .001$]. Bonferroni adjusted post hoc tests revealed that the mean sentence duration by the inexperienced Japanese (3493.9ms) were significantly longer than both the experienced Japanese (2629.1ms) and the native English (2318.7ms) talkers ($p < .001$). Similarly, the mean consonant duration of the inexperienced Japanese (110.1ms) were significantly longer than both the experienced Japanese (79.1ms) and the native English (78.2ms) talkers ($p < .001$). Mean vowel duration was significantly shorter for the native English (129.6ms) compared to both inexperienced (171.6ms) and experienced Japanese (156.2ms) talkers ($p < .001$). There were no other significant differences.

3.2. Durational variability (ΔC, ΔV, VarcoC, VarcoV)

Figure 2 shows mean values for the durational variability indices for inexperienced (NJ-I), experienced Japanese (NJ-E) talkers as well as native English (NE) talkers. Separate ANOVAs for ΔC, ΔV, VarcoC, and VarcoV with three talker groups (NJ-I, NJ-E, NE) as the between-group factor were conducted. The results showed significant effects of talker group on ΔC [$F(2, 57) = 5.660$, $p < .01$], VarcoV [$F(2, 57) = 8.736$, $p < .001$], but not for ΔV or VarcoC ($p > .05$).

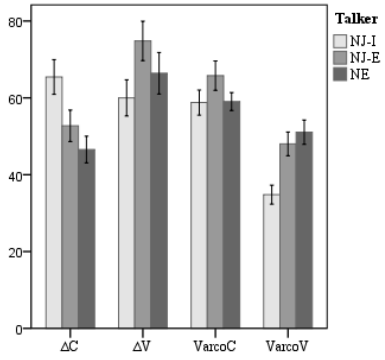

Figure 2: *Durational variability (ΔC, ΔV, VarcoC, VarcoV) of inexperienced Japanese (NJ-I), experienced Japanese (NJ-E) and native English (NE) talkers. Error bars indicate +/- one standard error.*

Bonferroni adjusted post hoc tests revealed that for ΔC, the native English talkers (65.4ms) showed the greater variability in consonant duration compared to the inexperienced Japanese talkers (46.6ms) ($p < .01$); but they did not differ significantly from the experienced Japanese talkers (52.7ms) ($p > .05$). For the VarcoC which also indexes the durational variability (controlled by the mean consonant intervals), there was no group difference across the talker groups.

For the variability of vowel duration, only the VarcoV showed a group difference. The native English exhibited greater variability in the vowel duration (51.1) compared to the inexperienced Japanese (34.8) ($p < .001$), but not compared to the experienced Japanese (48.0) ($p > .05$). In addition, among the Japanese talkers, the experienced Japanese showed greater variability in vowel duration compared to the inexperienced ones ($p < .01$).

3.3. Pairwise variability indices (rPVI-C, nPVI-V)

Figure 3 shows mean values for pairwise variability indices for consonant and vowel durations for the inexperienced (NJ-I), experienced Japanese (NJ-E) talkers as well as for the native English (NE) ones. Separate ANOVAs were conducted for rPVI-C, nPVI-V with three talker groups (NJ-I, NJ-E, NE) as the between-group factor. The results showed significant effects of talker group on nPVI-V [$F(2, 57) = 5.530$, $p < .01$], and on rPVI-C [$F(2, 57) = 3.859$, $p < .05$].

Bonferroni adjusted post hoc tests revealed that the native English talkers showed the larger pairwise vowel interval variability and less pairwise consonant interval variability (nPVI-V = 72.2; rPVI-C = 49.4) compared to the inexperienced Japanese talkers (nPVI-V = 50.3; rPVI-C = 63.5) ($p < .01$). The experienced Japanese were placed as the intermediate between the two groups and experienced Japanese (nPVI-V = 57.2; rPVI-C = 54.3), but the differences were not statistically differed from either of the other groups.

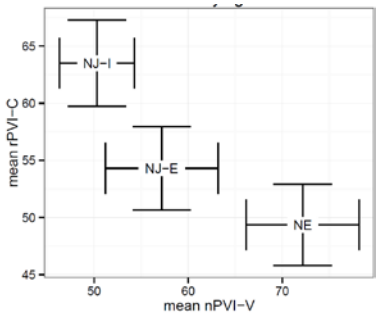

Figure 3: *Pairwise variability indices (rPVI-C, nPVI-V) of inexperienced Japanese (NJ-I), experienced Japanese (NJ-E) and native English (NE) Error bars indicate +/- one standard error.*

4. Discussion and conclusions

The aim of this study was to investigate speech rhythm for non-native English production. In particular, this study examined how different English experience levels among Japanese learners of English affected the variability of consonant and vowel intervals as these characterize speech rhythm. As previous L2 rhythm studies did not take L2 experience levels into account, potentially previous inconsistent findings may result from the variations in L2 production due to different L2 experience levels. To test this, experienced and inexperienced Japanese learners of English were recruited. Furthermore, the current research took into account the overall duration of consonant and vowel intervals (i.e., rate-normalized) when calculating their durational variability, since the relatively duration of L2 compared to L1 speech is problematic when only considering their standard deviations [e.g., 16].

The findings indicate that L2 rhythm is influenced by the learner's experience level, with this influence was clearest in the variability of vowel duration. It was found that the native English talkers and experienced Japanese talkers showed larger variability in vowel duration compared to the inexperienced Japanese talkers. The smaller variability of the latter group can be attributed to an influence of their L1, as Japanese does not employ stress or vowel reduction [3], it appears that the inexperienced Japanese talkers were not successful in producing vowels of variable duration. The overall vowel durations also reflect differences in the L1 vs L2 languages as well as L2 proficiency. Vowel duration overall was longest among the inexperienced Japanese, and shortest among the native English. The experienced Japanese also showed longer vowel durations (yet shorter than the inexperienced talkers) compared to native English talkers. Further research is necessary to investigate how vowel duration and durational variability are learned in L2; the current findings showed that experienced Japanese talkers did produce variability in vowel duration for English although their overall vowel duration was still longer than the native English ones.

Relative alternations of vowel duration on successive syllables were also found to be largest among the native English, of intermediate size for the experienced Japanese, and smallest for the inexperienced Japanese talkers. The current study found differences for both vowel-to-vowel variability and vowel durational variability whereas previous research on differences between native and non-native English talkers has only found difference in the latter (i.e., Spanish speakers of English in [10]). That is, in addition to variation in the variability of vowel duration, the Japanese talkers, particularly the inexperienced group, also showed less vowel-to-vowel variation on successive syllables. This result suggests that inexperienced Japanese not only exhibit a smaller durational variability in vowels, but also show less temporal alternation in vowel duration across syllables.

The current study also found variability in consonant duration (ΔC), which is consistent with [7]. The consonant duration variation for the inexperienced Japanese productions was significantly smaller compared to that found for native English talkers. However, the mean consonant duration among the inexperienced Japanese talkers was also significantly longer than the native English ones. The results revealed that VarcoC, the coefficient of variation of consonantal intervals, did not show a significant group difference across the native English, inexperienced and experienced Japanese talker groups, suggesting that consonant variability did not significantly differ when the overall durational differences across the groups were taken into account. However, further analyses with measuring consonant durations with respect to the syllable position (i.e., the first and second consonants) may exhibit the group difference given that previous findings suggest the difficulty in producing English consonant clusters among L2 Japanese [20-23].

These findings highlight the rhythm characteristics of Japanese-accented English and further reveal how L2 proficiency levels affect L2 rhythm. We are currently conducting further analyses to increase the number of talkers and to include Japanese data as well. Hopefully, the current results provide valuable insights into understanding L2 rhythm in Japanese-accented English and have implications for using rhythm measures, particularly in L2 studies. The findings demonstrate the importance of controlling for overall duration in order to accurately interpret the durational variability.

5. Acknowledgements

The authors would like to thank Leo Chong for assisting in preparation of stimuli. We would also like to extend our gratitude to the participants. This study has been supported by an ARC Discovery grant (DP130104447).

6. References

[1] K. Miyawaki, J. J. Jenkins, W. Strange, A. M. Liberman, R. Verbrugge, and O. Fujimura, "An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English," *Perception & Psychophysics*, vol. 18, pp. 331–340, 1975.

[2] A. Sheldon, and W. Strange, "The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception," *Applied Psycholinguistics*, vol. 3(3), pp. 243-261, 1982.

[3] R. M. Dauer, "Stress-timing and syllable-timing reanalyzed," *Journal of Phonetics*, vol.11, pp. 51-62, 1983.

[4] P. M. Carter, "Quantifying rhythmic differences between Spanish, English, and Hispanic English," *Amsterdam Studies in the Theory and History of Linguistic Science Series 4*, 272, pp. 63-75, 2005.

[5] I. Grenon, and L. White, "Acquiring rhythm: A comparison of L1 and L2 speakers of Canadian English and Japanese," *Proceedings of the 32nd Boston University conference on language development*, 2008.

[6] C. Lai, K. Evanini, and K. Zechner, "Applying rhythm metrics to non-native spontaneous speech," *Proceedings of SLaTE 2013*, 2013.

[7] H. Lin, and Q. Wang, "Vowel quantity and consonant variance: A comparison between Chinese and English," *Proceedings of between stress and tone*. Leiden, June 2005.

[8] P. P. Mok, and V. Dellwo, "Comparing native and non-native speech rhythm using acoustic rhythmic measures: Cantonese, Beijing Mandarin and English," *Proceedings of Speech Prosody*. 2008.

[9] A. Tortel, and D. Hirst, "Rhythm metrics and the production of English L1/L2." *Proceedings of Speech Prosody*, 2010.

[10] L. White, and S. L. Mattys, "Calibrating rhythm: First language and second language studies," *Journal of Phonetics*, vol. 35(4), pp. 501-522, 2007.

[11] O.-S. Bohn, and J. E. Flege, "The production of new and similar vowels by adult German learners of English," *Studies of Second Language Acquisition*, vol. 14, pp. 131–158, 1992.

[12] F. Ramus, M. Nespor, and J. Mehler, "Correlates of linguistic rhythm in the speech signal. *Cognition*, 73, pp. 265–292, 1999.

[13] V. Dellwo, and P. Wagner, "Relations between language rhythm and speech rate," *Proceedings of the 15th international congress of phonetic sciences*, pp. 471–474, 2003.

[14] V. Dellwo, "Rhythm and speech rate: A variation coefficient for delta C," In P. Karnowski, & I. Szigeti (Eds), *Language and language processing*, pp. 231–241, Frankfurt am Main: Peter Lang, 2006.

[15] E. Grabe, and E. L. Low, "Acoustic correlates of rhythm class," In C. Gussenhoven, & N.Warner (Eds.), *Laboratory phonology*, Berlin: Mouton de Gruyter, vol. 7, pp.515–546, 2002.

[16] M. Ordin, and L. Polyanskaya, "Development of timing patterns in first and second languages," *System*, 42, pp. 244-257, 2014.

[17] C.T. Best, and M.D. Tyler, "Nonnative and second-language speech perception: Commonalities and complementarities, in Second language speech learning: The role of language experience in speech perception and production," In M.J. Munro and O.-S. Bohn (Eds), John Benjamins: Amsterdam. pp.13-34, 2007

[18] K. Aoyama, and S. G. Guion, "Prosody in second language acquisition: Acoustic Analyses of duration and F0 range". In Bohn, O.-S., Munro, M.J. (Eds.), *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, Amsterdam, John Benjamins, pp. 282-297, 2007.

[19] Mochizuki-Sudo, and S. Kiritani, "Production and perception of stress-related durational patterns in Japanese learners of English," *Journal of Phonetics*, vol. 19(2), pp. 231-248, 1991.

[20] K. Tajima, D. Erickson, and K. Nagao, "Production of syllable structure in a second language: Factors affecting vowel epenthesis in Japanese-accented English," Speech Prosody and Timing: Dynamic Aspects of Speech (Indiana University Working Papers in Linguistics, Volume 4), pp. 77-92, 2003.

[21] K. Tajima, D. Erickson and K. Nagao, ''Phonetic analysis of vowel epenthesis in native Japanese speakers' production of English words,'' *Proceedings of 14th General Meeting of the Phonetics Society of Japan*, pp. 195–200, 2000.

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

[23] S. Funatsu and M. Fujimoto, M, "Mechanisms of vowel epenthesis in consonant clusters: an EMA study," *Proceedings of Acoustics 2012*, pp.341-346, 2012.

[24] P. Boersma, and D. Weenink, "Praat: doing phonetics by computer [Computer program]," Version 5.3.51, retrieved 2 June 2013 from http://www.praat.org/.