

Jaw dancing American style: A way to teach English rhythm

Ian Wilson¹, Donna Erickson^{2,3}, Timothy Vance⁴, Jeff Moore⁵

¹University of Aizu, Japan

²Haskins Laboratories, USA

³Kanazawa Medical University, Japan

⁴Komatsu University, Japan

⁵Toyo University INIAD, Japan

wilson@u-aizu.ac.jp, Ericksondonna2000@gmail.com, tjvance825@gmail.com, moore@iniad.org

Abstract

Teaching a second language (L2) involves teaching prominence patterns. Sentence prominence patterns vary from language to language. Traditionally, differences in language prominence patterns have been described as differences in timing: syllable-timing (e.g., French) versus stress-timing (e.g., English). Research by Erickson and others has shown that language prominence patterns vary across different languages and are reflected in the patterns of syllable jaw displacements (e.g., [1], [2], [3], and [4]). Moreover, recent studies ([5], [6], [7]) have shown that first language (L1) jaw displacement patterns tend to be transferred to speakers' L2. In this study, we focus on how 20 L1-Japanese speakers transfer jaw displacement patterns when speaking L2-English. We investigate three questions: (1) Can "teaching" jaw displacement patterns help the L2 learner to change their jaw displacement patterns to those of the new language? (2) Which method is better for teaching: showing jaw tracings or showing syllable magnitude patterns? (3) Can we see the effects of "jaw training" in terms of changes in formant frequencies, specifically, F1 and F2? Results suggest learners can quickly learn to alter their L2 jaw displacements, and that they seem to find jaw tracings more effective than syllable magnitude patterns as visual aid tools.

Index Terms: prominence patterns, jaw articulation, rhythm, L2 English by Japanese

1. Introduction

Teaching L2 learners how to speak fluently involves more than teaching vocabulary, segmental phonetics, grammar, etc. It also involves teaching prominence patterns, which vary from language to language. Traditionally, differences in language prominence patterns have been described as a difference in timing: syllable-timing (e.g., French) vs. stress-timing (e.g., English). This paper is motivated by the simple observation that for all languages, we speak by opening and closing our mouth, while phonating with our vocal folds. The result is a syllable. An utterance is made up of a series of syllables – in articulatory terms, a series of jaw openings (and closings); syllables are organized in phrase units.

Languages differ as to which syllables get the prominence in a phrase. Some languages like French, Chinese, Japanese are edge-strengthening languages, in which the final and optionally initial word receives "phrasal stress" (e.g., [8], [9]). For other languages like English and German, each syllable in the utterance receives a certain amount of prominence, with the largest prominence in the utterance assigned to the nuclear-stressed word/syllable (usually the last content word of an ut-

terance (e.g., [10]). This organization of syllable prominence patterns can be described in terms of metrical trees [11] or metrical grids [12], [13]. In this way of thinking, languages like English are n-ary syllable stressed languages, where each syllable gets a certain level of stress.

Work by Erickson and colleagues has shown that language prominence patterns across various languages are reflected in the patterns of jaw displacement for each syllable spoken, as measured using electromagnetic articulography (EMA). Jaw displacements are measured from a speaker's occlusal plane to the maximum lowering of the jaw for each syllable in an utterance. An acoustic consequence of increased jaw displacement is increased F1 values (for studies on English, see [1]; for Japanese, see [3]; for Spanish, see [14]; for French, see [6] and [4]).

Recent work has also indicated that the jaw displacement patterns of a speaker's L1 tend to be transferred to their L2 (for L1-Japanese speaking L2-English, see [15]; L1-English speaking L2-Japanese, see [5]; L1-French speaking L2-English, [6]; for L1-English speaking L2-French, see [4]). Moreover, for English, $F_2 - F_1$ measurements show that the vowels after a nuclear-stressed syllable become more schwa-like (see [16] and also Fig. 1).

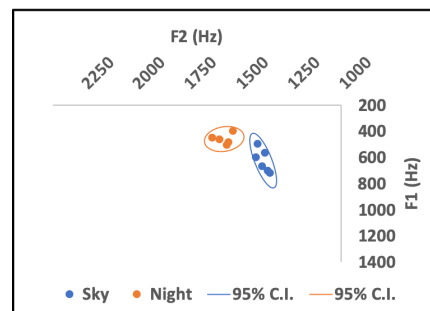


Figure 1: Scatter plot of $F_1 \times F_2$ for one English native speaker's /a/-vowels in the words 'sky' (dark blue dots) and 'night' (light orange dots) in 6 repetitions of the sentence "I saw five bright highlights in the sky tonight". Ovals show 95% confidence intervals.

In our current study, we focus on how intermediate L1-Japanese speakers negatively transfer their jaw displacement patterns when speaking L2-English. We study Japanese speakers here because of the differences between English and Japanese mentioned above, and because we have access to a large number of speakers. We investigate three questions:

1. Can “teaching” jaw displacement patterns help L2 learners change their jaw displacement patterns to those of the new language – specifically, can they learn to open their jaw on the nuclear-stressed item in an English sentence?
2. Which method is better for teaching: showing jaw tracings or showing syllable magnitude patterns?
3. Can we see the effects of “jaw training” in terms of changes in formant frequencies – specifically, F1 and F2?

2. Method

Twenty English-as-a-Foreign-Language (EFL) students at a university in Japan participated in the test. Each student was recorded individually in a quiet room. In order to make the learning exercise more fun as well as to help the learners focus on jaw patterns, the learners read the following statement: “Let’s practice ‘Jaw Dancing’! ‘Jaw Dancing’ is how you move your mouth to make speech sounds; it is how your mouth ‘dances’ as you speak. The mouth moves/dances differently for Japanese and for English. In this exercise, we hope to teach you how to do ‘American English jaw dancing’! If you do ‘American English Jaw Dancing’, you will sound more like a real American English speaker! Come on, let’s dance!!”

Very few participants were able to understand the English instructions – more than half did not know the English word *jaw*, so the bilingual experimenter provided explanation in Japanese about what *jaw dancing* means. He also demonstrated jaw opening/closing with his hands, emphasizing that only the lower jaw moves.

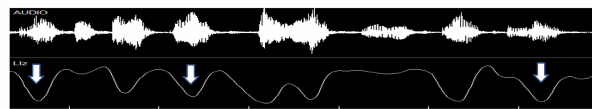
The students were then asked to read the English sentence, “I saw five bright highlights in the sky tonight” and record their voice using Praat software [17]. They were allowed to re-record if they made a word error or were not happy with their recording. Very few needed to re-record.

In order for us to assess which visual type of information was better for teaching jaw displacement patterns, students were divided into two groups. Group 1 was shown samples of continuous jaw tracings, made from electromagnetic articulograph (EMA) recordings, and Group 2 was shown syllable magnitude bar graphs, where the height of each discrete bar represented the measured amount of jaw opening for that syllable. The lowest jaw tracing/the largest syllable magnitude represented the point where the jaw was most open for the /a/ part of the /ai/ diphthongs in this utterance.

The instructions for Group 1 and Group 2 are shown in Fig. 2 and Fig. 3, respectively. The experimenter drew each participant’s attention to the fact that the top image was for a non-native speaker’s production and the bottom one was for a native speaker’s production. For the Group 1 graphs, he used his hands to show that the lower the curve, the farther down the lower jaw had moved. For the Group 2 graphs, he used his hands to show that the taller the bar, the wider the space between the upper and lower jaw.

In the instructions, the images of the native English speaker’s utterance have the nuclear stress on *sky*, as this sentence is being read naturally without contrastive stress; hence the largest jaw displacement is on *sky*, and consequently, the F1 and F2 formants (for the /a/ part of the diphthong) are low and back, while those for the following /a/ in *night* are centered and more schwa-like, as was shown in Fig. 1. After they read the instructions, learners were told to imitate the lower graph and to record again. Some practiced once or twice, some not at all,

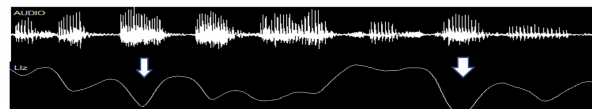
Step 2. Now, look at this picture of jaw dancing NOT AMERICAN style for this sentence.



I saw five bright highlights in the sky tonight

In the graph above, you see the sound wave, and then the “jaw wave”, which shows how much you opened your mouth for each word. Maybe this is what you did. You probably made big jaw opening at the beginning (“I”) and end of the sentence (“night”). You also probably made big jaw opening for “bright”.

Step 3. Now, look at this picture of jaw dancing AMERICAN style for this sentence.



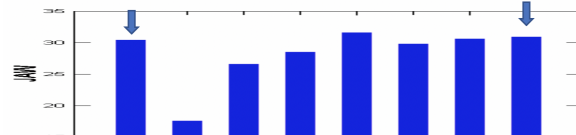
I saw five bright highlights in the sky tonight

Step 4. Try to make your first big jaw opening on five, and your biggest jaw opening on sky.

Step 5. Now record your voice again, using your new American Jaw Dancing rhythm!

Figure 2: Group 1 “Jaw Dancing” instructions

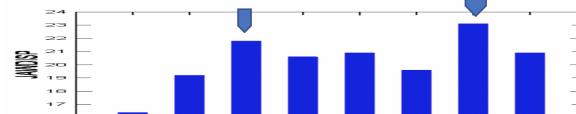
Step 2. Now, look at this picture of jaw dancing NOT AMERICAN style for this sentence.



I saw five bright highlights in the sky tonight

Each bar in the graph shows how big you opened your jaw for each word. Maybe this is what you did. You probably made big jaw opening at the beginning (“I”) and end of the sentence (“night”). You also probably made big jaw opening for “bright”.

Step 3. Now, look at this picture of jaw dancing AMERICAN style for this sentence.



I saw five bright highlights in the sky tonight

Step 4. Try to make your first big jaw opening on five, and your biggest jaw opening on sky.

Step 5. Now record your voice again, using your new American Jaw Dancing rhythm!

Figure 3: Group 2 “Jaw Dancing” instructions

before doing the second recording.

In order to assess an effect from the “dancing instructions” on the production of nuclear stress, F1 and F2 measurements were made of the two words, *sky* and *night* “before jaw dancing training” and “after jaw dancing training”.

3. Results

The results for both groups combined together are shown in scatter plots of $F1 \times F2$ for the final content words (*sky* and *night*) in the final phrase of the utterance. The left side of Fig. 4 indicates that before training, the $F1 \times F2$ values were not differentiated for the two words; the right side shows more differentiation for the formants after training.

Table 1 shows the mean F1, F2, and F2-F1 values of the two final content words for the combined groups before and after jaw dancing training. Table 2 shows the mean F1, F2, and F2-F1 values of the two final content words for each group (separately) before and after jaw dancing training. In both tables,

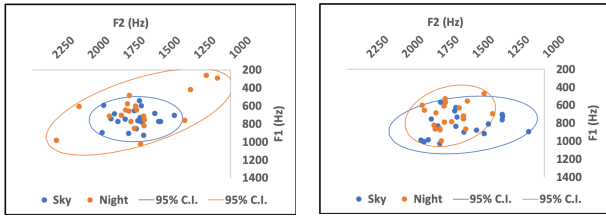
(a) *Before training*(b) *After all training*

Figure 4: *Before (a) and after (b) jaw dancing training – All data (Groups 1 & 2 together) scatter plots of $F1 \times F2$ for final content words ‘sky’ (dark blue circles) & ‘night’ (light orange circles) in the final phrase of the utterance*

the numbers in italics indicate increased values after training. Notice the general tendency for the nuclear stressed word, *sky*, to have higher F1, lower F2 and thus lower F2-F1 after training compared to before. Higher F1 indicates more jaw opening, while lower F2 can indicate lower tongue position. A smaller F2-F1 is to be expected for a stressed low vowel in English (e.g., [18] where it is also discussed that even stressed *high* vowels have more jaw opening than unstressed ones).

Table 1: *Mean vowel formants for final content words (both groups combined) before and after “jaw dancing” training*

Both groups combined				
word	before/after training	F1	F2	F2-F1
sky	before	746	1670	924
	after	820	1577	757
night	before	652	1643	977
	after	717	1669	951

Table 2: *Mean vowel formants for final content words (each group separately) before and after “jaw dancing” training*

Group 1 (jaw tracings)				
word	before/after training	F1	F2	F2-F1
sky	before	756	1714	958
	after	831	1604	774
night	before	652	1706	1054
	after	709	1731	1022
Group 2 (syll. mag. patterns)				
sky	before	737	1626	890
	after	808	1549	741
night	before	680	1580	900
	after	726	1606	880

We see a similar pattern of higher F1 and lower F2-F1 for *night*. However, F2 is higher after training, yet F2-F1 is lower. Future studies using electromagnetic articulographs (EMA) are needed to better understand this. A two-sample t-test with Bonferroni-adjusted pooled variance p-values shows a

significant difference ($p = 0.011$) in the values of F2-F1 for the nuclear-stressed word *sky*. These changes in formant patterns suggest that Japanese learners of English are indeed paying attention to “jaw dancing”, and are opening their mouths more for both words, but especially for the nuclear stressed word.

As for which jaw dancing training was a more effective visual tool for the Japanese learners of English, the results suggest that the learners were perhaps better able to do English-style jaw movements when they were trained with the jaw tracings, rather than the syllable magnitude bar graphs. Fig. 5 shows formant scatter plots of Group 1 before (left) and after (right) training, respectively. Fig. 6 shows formant scatter plots of Group 2 before (left) and after (right) training, respectively.

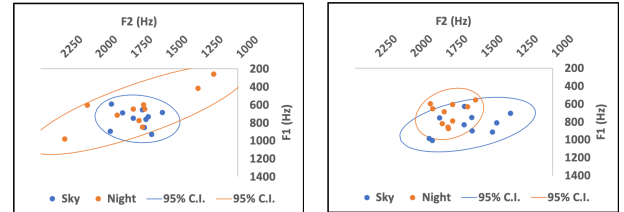
(a) *Before training with jaw tracings*(b) *After training with jaw tracings*

Figure 5: *Before (a) and after (b) training with jaw tracings – Group 1 scatter plots of $F1 \times F2$ for content words ‘sky’ (blue circles) & ‘night’ (orange circles) in the final phrase of the utterance*

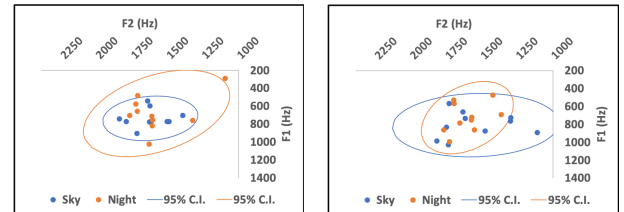
(a) *Before training with syllable magnitude bar graphs*(b) *After training with syllable magnitude bar graphs*

Figure 6: *Before (a) and after (b) training with syllable magnitude bar graphs – Group 2 scatter plots of $F1 \times F2$ for content words ‘sky’ (blue circles) & ‘night’ (orange circles) in the final phrase of the utterance*

For both groups, formants of the two words are not clearly separated before training; but after training, there is clearer separation for Group 1 (trained with jaw tracings) than Group 2 (trained with syllable magnitude bar graphs). A two-sample t-test with Bonferroni-adjusted pooled variance p-values shows a significant difference ($p = 0.032$) in the values of $F2 - F1$ for the nuclear-stressed word *sky*, but only for Group 1, not for Group 2. This suggests that jaw tracings may provide better guidance to *jaw dancing* than syllable magnitude patterns.

4. Discussion

The results of this 20-participant pilot study with *jaw dancing* suggest that Japanese learners of English can learn to change their jaw displacement patterns to those of the new language, at least to some extent. The results suggest that they learn to open their jaw more on the nuclear-stressed item in an English

utterance after jaw dancing training, as evidenced by increased F1 and decreased F2 on the stressed word.

As far as which jaw dancing training is most effective for Japanese learners of English, the results from this study suggest that jaw tracings may be better than syllable magnitude patterns. Japanese learners may have interpreted the syllable magnitude bar graphs to indicate they needed to increase intensity on syllables with larger magnitudes; however, since intensity analysis was not done, because microphone distance was not controlled during the recordings, this is a topic for future investigation.

The results of this paper also suggest that approaching prosody through kinematic training may be as effective or even more than the more traditional method of instructing Japanese learners of English about nuclear stress in English. Future work needs to be done to test the relative effectiveness of kinematic training versus cognitive training by giving linguistic instructions, or versus training with intonational patterns.

Jaw dancing applications can be extended to other language learning contexts, e.g., English speakers learning Japanese. A recent study based on acoustic and articulatory (EMA) data reported that English speakers of Japanese tend to transfer their L1 lexical stress patterns to their spoken Japanese, especially when the Japanese word is an English cognate [5]. For instance, in the sentence *aka pajama da* (“they are red pajamas”), even English speakers who are relatively advanced in L2 Japanese tend to reduce the second syllable of *pajama*. Preliminary work with jaw dancing training [16] using syllable magnitude patterns suggests that jaw dancing training indeed helps learners of Japanese to avoid making stressed syllables followed by reduced schwa-like vowels. Fig. 7 shows results for one English speaker of Japanese before and after jaw dancing training. Notice that before training, the middle syllable *ja* had distinctly lower F1 and F2 values than the other two syllables, which were more centered in the vowel space. After jaw dancing training, the $F1 \times F2$ values of all three syllables clustered together. Thus, it seems that this American English speaker reduced the overall amount of jaw movement after training.

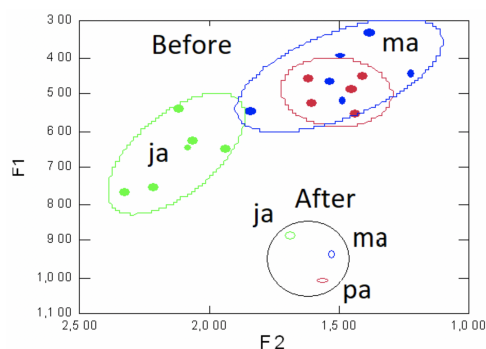


Figure 7: Scatter plots of $F1 \times F2$ for three syllables in ‘pajama’. Filled circles (top section) are before jaw dancing training and empty circles (lower section) are after training

English speakers, especially teachers of English, perhaps tend to open their jaw more in order to speak more clearly. So, jaw training may help these teachers improve their spoken Japanese. Another point is that maybe syllable magnitude training is better for English speakers of Japanese, while jaw tracing training is better for Japanese speakers of English. The rationale is that Japanese may not necessarily have a concept of what a syllable is, but most English speakers do. Patterns of syl-

lable magnitude to English speakers may indicate information to them about how small a syllable is, which will lead them to reduce the amount they open their jaw.

Jaw dancing training may also help other language learners, e.g., French learners of English and vice versa. These language learners face the same problem in that English learners of French need to open their jaw less, except for the end of an utterance, and speakers of French learning English, need to open their jaw more on stressed items, and less on the following reduced item.

Regarding the mechanics of jaw opening, it is probably difficult to open the jaw the exact same amount for a series of syllables. So, in English it makes sense that after a large jaw opening, the next syllable will have a smaller jaw opening. Thus, for training English speakers to speak a non-stress language like Japanese, they need to practice *not* opening their jaw more, especially for cognates with lexical stress. By *not* opening the jaw more, this will automatically result in not having a difference between full vowels and reduced schwa-like vowels. Conversely for Japanese speakers of English, if they can follow jaw tracings, they will see that the amount of jaw opening increases for certain syllables, and then if they open their jaw a lot for the stressed syllable, the mechanical nature of the jaw will lead to a somewhat reduced jaw opening (to produce a schwa-like vowel) for the following syllable. In this way, kinematic prosody training may automatically lead to better prosody, without explicit instructions about where to put stress, or how to acoustically implement stress.

5. Conclusions

This paper proposes a kinematic method of teaching prosody to L2 learners, referred to here as *jaw dancing*. The results suggest that Japanese learners of English change their jaw displacement patterns to a more English-like pattern of increased jaw displacement for nuclear stress, as evidenced by increased F1 and decreased F2. Interestingly, Japanese learners may perform better after being trained with jaw tracings than with syllable magnitude bar graphs – more complex continuous data winning out over simpler discrete data.

Future work needs to include more speakers and sentences. We will conduct a parallel experiment with English learners of Japanese: do they perform Japanese prosody better with jaw displacement training? Future work also will involve EMA and/or video studies, to directly measure the jaw before/after training, and to determine the extent that the concept “lower jaw = greater prominence” generalizes to all other vowels. Future work may also elucidate more differences between high- and low-proficiency L2 speakers’ jaw movements [19].

Finally, it is obvious that most language instructors do not have access to jaw training materials (e.g., jaw traces or EMA equipment to get them). However, since jaw tracings/syllable magnitude patterns reflect the stress patterns of an utterance, it would be possible to draw pseudo jaw tracings or syllable magnitudes, based on a phonologically derived metrical grid, and use these for jaw dancing training. One could also use video camera tracking of the jaw to create realistic jaw tracings.

6. Acknowledgements

This work was supported in part by two Grants-in-Aid for Scientific Research (C) from the Japan Society for the Promotion of Science: No. 22520412 and No. 25370444. We greatly appreciate the helpful suggestions we got from four reviewers.

7. References

- [1] D. Erickson, A. Suemitsu, Y. Shibuya, and M. Tiede, "Metrical structure and production of English rhythm," *Phonetica*, vol. 69, no. 3, pp. 180–190, 2012.
- [2] D. Erickson and S. Kawahara, "Articulatory correlates of metrical structure: Studying jaw displacement patterns," *Linguistics Vanguard*, vol. 2, no. 1, 2016.
- [3] S. Kawahara, D. Erickson, J. Moore, Y. Shibuya, and A. Suemitsu, "Jaw displacement and metrical structure in Japanese: The effect of pitch accent, foot structure, and phrasal stress," *Journal of the Phonetic Society of Japan*, vol. 18, no. 2, pp. 77–87, 2014.
- [4] C. L. Smith, D. Erickson, and C. Savariaux, "Articulatory and acoustic correlates of prominence in French: Comparing L1 and L2 speakers," *Journal of Phonetics*, vol. 77, p. 100938, 2019.
- [5] I. Wilson, D. Erickson, S. Kawahara, and T. Monou, "Acquiring jaw movement patterns in a second language: Some lexical factors," in *Proceedings of the 33rd General Meeting of the Phonetic Society of Japan*, K. Maekawa, Ed., Seisen University, Japan, 2019, pp. 32–37.
- [6] T. Huang and D. Erickson, "Articulation of English "prominence" by L1 (English) and L2 (French) speakers," in *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia*, S. Calhoun, P. Escudero, M. Tabain, and P. Warren, Eds., 2019, pp. 2480–2484.
- [7] D. Erickson, T. Huang, C. Menezes, and S. Kawahara, "Using jaw movement patterns to visualize prosody [One of three official posters entered in the Prosody Visualisation Challenge 2 (PVC2), International Congress of Phonetic Sciences, Melbourne, Australia]," 2019.
- [8] S.-A. Jun, "Prosodic typology," in *Prosodic typology: The phonology of intonation and phrasing*, S.-A. Jun, Ed. Oxford, UK: Oxford University Press, 2005, pp. 430–458.
- [9] —, "Prosodic typology: by prominence type, word prosody, and macro-rhythm," in *Prosodic typology II: The phonology of intonation and phrasing*, S.-A. Jun, Ed. Oxford, UK: Oxford University Press, 2014, pp. 520–539.
- [10] J. Cole, J. I. Hualde, C. L. Smith, C. Eager, T. Mahrt, and R. N. de Souza, "Sound, structure and meaning: The bases of prominence ratings in English, French and Spanish," *Journal of Phonetics*, vol. 75, pp. 113–147, 2019.
- [11] M. Liberman and A. Prince, "On stress and linguistic rhythm," *Linguistic Inquiry*, vol. 8, pp. 249–336, 1977.
- [12] B. Hayes, *Metrical stress theory: Principles and case studies*. University of Chicago Press, 1995.
- [13] E. O. Selkirk, "The syllable," in *The structure of phonological representations, Part II*, H. van der Hulst and N. Smith, Eds. Foris Publications, 1982, pp. 337–383.
- [14] D. Erickson, J. Villegas, I. Wilson, Y. Iguro, J. Moore, and D. Erker, "Some acoustic and articulatory correlates of phrasal stress in Spanish," in *Speech Prosody 2016*, 2016, pp. 450–454. [Online]. Available: <http://dx.doi.org/10.21437/SpeechProsody.2016-92>
- [15] D. Erickson, S. Kawahara, Y. Shibuya, A. Suemitsu, and M. Tiede, "Comparison of jaw displacement patterns of Japanese and American speakers of English: A preliminary report," *Journal of the Phonetic Society of Japan*, vol. 18, no. 2, pp. 88–94, 2014.
- [16] D. Erickson, "Acquisition of prominence patterns in a second language [Presentation at Sophia University, Tokyo on 15 Oct 2019] doi=10.13140/rg.2.2.17379.71209," 2019.
- [17] P. Boersma and D. Weenink, "Praat: doing phonetics by computer [computer program] (version 6.1.08)," 2019. [Online]. Available: <http://www.praat.org/>
- [18] D. Erickson, "Articulation of extreme formant patterns for emphasized vowels," *Phonetica*, vol. 59, pp. 134–149, 2002.
- [19] I. Wilson, D. Erickson, and N. Horiguchi, "Articulating rhythm in L1 and L2 English: Focus on jaw and F0," in *Proceedings of the 2012 autumn meeting of the Acoustical Society of Japan (ASJ)*. Acoustical Society of Japan, 2012, pp. 319–322.