ISAPh 2018 International Symposium on Applied Phonetics
19-21 September 2018, Aizuwakamatsu, Japan

Initial Cluster Timing in Japanese English

Jeff Moore

Sophia University, Japan
moore@iniad.org

Abstract

Previous work on consonant cluster timing has found evidence of vowel insertion when second language speakers attempt foreign cluster pronunciations. It is also theorized that sonority should have an effect on articulation. An experiment is presented to examine whether these ideas hold true for Japanese English speakers. Results indicate that the sampled Japanese English speakers time their clusters very closely, without evidence of epenthesis. There seem to be minor effects of voicing and sonority, though not the effects that major theories on sonority would predict.

Keywords: clusters, articulation, L2, Japanese, English

1. Introduction

Japanese is traditionally seen as a 'mora-timed' language [1]–[4]. Moras, or morae, are small units of speech, smaller than a syllable, but larger than a segment. In Japanese, a mora may consist of either a vowel (e.g. あ/a), a consonant-vowel pair (e.g. か/ka), a consonant-glide-vowel set (e.g. きゃ/kja), or a nasal (ん/ɴ). In many instances, the rhythm of Japanese is constructed from patterns of moras. When Japanese speakers' judgments on timing are assessed by clapping or counting exercises, their perceptions very often map to moras.

The Japanese language permits only a very limited set of syllable structures. Every Japanese syllable must have at least one vowel (V) as a syllable nucleus. Optionally, a syllable may also contain one initial consonant (C), possibly followed by a glide (G) in onset position. The final position, or coda, may be an additional vowel, a nasal, or a geminate—a duplicate of the following consonant.

Vowel devoicing provides one additional wrinkle on this otherwise fairly simple syllable structure. When one of the two high vowels, /i/ and /u/, is between two voiceless consonants, such as /s/ or /k/, that vowel is 'devoiced', produced like a whispered version of itself, e.g. [1], [3]. However, recent evidence suggests that the devoiced vowel may be 'targetless', meaning that the speaker does not move the tongue in the direction of a vowel target [5].

Under moraic timing, the devoiced mora is considered equivalent to any other mora, but rhythmic effects are observed in syllable timing. In sung Japanese, for instance, and especially songs translated from English, those devoiced moras can be combined into the following mora to form a syllable with a complex onset [6]. Speakers may move the tongue in the direction of the voiceless vowel, or they may simply insert a small gap without any articulatory target [5]. These phenomena cast some doubt on how truly 'moraic' the timing of Japanese is, and whether Japanese permits some form of complex onset by way of these devoiced vowels.

1.1. Cross-linguistic patterns in syllable structure

Under the Minimal Sonority Distance framework [8], the difference in sonority between two adjacent segments matters. A language might allow a nasal next to an approximant, but disallow a stop-fricative cluster. These classifications begin to break down when the analysis gets detailed, however. English permits a fricative-stop cluster like /st/ in initial position, but only with /s/, and it allows very different clusters in onset and coda positions. Japanese allows a nasal-glide onset, /nj/, but not a fricative-liquid onset or stop-nasal onset, although the sonority distance ought to be the same.

Minimal Sonority Distance describes some observed phenomena clearly, but as the above examples demonstrate, there are some serious gaps. It appears that typological markedness predicts phonotactic constraints better than Minimal Sonority Distance, particularly with respect to interlanguage development [9]–[11]. In fact, some evidence indicates an opposite pattern, that L2 learners are more successful producing onset clusters with greater sonority distance [12].

1.2. L2 Cluster Articulation

Lisa Davidson and Maureen Stone [13] asked English speakers to pronounce Polish consonant

10.21437/ISAPh.2018-18

clusters that aren't present in English, and measured the gaps between the consonants. From their work, we get three basic patterns of cluster timing. In a close transition, there is no gap at all between the hold phase of the first consonant and the hold phase of the second. A 'vocoid' is produced when the first and second consonant are close enough that the release of the first overlaps with the approach of the second, and we may hear a small, voiced period between the two, a brief, targetless, coarticulated vowel. A full on epenthetic vowel occurs when consonants do not overlap at all, and the space between is filled with a voiced period that may look like a normal vowel.

1.3. Research Questions

1. Do Japanese English speakers use full vowel epenthesis or vocoids?

Are the consonants in Japanese English consonant clusters tightly timed, with hold phases touching, loosely timed with the release and approach phases overlapping, or resyllabified with an epenthetic vowel?

2. Do JE speakers produce tighter stop-liquid clusters, or sibilant-stop clusters?

The Sonority Sequencing Principle would suggest that stop-liquid clusters are less marked than sibilant-stop clusters. Under that framework, sibilant-stop clusters should be more difficult.

2. Experiment

An experiment was conducted using Electromagnetic Articulography (EMA) to measure the speech of five participants.

2.1. Participants

Four native speakers of Tokyo Japanese were asked to sit for recordings. Three participants were graduate and undergraduate students, aged 19 to 23, while one participant (JFA) was an instructor at a major university in Tokyo. Three female speakers and one male speaker were recruited, with varying levels of English proficiency. One native speaker (this author) was recorded for comparison Procedures were explained to participants in their respective native languages.

Table 1: Participants

AMN	American Male, Native
JFA	Japanese Female, Advanced
JFUI	Japanese Female, Upper Intermediate
JFLI	Japanese Female, Lower Intermediate
JML	Japanese Male, Lower

Proficiency ratings are based on judgments by 245 native English listeners listening to audio clips in an online survey. They rated the speakers for comprehensibility and accentedness, and wrote the words they thought they heard to assess intelligibility.

2.2. Materials

Participants were asked to read aloud short phrases from a monitor. Target words were presented in a carrier phrase, "Okay, <target>", and displayed using ePrime software. Six tokens with stop-liquid clusters were elicited: 'braid', 'blade', 'prayed', 'played', 'clay', and 'crane'. Four tokens with fricative-stop liquid clusters were also elicited: 'ski', 'sky', 'speak', 'spy'.

2.3. Data Collection

Their articulations were recorded using an NDI Wave Electromagnetic Articulograph (EMA), recording at 100 Hz. Five transducers were affixed to the tongue. Three were affixed along the sagittal midline. The most anterior sagittal transducer was one centimeter from the tip of the tongue. The most posterior was affixed as far back as was comfortable for the participant, ranging from 5 to 6 centimeters from the tongue tip at furthest extension. A mid-tongue transducer was affixed at the midpoint between these two. Two lateral transducers were affixed in line with the mid-tongue transducer, along a coronal plane perpendicular to the sagittal midline, placed one centimeter from the edge of the tongue.

2.4. Data Analysis

The gaps between articulations were measured in two different ways. Hold Gap measures the time in milliseconds between the beginning of the initial plosive's release phase and the beginning of the liquid's hold phase. Iceberg Gap measures the time in milliseconds between the moment of highest velocity in the release phase of the plosive and the moment of highest velocity in the approach phase of the liquid. Iceberg Gaps are named for the concept of iceberg points in Osamu Fujimura's C/D Model, which posits that these points are most stable across multiple articulations—frozen, and making immobile triangles for each articulation.

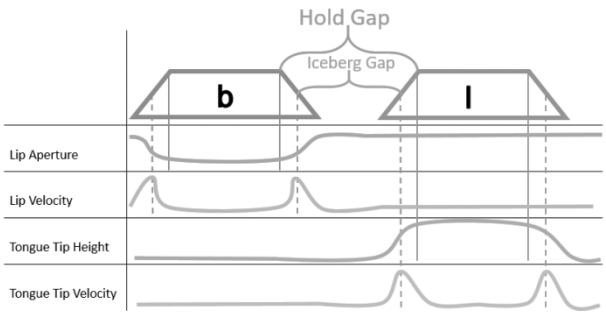

Figure 1: Hold Gaps and Iceberg Gaps

Articulatory data was head corrected using Donald Derrick's 2012 head correction scripts and analyzed using the MView package for MATLAB (Tiede, 2005) and R (R Core Team, 2017).

3. Results

3.1. Results by Word

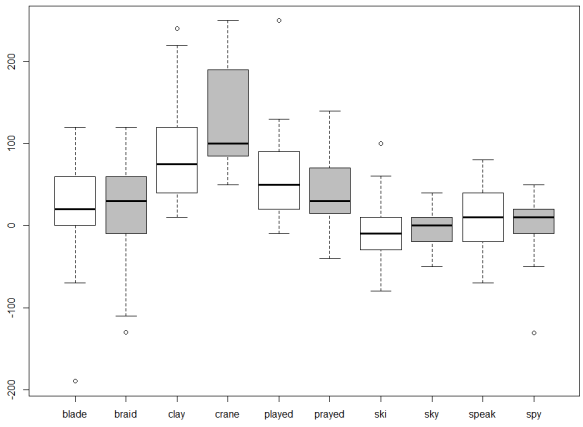

Figure 2: Hold Gaps by Word, Japanese Speakers

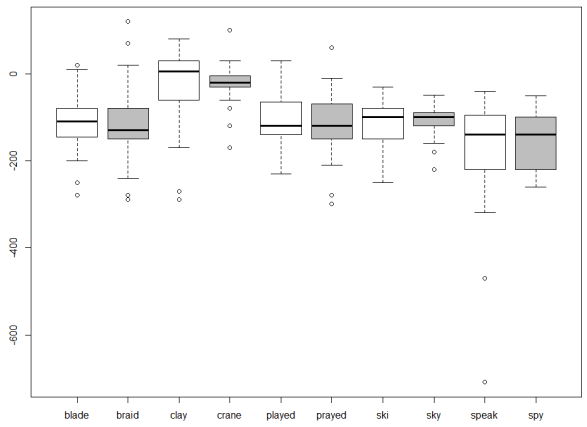

Figure 3: Iceberg Gaps by Word, Japanese Speakers

Here we see the hold gaps by word, combining all 4 Japanese speakers. Sibilants are on the average more closely timed than liquids, and quite often they overlap. Further, "ski" is the most overlapped of the bunch.

Iceberg gaps look a bit different. Nearly every token overlaps, indicating that these four Japanese English speakers are using vocoids, rather then full epenthetic vowels. Only clay has a significant number of tokens with a gap. Interestingly, ski does not overlap more than other tokens in terms of iceberg points.

Clay and crane have larger gaps than the rest. They are both homorganic and dorsal, so this result is expected. The tongue needs time to move from the dorsal closure to the apical gesture in clay, or to retract further for a bunched /r/ in crane. This is true of the native speaker's data as well.

3.2. Results by Speaker

3.2.1. JFA

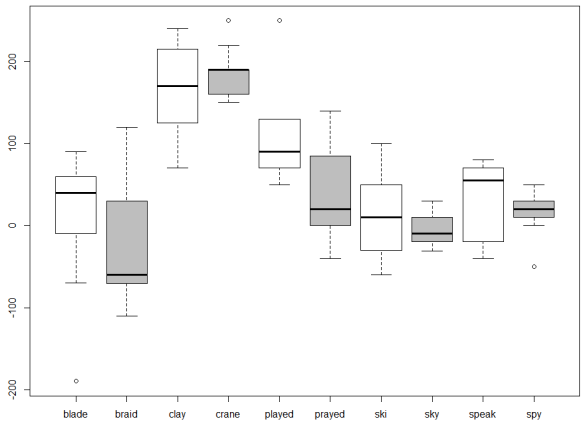

Figure 4: Hold Gaps, JFA

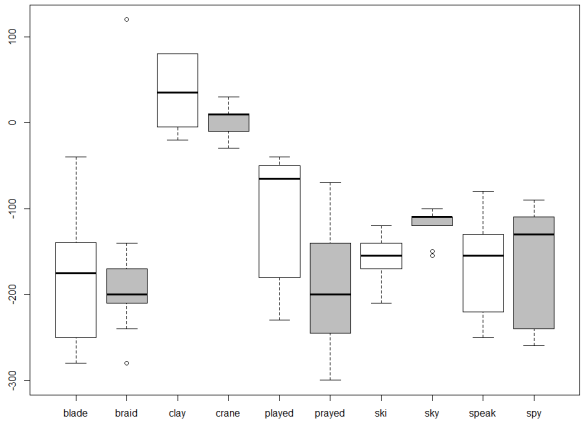

Figure 5: Iceberg Gaps, JFA

Surprisingly, JFA had fairly large hold phase gaps in her productions, despite being the most experienced speaker. The only word that consistently showed an

overlap was braid—which had some interesting properties for many speakers. There is a clear effect of voicing here. 'Blade' and 'braid' have much smaller gaps than 'played' and 'prayed.

JFA had quite a slow speaking rate, which may account for the disparity between her hold and iceberg gaps. She rarely overlaps the hold phases of her clusters, but most of her productions overlap in the release and approach phases. She is not categorically better at sibilant clusters than liquid clusters, and she also does not show any evidence of an advantage for ski

. It could be that she has been using English long enough that her productions are governed less by transfer from Japanese, and more by physical constraints on articulation.

3.2.2. JFUI

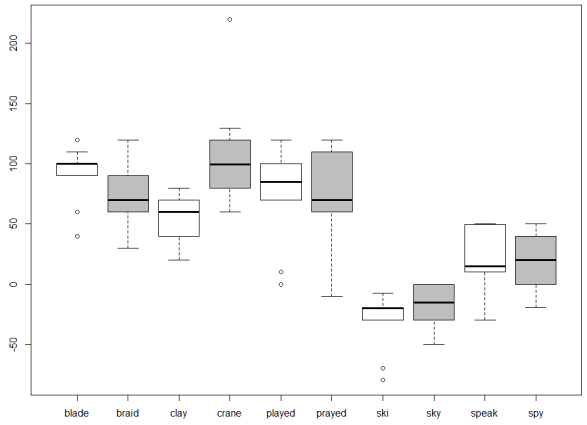

Figure 6: Hold Gaps, JFUI

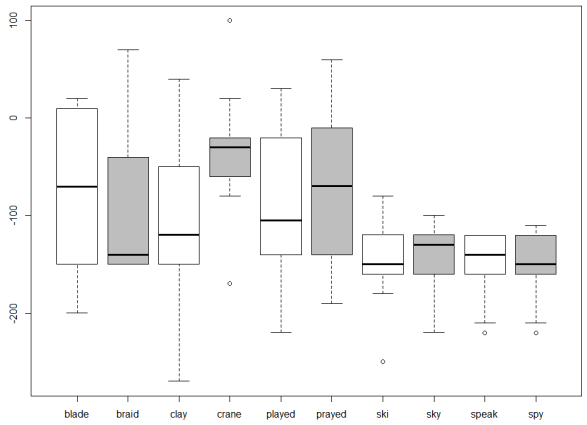

Figure 7: Iceberg Gaps, JFUI

JFUI is more consistent in her hold gaps, as they all average somewhere between 60 and 100 milliseconds. 'Prayed' shows a huge amount of variation, however, and on some samples even overlaps, which she never does for 'braid'.

Iceberg gaps tell a different story. 'Clay' and 'crane' are the only samples without any overlap at all, and 'blade' clearly overlaps much more than 'played' does. Just as in JFA's samples, 'prayed' has wide variation, while 'braid' is more consistent. JFUI shows a big advantage in hold gaps for ski, and is categorically better at sibilant clusters than liquid clusters.

3.2.3. JFLI

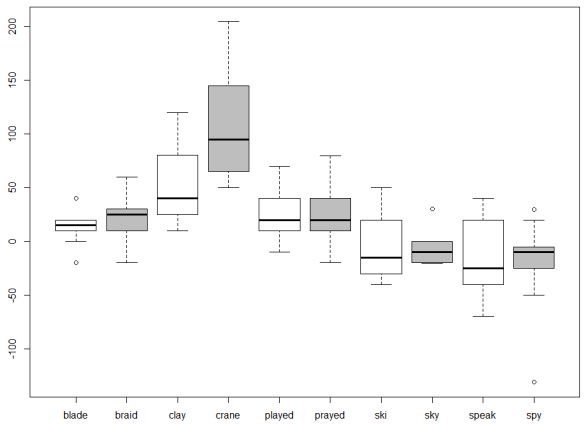

Figure 8: Hold Gaps, JFLI

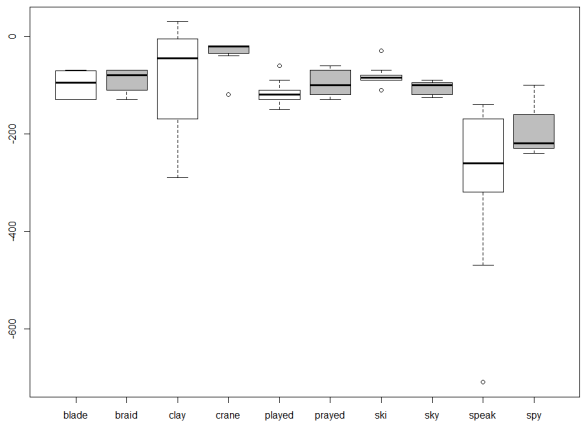

Figure 9: Iceberg Gaps, JFLI

JFLI's data is fairly similar to JFUI's: better at sibilants, and nearly every token overlaps in iceberg points. She is not especially good at 'ski', however.

3.2.4. JML

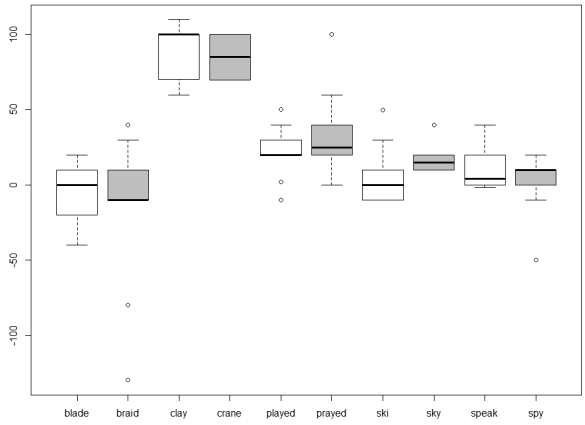

Figure 10: Hold Gaps, JML

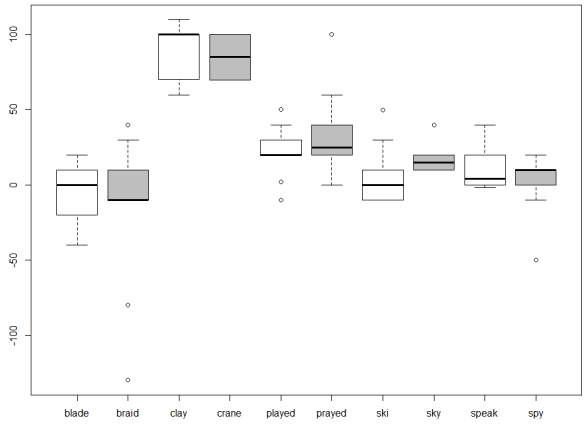

Figure 11: Iceberg Gaps, JML

JML has fairly small gaps, smaller in many cases than JFA and JFUI, the more advanced speakers. This may be in part due to a faster articulation rate, although there was only a mild correlation (0.3678, p = 8.764e-10) between articulation rate and gap times.

Once again 'clay' and 'crane' are the outliers, and once again we can see a mild effect of voicing on the hold gaps in 'played' and 'prayed'. Ski is a bit closer timed than other words, but he is relatively skilled at producing a close transition in 'blade' and 'braid', so there is less of a clear advantage in pronouncing sibilant clusters. He also seems to show an effect of voicing, producing 'played' and 'prayed' with larger gaps than in 'blade' and 'braid'.

3.2.5. AMN

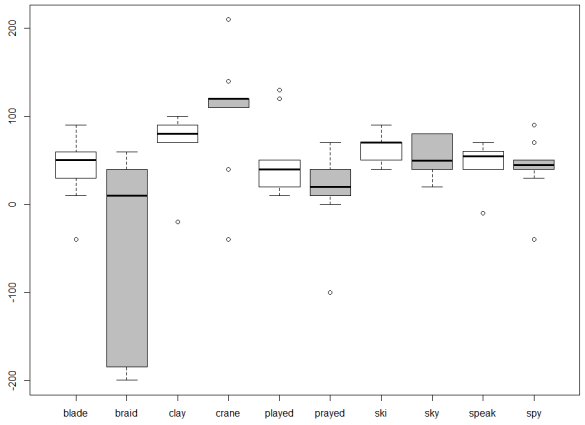

Figure 12: Hold Gaps, AMN

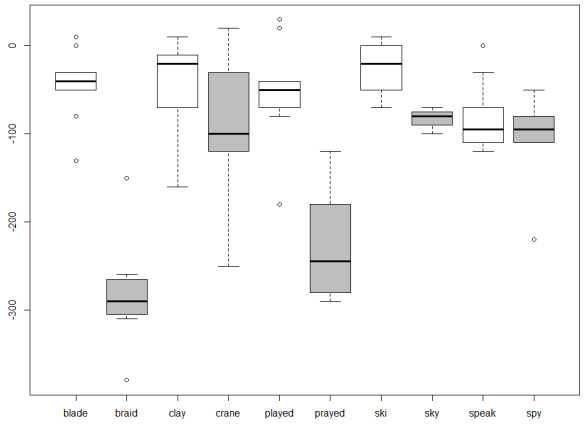

Figure 13: Iceberg Gaps, AMN

The native speaker, AMN, may be the most surprising of the speakers presented here. In terms of hold gaps, many of his productions show larger gaps than in JML's productions. Iceberg gaps tell the real story here, though, as AMN has almost no samples that do not overlap the release and approach phases to some extent. AMN exclusively uses a bunched articulation for /r/, so this may explain the large overlap with bilabial gestures, as well as the relatively broad overlap time in 'crane'. AMN has significant hold gaps across the board, although the articulations overlap in the releases and approaches. Perhaps the idealized 'native' pronunciation is not tight articulation between two consonants after all.

4. Discussion

Timing data suggests complications to the standard description of excrescent vowels as a means of repair in Japanese English clusters. All three female Japanese speakers had gaps between the release of the initial stop and the approach of the liquid. The

lower-level, male speaker, however, had an extremely short gap between the stop and liquid, just over half the duration of the native speaker's. His speech rate was much higher than that of any other speaker, and although timing data was normalized for duration, his fast production led to considerably greater overlap than that of any other speaker.

Clearly there is a large effect of the word on these timings, as words including dorsal gestures had much larger gaps on average and longer durations. Curiously, voicing seems to have an effect on the Japanese speakers, even though there is no physical reason that they should draw out a voiceless onset cluster or be more likely to insert a vocoid here. The four Japanese speakers time their sibilant-stop clusters a bit more closely than stop-liquid clusters, but there was significant variation between speakers. A larger sample of speakers will be needed to make strong claims about the relative timing of these two types of clusters.

In most cases, the approach of the liquid completely overlapped the release of the stop. This is considered a target-like timing, e.g.(Shaw & Davidson, 2011), as there is no gap between the release of the first consonant and the approach phase of the second consonant. Further, speakers showed no evidence of a backward movement toward an /u/-like production. This suggests that if there is any sort of inserted voiced period present in these speakers' productions, it is a targetless vocoid of the sort described by Shaw and Kawahara (Shaw & Kawahara, 2018). Thus the present study indicates that the four Japanese English speakers studied here produced target-like timings, and if there is mistiming present at all, it is in the form of a targetless vocoid.

5. Conclusions

1. Do Japanese English speakers use full vowel epenthesis or vocoids?

The speakers presented here show very few instances in which their timing could be described as an epenthesis. The vast majority of productions were either tightly timed or used short vocoids.

2. Do JE speakers produce tighter stop-liquid clusters, or sibilant-stop clusters?

Sibilant-stop clusters were on the average more closely timed, but given the amount of variation between the four speakers, it would be unwise to make strong claims about these clusters for all Japanese English speakers.

6. Acknowledgements

I would like to thank my advisor, Takayuki Arai, for his guidance in shaping this paper. I would also like to thank Shigeto Kawahara and Jason Shaw for providing the funding, space, education and opportunity necessary for this experiment to take place.

7. References

[1] M. E. Beckman, "Segmental duration and the 'mora' in Japanese," *Phonetica*, vol. 39, pp. 113–135, 1982.

[2] J. Ito and A. Mester, "Japanese phonology," in *The handbook of phonological theory*, J. Goldsmith, Ed. Oxford: Blackwell, 1995, pp. 817–838.

[3] P. Keating and M. Hoffman, "Vowel variation in Japanese," *Phonetica*, vol. 41, pp. 191–207, 1984.

[4] T. J. Vance, *The Sounds of Japanese*. Cambridge: Cambridge University Press, 2008.

[5] J. A. Shaw and S. Kawahara, "The lingual articulation of devoiced /u/ in Tokyo Japanese," *Journal of Phonetics*, vol. 66, pp. 100–119, Jan. 2018.

[6] R. L. Starr and S. S. Shih, "The syllable as a prosodic unit in Japanese lexical strata: Evidence from text-setting," *Ms. National University of Singapore and University of California, Merced.(Also talk presented at LabPhon 14)*, 2014.

[7] E. Selkirk, "On the major class features and syllable theory," in *Language Sound Structure: Studies in Phonology*, M. Aronoff and R. T. Oehrle, Eds. Cambridge, MA: MIT Press, 1984, pp. 107–136.

[8] E. Selkirk, "The syllable," in *The structure of phonological representations*, vol. 2, Dordrecht, The Netherlands: Foris, 1982, pp. 337–383.

[9] F. R. Eckman and G. K. Iverson, "Sonority and markedness among onset clusters in the interlanguage of ESL learners.," *Second Language Research*, vol. 9, pp. 234–252, 1993.

[10] F. R. Eckman and G. K. Iverson, "Pronunciation difficulties in ESL: Coda consonants in English interlanguage," in *First and second language phonology*, M. Yavas, Ed. San Diego, CA: Singular Publishing, 1994, pp. 251–265.

[11] B. Hancin-Bhatt and R. M. Bhatt, "Optimal L2 syllables: interactions of transfer and developmental effects," *Studies in Second Language Acquisition*, vol. 19, pp. 331–378, 1997.

[12] E. Broselow and D. Finer, "Parameter setting in second language phonology and syntax," *Second Language Research*, vol. 7, pp. 35–59, 1991.

[13] L. Davidson and M. Stone, "Epenthesis versus gestural mistiming in consonant cluster production: an ultrasound study," presented at the Proc. of the West Coast Conference on Formal Linguistics, 2003, vol. 22, pp. 165–178.