The production of Mandarin /r/ by early and late Japanese-Mandarin bilinguals: articulatory and acoustic findings

Zhiqiang Zhu, Peggy Pik Ki Mok

The Chinese University of Hong Kong
zhiqiangzhu@link.cuhk.edu.hk, peggymok@cuhk.edu.hk

ABSTRACT

The acquisition of English /r/ by Japanese learners is a well-known issue in speech acquisition. However, few studies have been done on how Japanese learners acquire rhotic sounds in other languages. This study aims to fill this gap by collecting both articulatory and acoustic data of Mandarin /r/ sounds by Japanese learners and native Beijing Mandarin speakers. Past studies reported that early language experience could benefit the speakers in the long run. This study also focuses on comparison between early bilinguals and late bilinguals. The results suggested that the early bilinguals had an overall advantage over the late bilinguals. The late bilinguals tended to use similar articulation gestures for all allophones of Mandarin /r/, have longer duration for postvocalic /r/ and syllabic /r/, and produce Mandarin /r/ with higher F2 and F3. The results will be discussed under the theoretical frameworks of the SLM, PAM, and NLM-e.

Keywords: Mandarin /r/, Japanese learners, simultaneous bilingual, articulation, speech acquisition

1. INTRODUCTION

Since the 1980s, the acquisition of English /r/ by Japanese learners has become a well-known topic in second language acquisition [1], [2], which directly influenced the development of many popular speech acquisition theories, e.g., SLM [3], [4], PAM [5], [6], NLM [7], [8]. There are two main reasons why Japanese learners have difficulty learning English /r/. First, rhotic sounds do not exist in standard Japanese. Second, Japanese learners typically tend to assimilate both English /r/ and /l/ into a single sound category, the Japanese /ɾ/ [9]. However, the assimilation is asymmetric, Japanese learners are more likely to assimilate English /l/ than English /r/ to Japanese /ɾ/ [10], [11] .

Despite intensive studies on the acquisition of English /r/ by Japanese, only a few studies have been done on Japanese learners' acquisition of rhotic sounds in other languages [12], for example, Mandarin Chinese /r/. Table 1 compares the Mandarin /r/, English /r/, and Japanese /ɾ/ in terms of syllable position, acoustic feature, and articulatory feature. Given the similarities and differences between Mandarin /r/ and English /r/, it would be theoretically interesting to examine Japanese learners' production of Mandarin /r/.

Another focus of this study is the comparison between early language learners (Japanese-Mandarin simultaneous bilingual, SB speakers) and late language learners (advanced Japanese learners of Mandarin, AJ speakers). The differences between early and late language learners in speech acquisition are a long-debated issue. Compared with late bilinguals, cross-linguistic influences between the two languages start early on and last for a longer time for early bilinguals, which can lead to non-native performance in vulnerable domains [15]. However, some studies have shown that early language experience, even over-hearing, can significantly boost a speaker's production and perception of that language later in life compared to L2 learners with no prior experience [16], [17]. Nevertheless, it is still unclear whether such an advantage can be extended to articulation. Using ultrasound imaging, this study can help to see, first, if early bilinguals generally perform better than the late bilinguals in producing Mandarin /r/; second, whether such an advantage of early bilinguals can be seen in Mandarin /r/ articulation.

	Mandarin /r/	**English /r/**	**Japanese /ɾ/**
Syllable position	Prevocalic, syllabic, postvocalic.	Prevocalic, syllabic, postvocalic.	Only prevocalic.
Acoustic feature	Low F3, but relatively higher than English /r/.	Low F3.	High F2 and F3 [13].
Articulatory feature	Prevocalic /r/: only bunched gesture. Syllabic and postvocalic /r/: both bunched and retroflex gestures [14].	Bunched and retroflex gestures are used in all syllable positions.	Apico-alveolar tap.

Table 1: Comparisons between Mandarin /r/, English /r/, and Japanese /ɾ/.

For theoretical predictions, the SLM and the PAM differ regarding Japanese learners' articulation of Mandarin /r/ sounds. The SLM claims that once the learners can discern the difference between an L2 sound from its closest native sound, they can establish a new sound category. However, sound categories are established based on auditory cues; therefore, Japanese learners may not acquire the subtle articulatory variation within Mandarin /r/. However, the PAM claims that articulatory gestures serve as the primitives for speech perception; thus, the Japanese learners can learn articulatory variation for the Mandarin /r/ sounds. In addition, the NLM-e claims that simultaneous bilinguals can successfully map phonetic features of two languages onto separate perceptual spaces like the monolinguals did for their one native language. Therefore, compared with the AJ speakers, the SB speakers have already well-established the Mandarin /r/ category in their mental representation and should have a more nativelike production performance than the AJ speakers.

2. METHOD

2.1. Participants

This study consists of three groups: 8 advanced L1 Japanese learners of Mandarin (AJ speakers), 8 Japanese-Mandarin simultaneous bilinguals (SB speakers) and 8 native Beijing Mandarin speakers (NM speakers). Details of the participants are listed in Table 2. Notably, the AJ speakers (late bilinguals) acquired Mandarin after adulthood, they were classified as "advanced" by two main criteria: they must have over one year of immersion in Beijing and have passed the HSK-6 test, which is the highest level for Mandarin learners. This study chose "advanced" learners to minimize language proficiency's influence on comparison between early and late bilinguals. The SB learners are early bilinguals of both Japanese and Mandarin because at least one of the SB speakers' parents was a native Mandarin speaker. The SB learners were exposed to both Japanese and Mandarin from an early age.

Group of speakers	Details
Advanced L1 Japanese speakers (AJ speakers)	8 participants (2M, *Mean age* = 31, *SD* = 5.55), > 1-year immersion in Beijing and with HSK-6 level.
Japanese-Mandarin simultaneous bilingual speakers (SB speakers)	8 participants (4M, *Mean age* = 24, *SD* = 2.98). All of them were exposed to both Japanese and Mandarin from an early age, with their mothers being native Mandarin speakers. All with HSK-6 level.
Native Beijing Mandarin speakers (NM speakers)	8 participants (3M, *Mean age* = 21, *SD* = 2.06). All of them were born and raised in Beijing.

Table 2: Information of participants.

2.2. Stimuli and Procedure

Three sets of target stimuli were designed for the Mandarin /r/ sounds, including 14 tokens for Chinese prevocalic /r/ (e.g., "如, [ɹu]), 3 tokens for Chinese syllabic /r/ (e.g., "儿," [ɚ]), 32 tokens for Chinese postvocalic /r/ (e.g., "皮儿," [pʰiɹ]). All the stimuli were produced together with a carrier phrase, "他答__吧" (he answered__). All the stimuli were randomized with three repetitions. Thus, for each participant, there were (14+3+32) tokens × 3 repetitions = 147 tokens in total.

The recording sessions took place at a linguistic laboratory in Beijing. The articulatory data was collected with the Telemed Echo B ultrasound system (framerate at 81.6 fps, probe field of view at 92 degrees, depth at 80mm). The acoustic data was then gained by synchronizing with the ultrasound video using the AAA software (44.1kHz/16-bit sampling rate) [18].

The participants were seated on a chair with their jaws resting on the ultrasound probe during experiment. The participants also wore a specially designed helmet for ultrasound stabilization. Target words embedded into carrier phrases were shown in PPT slides on a laptop screen in front of the participants. The participants were instructed to read sentences at a normal speech rate without any pauses. Before the recording, the experimenter would fine-tune the placement of the ultrasound settings for an optimal image.

2.3. Measurement and data analysis

The articulatory analysis included both tongue shape categorization and tongue contour comparison. Following the tongue shape categorization criteria of previous studies [19], [20], this study classifies the Mandarin /r/ into two types, retroflex gesture (the tongue tip curling up) and bunched gesture (no sign of retroflex, the tongue dorsum is bunched). The tongue splines were drawn manually at the key frame where the maximal constriction of Mandarin /r/ occurred. 42 equally spaced data points were exported as x–y Cartesian coordinates without rotation for each spline. The tongue contours comparison was then made via smoothing spline ANOVA using the "ggplot2" and "ggs" packages in R [21], [22]

For the acoustic data, the /r/ sound, together with its adjacent vowel, was labelled in Praat, and the acoustic target of the Mandarin /r/ was identified at the minimum F3 point. Duration of the /r/ sound, F1, F2, F3, and F3-F2 of the acoustic target was modelled in linear mixed effect (LME) models, respectively [23]. Due to page limits, only the results of duration, F2, and F3 are reported here.

In addition, four native Beijing Mandarin speakers (4F, mean age = 22.5, SD = 1.91) were invited on giving ratings to all the target stimuli. The native judges were university students without hearing impairment. The target stimuli with their carrier phrase were randomized across all the speakers. Before the experiment, the native judges were told that they would hear a series of utterances produced by either native Beijing Mandarin speakers or Japanese learners. The native judges were then instructed to give ratings to the target stimuli on goodness and accentedness based on 7-point scales, from 1 (very good; no accent) to 7 (very bad; heavy accent). The ratings averaged across the four native speakers were reported for each group.

3. RESULTS

3.1. Articulatory results

The results of tongue shape categorization showed that the SB and NM groups have similar patterns for the Mandarin /r/ gestures, which differed from the AJ speakers. For the SB and NM groups, only bunched gesture is used for prevocalic /r/. Both retroflex and bunched gestures are used for syllabic /r/ and postvocalic /r/. However, retroflex gestures and bunched gestures are found in all phonological positions by the AJ speakers; four out of eight AJ speakers have used retroflex gesture for the prevocalic /r/. For illustration, Figure 1 shows the raw ultrasound images of Mandarin /ri/ by three speakers in separate groups. It is evident that AJ3 curled up the tongue to produce the Mandarin prevocalic /r/ (/ri/), while SB2 and NM1 both used bunched gestures for the Mandarin /ri/.

Figure 2 shows the tongue contours of the Mandarin /r/ sounds by the AJ2, SB2, and NM2 speakers, respectively. Due to page limits, this study cannot list all the participants' SSANOVA tongue contour images. The AJ2, SB2 and NM2 speakers were selected as typical productions from

speakers to illustrate the different tongue configurations. The results also yielded similar patterns between the SB and NM groups but not for the AJ group. As displayed in Figure 2, the SB and NM speakers (represented by SB2 and NM2) employed different articulatory gestures to produce the Mandarin /r/ sounds. They adopted similar tongue gestures for syllabic /r/ and postvocalic /r/, but not for prevocalic /r/. However, five out of eight AJ speakers, like AJ2, used similar tongue gestures for the Mandarin /r/ sounds at all syllable positions.

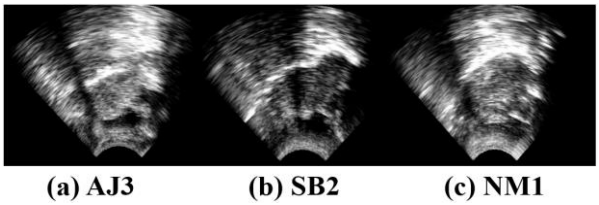

Figure 1: Raw ultrasound images of Mandarin /ri/ by the AJ3, SB2 and NM1 speakers respectively.

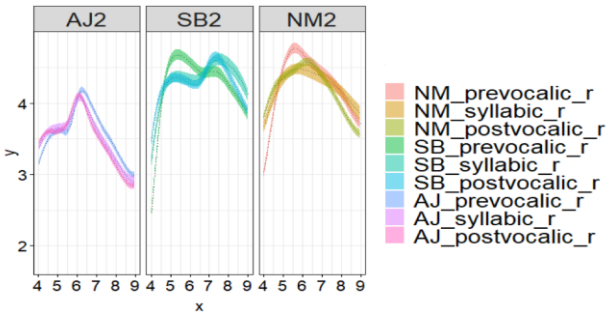

Figure 2: Tongue contours of the Mandarin /r/ sounds by the NM2, SB2, and AJ2 speakers, respectively.

3.2. Acoustic results

Duration: Figure 3 shows the duration of Mandarin /ɹ/ sounds at different syllable positions by the three groups. The best model for duration included Group, Syllable position, and the interaction between Group and Syllable position as fixed effects, and Participant and Utterance as random intercepts, with a random slope for Participant and Utterance on Syllable position, respectively. The model results showed that there was a main effect on Group and a main effect on Syllable position. The model results also showed a significant two-way interaction between Group and Syllable position. Post-hoc pairwise comparisons were performed on the duration in each syllable. The results showed that, in the prevocalic position, the duration between each two groups remained non-significant (p > .05). In the syllabic position and the postvocalic position, the duration of the AJ group was significantly longer than that of the NM group (for the syllabic position, Estimate =61.259, SE = 22.5, t = 2.717, p = .027; for the postvocalic position, Estimate =76.632, SE = 21.3, t = 3.600, p = .003). However, in the syllabic and postvocalic positions, the duration between the SB group and the NM group remained non-significant (p > .05).

F2: Figure 4 shows the F2 values of Mandarin /ɹ/ sounds at different syllable positions by the AJ, SB, and NM groups. The best model for F2 included Group, Syllable position, and the interaction between Group and Syllable position as fixed effects, and Participant and Utterance as random intercepts, with a random slope for Participant and Utterance on Syllable position, respectively. The model results showed that there was a main effect on Group and a main effect on Syllable position. The model results also yielded a significant two-way interaction between Group and Syllable position. Post-hoc pairwise comparisons were performed on formant values in each syllable. The results suggested that the F2 values of the AJ group were significantly higher than that of the NM and SB groups in all three syllable positions (compared with the NM group, in the prevocalic position, Estimate =1.079, SE = 0.238, t = 4.530, p < .001; in the syllabic position, Estimate =0.675, SE = 0.267, t = 2.528, p =.038; in the postvocalic position, Estimate =0.655, SE = 0.235, t = 2.785, p =.024). However, the F2 values remain non-significant between the SB and the NM group in all syllable positions (p > .05).

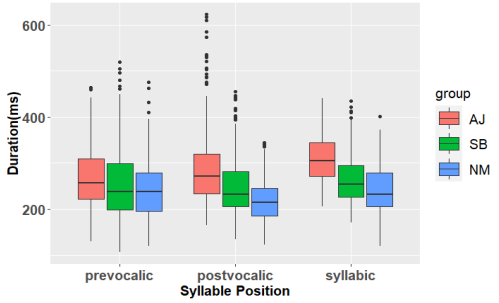

Figure 3: Duration of Mandarin /ɹ/ sounds at different syllable positions by the AJ, SB, and NM groups.

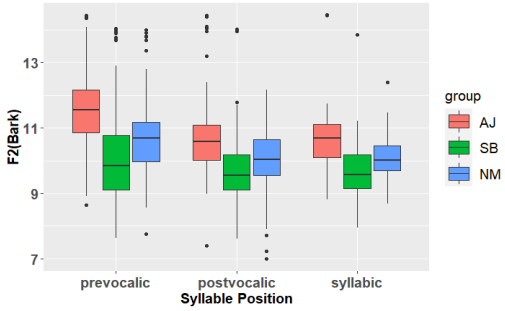

Figure 4: F2 values of Mandarin /ɹ/ sounds at different syllable positions by the AJ, SB, and NM groups.

F3: Figure 5 shows the F3 values of Mandarin /ɹ/ sounds at different syllable positions by the AJ, SB, and NM groups. The best model for F3 included Group and Syllable position as fixed effects, and Participant and Utterance as random intercepts, with a random slope for Participant and Utterance on Syllable position, respectively. The model results showed that there was a main effect on Group and a main effect on Syllable position. Post-hoc pairwise comparisons were performed on formant values in each syllable. The results were similar to the F2 values, which suggested that the F3 values of the AJ group were significantly higher than that of the NM and SB groups in all three syllable positions (compared with the NM group, in the prevocalic position, Estimate =1.042, SE = 0.245, t = 4.261, p < .001; in the syllabic position, Estimate =0.918, SE = 0.256, t = 3.582, p = .003; in the postvocalic position, Estimate =1.103, SE = 0.244, t = 4.526, p < .001). However,

the F3 values remain non-significant between the SB and the NM group in all syllable positions ($p > .05$).

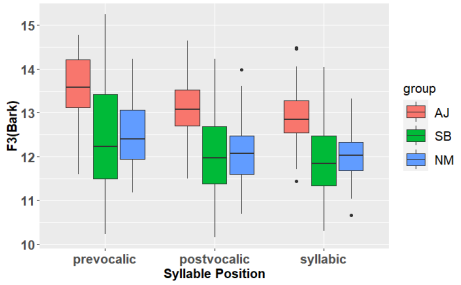

Figure 5: F3 values of Mandarin /ɹ/ sounds at different syllable positions by the AJ, SB, and NM groups.

3.3. Ratings by native speakers

Figure 6 shows the averaged ratings of the target stimuli by the four native Mandarin speakers. Higher scores in the accentedness and goodness ratings indicated that the target stimuli were judged as more accented and less perfect, respectively. For the goodness rating, the scores in all syllable positions ranked as AJ group > SB group > NM group. It was the same case for the accentedness rating in all syllable positions. The rating results suggested that the native Mandarin speakers gave better scores to the SB group than that of the AJ group for all allophones of Mandarin /r/ sounds.

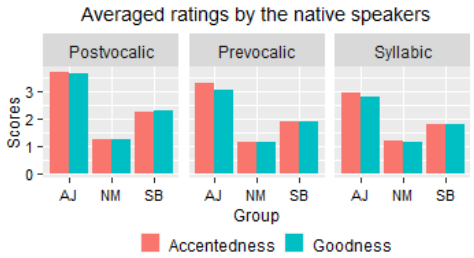

Figure 6: Averaged ratings of goodness and accentedness by the native Mandarin speakers.

4. DISCUSSION

This study examines the production of Mandarin /r/ by Japanese-Mandarin simultaneous bilinguals and advanced Japanese learners of Mandarin, which confirms that simultaneous bilinguals (early bilinguals, SB) have an overall advantage over advanced Japanese learners (late bilinguals, AJ). Compared with the SB and NM groups, the production differences of the AJ group are mainly reflected in four aspects: 1) using the retroflex gesture for prevocalic /r/, 2) employing similar gestures for all allophones of Mandarin /r/, 3) having excessively longer duration for syllabic /r/ and postvocalic /r/, 4) producing Mandarin /r/ sounds with significantly higher F2 and F3. Also, the ratings by the native Mandarin speakers showed that scores of the SB group were better than that of the AJ group in both goodness and accentedness. The advantage of the early bilinguals over the late bilinguals corroborated the NLM-e, which claims that simultaneous bilinguals can successfully acquire phonetic features in both languages. This study demonstrates that the SB group successfully acquired phonetic features of Mandarin /r/, while the AJ group showed more production biases. The following discussion will compare the production performance between the AJ and SB groups in terms of articulation, duration, and formant values, respectively.

Articulation: the articulatory patterns suggested that the SB and NM speakers could discern subtle articulatory variations between Mandarin prevocalic /r/ and Mandarin syllabic/postvocalic /r/, while some AJ speakers could not. Therefore, most of the AJ speakers used similar gestures to produce all the allophones of Mandarin /r/ sounds. In addition, half of AJ speakers transferred their retroflex gesture for the syllabic/postvocalic /r/ into the prevocalic /r/. The results partially supported both the PAM and the SLM. The PAM was supported by the SB speakers and a few AJ speakers that employed different gestures for the Mandarin /r/ sounds. The SLM was supported by most of the AJ speakers that they could not acquire the subtle articulatory variations within Mandarin /r/.

Duration: the AJ group demonstrated significantly longer durations in producing the syllabic /r/ and the postvocalic /r/ compared to the SB and NM groups. Excessively longer durations in segments are a common production bias observed in second language acquisition [24]. It is possible that the AJ speakers exaggerated the duration of these sounds to intensify the rhoticity of the syllabic /r/ and postvocalic /r/. However, the SB and NM groups did not use this technique. Instead, the rhoticity of the Mandarin /r/ for these groups was primarily reflected in the low F3 and small distance between F3 and F2.

Formant values: compared with the SB and NM speakers, the AJ speakers produced Mandarin /r/ in all syllable positions with higher F2 and F3. Since a low F3 is an essential indicator of rhotic sounds [25], a higher F3 suggested that the Mandarin /r/ sounds produced by the AJ speakers were less rhotic compared with that of the SB and NM groups. In addition, Japanese /ɾ/ is characterized by high F2 and F3. Compared with the SB speakers, the Mandarin /r/ sounds produced by the AJ speakers were more heavily influenced by their Japanese /ɾ/. This finding is also compatible with the claims of NLM-e, because the SB speakers had better performance in acquiring the phonetic features of the Mandarin /r/ than the AJ speakers.

In conclusion, this study demonstrates that the Japanese-Mandarin simultaneous bilinguals (early bilinguals) outperformed the advanced Japanese of Mandarin (late bilinguals) in Mandarin /r/ production. The production biases of the late bilinguals were manifested in articulation, duration, and formant values. The results of this study corroborated the NLM-e, while the SLM and the PAM were only partially supported. This study has three main implications for future research. First, to further investigate the assimilation relationship between Mandarin /r/-/l/ to the Japanese /ɾ/, the production of Mandarin /l/ and Japanese /ɾ/ by Japanese learners and native speakers should be examined as well. Second, more participants should be recruited for both Japanese-dominant simultaneous bilinguals and Mandarin-dominant simultaneous bilinguals to investigate the effect of language dominance. Third, perception experiments should be conducted to test perceptual relationship between Mandarin /r/-/l/ and Japanese /ɾ/.

5. 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," *Percept. Psychophys.*, vol. 18, no. 5, pp. 331–340, 1975.

[2] S. G. Guion, J. E. Flege, R. Akahane-Yamada, and J. C. Pruitt, "An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants," *J. Acoust. Soc. Am.*, vol. 107, no. 5, pp. 2711–2724, 2000.

[3] J. E. Flege, "Second language speech learning: Theory, findings, and problems," in *Speech perception and linguistic experience: Issues in cross-language research*, W. Strange, Ed. Timonium, MD: York Press, 1995, pp. 233–277.

[4] J. E. Flege and O.-S. Bohn, "The revised Speech Learning Model (SLM-r)," doi: 10.13140/RG.2.2.27529.06249.

[5] C. T. Best, "A direct realist view of cross-language speech perception," in *Speech perception and linguistic experience: Issues in cross-language research*, W. Strange, Ed. York Press, 1995, pp. 171–204.

[6] C. T. Best and M. D. Tyler, "Nonnative and second-language speech perception: Commonalities and complementarities," in *Language experience in second language speech learning: In honor of James Emil Flege*, O.-S. Bohn and M. J. Munro, Eds. John Benjamins, 2007, pp. 13–34.

[7] P. K. Kuhl and P. Iverson, "Linguistic experience and the 'perceptual magnet effect,'" in *Speech perception and linguistic experience: Issues in cross-language research*, W. Strange, Ed. Timonium, MD: York Press, 1995, pp. 121–154.

[8] P. K. Kuhl, B. T. Conboy, S. Coffey-Corina, D. Padden, M. Rivera-Gaxiola, and T. Nelson, "Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e)," *Philos. Trans. Biol. Sci.*, vol. 363, no. 1493, pp. 979–1000, 2008.

[9] N. Takagi and V. Mann, "The limits of extended naturalistic exposure on the perceptual mastery of English/r/and/l/by adult Japanese learners of English," *Appl. Psycholinguist.*, vol. 16, no. 4, pp. 380–406, 1995.

[10] K. Aoyama, J. E. Flege, S. G. Guion, R. Akahane-Yamada, and T. Yamada, "Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese/r/and English/l/and/r," *J. Phon.*, vol. 32, no. 2, pp. 233–250, 2004.

[11] K. Hattori and P. Iverson, "English/r/-/l/category assimilation by Japanese adults: Individual differences and the link to identification accuracy," *J. Acoust. Soc. Am.*, vol. 125, no. 1, pp. 469–479, 2009.

[12] J. Larson-Hall, "Predicting perceptual success with segments: a test of Japanese speakers of Russian," 2004.

[13] P. Iverson, V. Hazan, and K. Bannister, "Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English/r/-/l/to Japanese adults," *J. Acoust. Soc. Am.*, vol. 118, no. 5, pp. 3267–3278, 2005.

[14] S. Chen and P. P. K. Mok, "Articulatory and acoustic features of Mandarin/ɹ: A preliminary study," in *2021 12th International Symposium on Chinese Spoken Language Processing (ISCSLP)*, 2021, pp. 1–5.

[15] V. Yip and S. Matthews, *The bilingual child: Early development and language contact*. Cambridge University Press, 2007.

[16] L. M. Knightly, S.-A. Jun, J. S. Oh, and T. K. Au, "Production benefits of childhood overhearing," *J. Acoust. Soc. Am.*, vol. 114, no. 1, pp. 465–474, 2003.

[17] S. A. Montrul, "Incomplete Acquisition in Bilingualism: Re-examining the Age Factor," vol. 39, Sep. 2008, doi: 10.1075/SIBIL.39.

[18] Articulate Instruments Ltd, "Articulate Assistant Advanced User Guide: Version 2.16." Edinburgh, UK, 2012.

[19] H. B. Klein, T. M. Byun, L. Davidson, and M. I. Grigos, "A multidimensional investigation of children's/r/productions: Perceptual, ultrasound, and acoustic measures," *Am. J. Speech-Language Pathol.*, vol. 22, pp. 540–553, 2013.

[20] P. Delattre and D. C. Freeman, "A dialect study of American r's by x-ray motion picture," *Linguistics*, vol. 6, no. 44, pp. 29–68, 1968.

[21] H. Wickham, *ggplot2: elegant graphics for data analysis*. New York: Springer, 2009.

[22] C. Gu, "Smoothing spline ANOVA models: R package gss," *J. Stat. Softw.*, vol. 58, pp. 1–25, 2014.

[23] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *J. Stat. Softw.*, vol. 67, no. 1, pp. 1–48, 2014.

[24] R. E. Baker, M. Baese-Berk, L. Bonnasse-Gahot, M. Kim, K. J. Van Engen, and A. R. Bradlow, "Word durations in non-native English," *J. Phon.*, vol. 39, no. 1, pp. 1–17, 2011.

[25] C. Y. Espy-Wilson, S. E. Boyce, M. Jackson, S. Narayanan, and A. Alwan, "Acoustic modeling of American English/r," *J. Acoust. Soc. Am.*, vol. 108, no. 1, pp. 343–356, 2000.