ANALYSIS OF L2 ENGLISH VOWEL PRODUCTION BY NATIVE JAPANESE CHILDREN IN DOMESTIC ELEMENTARY SCHOOL

Tsuneo Kato, Fumika Mizutani, Yuumi Mizukoshi, Kohei Kitamura and Seiichi Yamamoto

Doshisha University
tsukato@mail.doshisha.ac.jp

ABSTRACT

Children starting at an early age learn to produce second language (L2) speech faster than adults, but their production is nonetheless affected by their native language (L1). We are conducting a longitudinal collection of L2 English speech produced by native Japanese children in a domestic elementary school to reveal how children's L2 speech changes. This paper focuses on their vowel production, one of the major differences between Japanese and English, and reports initial results of analyses on its formant and duration. As in previous studies, Japanese children had difficulty distinguishing some neighboring tense and lax vowels by articulation, but differentiated the vowels by controlling the duration. Furthermore, the analysis suggested that many of the children attempted to distinguish /æ/ and /ɚ/ from /ʌ/ and /ɑ/ by articulation.

Keywords: L2 English vowel production, native Japanese children, formant, duration

1. INTRODUCTION

Many studies have shown that learning a second language earlier is better in learning production of speech sounds [1, 2, 3]. In Japan, compulsory English education is going to start from the 5th grade of elementary schools, i.e. 11 years old, two years earlier than at present to enable the children to learn communication skills in English more effectively. However, English speech of native Japanese children who learn English as a foreign language domestically has not been analyzed or recorded longitudinally. Therefore, we started collecting English speech produced by 89 native Japanese children in an elementary school two years ago and have been collecting their speech every six months.

One of the major differences in speech sounds between Japanese and English is the vowel system. Japanese has two short and long sets of five monophthongal vowels: /i/, /e/, /a/, /o/ and /ɯ/. The short and long vowels do not differ in quality, but are mono-moraic and bi-moraic in duration. On the other hand, American English has more than ten monophthongal vowels. We set five tense vowels /i/, /ɑ/, /ɚ/, /ɔ/, /u/, and five lax vowels /ɪ/, /e/, /æ/, /ʌ/, /ʊ/ in this study.

The accuracy of L2 speech production is thought to be limited by perceptual factors. Best's Perceptual Assimilation Model (PAM) explains that the discrimination of the difference between pairs of phonemically distinct L2 sounds is influenced by the perceptual similarity between specific L1 and L2 sounds [4]. Flege's Speech Learning Model (SLM) explains that category assimilation occurs when a L2 learner fails to form a new category for an L2 sound and category dissimilation occurs when the learner forms a new category for an L2 sound and differentiates the contrast between an existing L1 category and the new L2 category [6].

L2 English vowel production of Japanese adults was studied intensively by Tsukada [12]. The thesis reported that Japanese adults have difficulty in distinguishing neighboring vowels, such as /ɑ/-/ʌ/ and /ʊ/-/u/. Lambecher et al. examined if Japanese speakers were able to improve in identifying and producing English low / mid vowels, /æ/, /ɑ/, /ʌ/, /ɔ/ and /ɚ/ [9]. Ingram and Park showed that Japanese learners of English had difficulty distinguishing spectral quality of some neighboring tense and lax vowels in production, but were good at distinguishing their duration [7]. Oh et al. showed that Japanese children learned how to produce English vowels in a native manner faster than adults through a one-year longitudinal measurement of the formant shortly after their arrival in the US and one year after the first recording [11]. However, learning of L2 production of English as a foreign language outside the Anglosphere in childhood has not been analyzed or recorded longitudinally.

This paper introduces an overview of our longitudinal L2 English speech collection in an elementary school and then analyzes the spectral quality of the learners' production of English vowels by formant analysis. We focus on two pairs /i/-/ɪ/ and /u/-/ʊ/, and low / mid vowels. Finally, we analyze duration of the vowels in comparison with native American children and try to understand Japanese children's strategies for producing L2 English vowels.

2. COLLECTION OF ENGLISH SPEECH

2.1. L2 English speech corpus of Japanese elementary school children

We collected English read-aloud speech of native Japanese children aged 10 to 11 years old biannually. The first recording was made on 63 children out of 89 at the age of 10. The second, third and fourth recordings were made on all 89 children at six-month intervals. The children were regularly taking English classes given by Japanese, American and British teachers of English three hours a week during the period. The teachers mainly taught oral communication skills in American English through conversation, role plays, and presentations. The children had learned the Latin alphabets at the age of 7 and basic phonics at the age of 8.

The texts of the recorded speech were 60 basic English words from GFTA3 (Goldman-Fristoe Test of Articulation Ver. 3), 10 English short sentences and 30 basic Japanese words for comparing L1 and L2 speech sounds. The children read aloud designated texts by viewing a slide on a tablet that showed pictures with the texts. They participated in the recordings without any practice in advance. Thus, they could not read some words or sentences. Every slide had a button to play a model utterance produced by one of the native American teachers. The children could repeat unreadable words or sentences after listening to the model utterance if they wished to. All the utterances including speech errors were saved in the corpus with a label indicating order of time and type of the utterance. However, the utterances produced after listening to the model utterance were excluded from the list for analysis. There were inevitably various consonantal contexts in the analysis of vowel production due to the children's limited vocabulary. Instead, the formant frequencies were measured at the midpoint of vowel segments and averaged over various consonantal contexts.

2.2. English speech corpus of American children

The CID speech corpus, which contains English words and short sentences produced by 436 native American children aged between 5 and 18, was used for comparison [10]. All the speakers uttered the same words and sentences. The words are "bead" (/i/), "bit" (/ɪ/), "bed" (/eh/), "bat" (/æ/), "pot" (/ɑ/), "ball" (/ɔ/), "but" (/ʌ/), "put" (/ʊ/), "boot" (/u/) and "bird" (/ɚ/) embedded in a carrier sentence "I say uh ... again". The consonantal contexts of the focused vowels were the same across the words.

3. FORMANT ANALYSIS OF VOWELS

3.1. Measuring method and data for analysis

The word utterances were processed phoneme segmentation manually using praat. The canonical phoneme sequence of a word was given by the CMU pronunciation dictionary [5]. The frequencies of the first and second formants (F1 and F2) were measured at the midpoint of every vowel segment.

The speech data of 36 randomly-sampled children, 18 males and 18 females, were processed for the analysis. Utterances repeated after listening to a model utterance, utterances with stammers and utterances with ambient noises were excluded, and only utterances in which the speaker knew the correct pronunciation and produced correctly without referring a model utterance were analyzed.

For comparison, speech data of 61 age-matched native American children, (13 female and 15 male 10-year-olds, and 18 female and 15 male 11-year-olds) were randomly sampled and analyzed.

3.2. Formant distribution of groups

The F1 and F2 frequency distributions of 10 English monophthongal vowels produced by the Japanese female and male children are shown in Figures 1 and 2, respectively. The centroids of each vowel in three terms are marked in '+', '□', and '•' with the same color, respectively. One standard deviation ellipses are drawn with a dashed line only for the last term. In the figures, the phonemes are represented in ARPABET instead of IPA, that is, "iy" for /i/, "ih" for /ɪ/, "uw" for /u/, "uh" for /ʊ/, "eh" for /e/, "ae" for /æ/, "aa" for /ɑ/, "ah" for /ʌ/, "ao" for /ɔ/ and "er" for /ɚ/. For comparison, the F1 and F2 frequency distributions of the female and the male native American children are shown in Figures 3 and 4, respectively. The centroids of 10- and 11-year-olds are marked in '+' and '∘', respectively, and one standard deviation ellipses are drawn for the 11-year-olds. Note that the distributions of the 10- and 11-year-old groups are derived from different children in the CID corpus, whereas all the distributions are derived from the same children for the Japanese case.

At first glance, the Japanese children show a narrower range of F1 and F2 frequencies than the native American children. This is not because of L2 speech. The range of their L1 speech, Japanese, was as narrow as that of their L2 speech. Notably, substantial overlaps are observed between particular pairs of a tense and a lax vowels such as, /i/(iy)-/ɪ/(ih), /u/(uw)-/ʊ/(uh) and /ɑ/(aa)-/ʌ/(ah).

We conducted MANOVA tests on F1 and F2 fre-

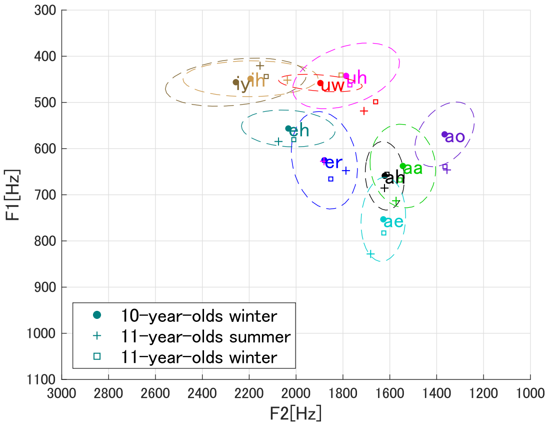

Figure 1: F1 and F2 frequencies of 10 English vowels produced by 18 Japanese female 10- and 11-year-olds.

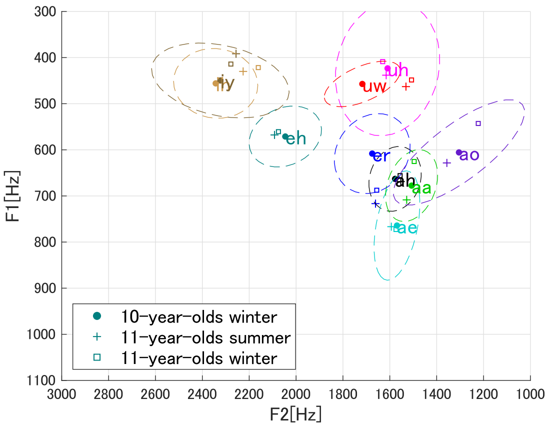

Figure 2: F1 and F2 frequencies of 10 English vowels produced by 18 Japanese male 10- and 11-year-olds.

quencies with gender being a between-subject factor, and a type of vowel and a term being within-subject factors. There were no significant differences between /i/ and /ɪ/ (p = 0.3548) or between /u/ and /ʊ/ (p = 0.3598). There was a significant difference ($p < 0.01$) among five low / mid vowels: /æ/, /ɑ/, /ʌ/, /ɔ/ and /ɚ/. Looking at the difference in detail with separate ANOVA tests on F1 and F2 frequencies, the low / mid vowels showed no significant difference in F1, but in F2 (F1: $F_{(4,132)} = 30.3$, F2: $F_{(4,132)} = 133.3$, $p < 0.01$). Bonferroni's post hoc tests in F1 showed no significant differences at four pairs: /ɑ/-/ʌ/, /ɑ/-/ɚ/, /ʌ/-/ɚ/ and /ɔ/-/ɚ/ with a significant level at 0.05. Bonferroni's post hoc tests in F2 showed no significant difference only at a pair /æ/-/ʌ/.

Furthermore, we verified the simple main effect of a type of the low / mid vowels in each term and in F1 and F2, respectively to check if the discrimina-

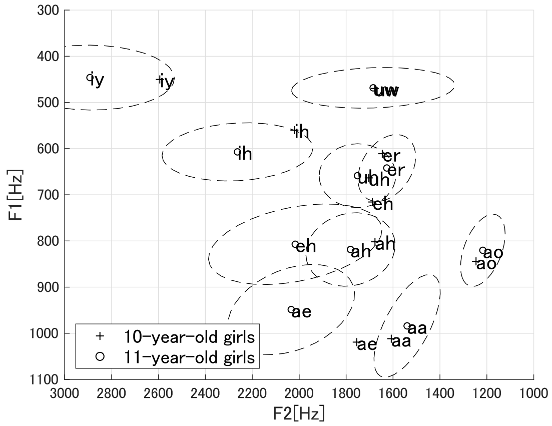

Figure 3: F1 and F2 frequencies of 10 English vowels produced by 31 American female 10- and 11-year-olds.

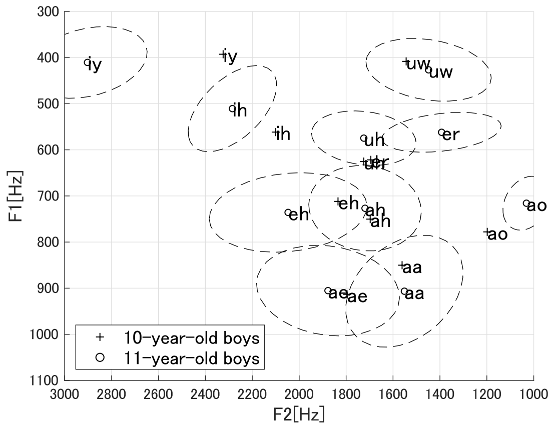

Figure 4: F1 and F2 frequencies of 10 English vowels produced by 30 American male 10- and 11-year-olds.

tion of the low / mid vowels improved over time. The F-value on F1 frequency increased over time (10-year-olds winter: $F_{(4,132)} = 13.4$, 11-year-olds summer: $F_{(4,132)} = 13.5$ and 11-year-olds winter: $F_{(4,132)} = 28.4$). The F-value on F2 frequency also increased (10-year-olds winter: $F_{(4,132)} = 40.4$, 11-year-olds summer: $F_{(4,132)} = 72.7$ and 11-year-olds winter: $F_{(4,132)} = 74.3$). This result suggested that many of the Japanese children attempted to differentiate the low / mid vowels. The centroids of /ɚ/(er) and /æ/ are separating from those of /ɑ/ and /ʌ/ in Figures 1 and 2.

We conducted MANOVA tests on F1 and F2 frequencies of the native American children with gender and an age being between-subject factors and a type of vowel being a within-subject factor. Both of the pairs /i/-/ɪ/ and /u/-/ʊ/, and the low / mid vowels showed significant differences (/i/-/ɪ/: $p < 0.01$, /u/-/ʊ/: $p < 0.01$, and the low / mid vowels: $p < 0.01$).

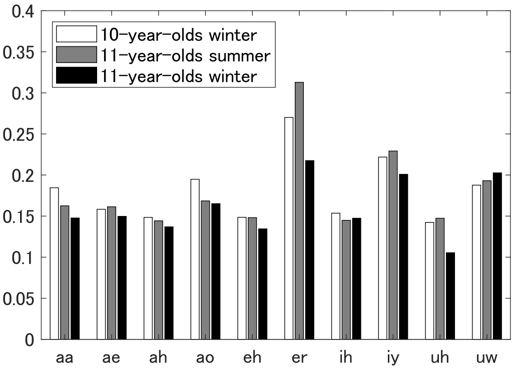

Figure 5: Mean durations of 10 English vowels produced by 36 Japanese 10- and 11-year-olds.

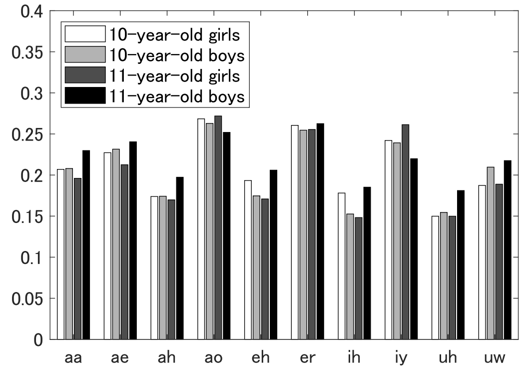

Figure 6: Mean durations of 10 English vowels produced by 61 American 10- and 11-year-olds.

Considering the results all together, while it was difficult for Japanese children to distinguish neighboring vowels, such as /i/-/ɪ/, /u/-/ʊ/, and low / mid vowels, many of them attempted to differentiate /æ/ and /ɚ/ from /ʌ/ and /ɑ/. The ANOVA test showed that the differentiation of the phonemes in formant frequencies increased over one year on average.

4. DURATION ANALYSIS OF VOWELS

Duration of the vowel segments was measured on the basis of the manual segmentation of phonemes. Figure 5 shows the mean duration of the vowels produced by native Japanese children for three terms. Figure 6 shows those measured on the CID corpus for comparison. Figure 5 does not separate gender because there was no significant difference. Comparing two pairs of a tense and a lax vowels /i/-/ɪ/ and /u/-/ʊ/, Japanese and American children were similar in that mean durations of the tense vowels /i/ and /u/ were significantly longer than those of the lax vowels /ɪ/ and /ʊ/. On the other hand, Japanese children showed mostly the same duration for four low / mid vowels /ɑ/, /æ/, /ʌ/, and /ɔ/, whereas American children differentiated the duration of the vowels.

Lingua franca core postulated that discrimination of the tense and lax vowels with duration is important for L2 learners who cannot distinguish the neighboring vowel sounds by articulation [8]. An ANOVA test was conducted on the duration with gender being a between-subject factor, and a type of vowel and a term being within-subject factors. There were significant differences in duration between /i/ and /ɪ/ ($F_{(1,34)} = 488.6$, $p < 0.01$) and /u/ and /ʊ/ ($F_{(1,28)} = 51.5$, $p < 0.01$). There was a significant difference among four low / mid vowels /æ/, /ɑ/, /ʌ/ and /ɔ/ ($F_{(3,99)} = 20.4$, $p < 0.01$). This is because the standard deviation of each vowel was as small as 0.005 second. Bonferroni's post hoc tests showed significant differences for all combinations except for two: /æ/-/ɑ/ and /ɑ/-/ɔ/ at a significant level at 0.05. Note that F-values over the three terms did not show a clear tendency for durations between /i/ and /ɪ/, or between /u/ and /ʊ/, or across the low / mid vowels.

Considering the results together, the Japanese children differentiated producing /i/-/ɪ/ and /u/-/ʊ/ by duration control while they attempted to differentiate producing the low / mid vowels by articulation.

5. CONCLUSIONS

Production of L2 English vowels by native Japanese children was analyzed through measurement of F1 and F2 frequencies and duration on English word utterances biannually collected in an elementary school in Japan. The analysis showed some of their strategies for producing English vowels. Although Japanese children generally had difficulty distinguishing the spectral quality of the neighboring tense and lax vowels /i/-/ɪ/, /u/-/ʊ/ and a group of low / mid vowels /ɑ/, /æ/, /ʌ/, /ɔ/ and /ɚ/, they differentiated the pairs of /i/-/ɪ/ and /u/-/ʊ/ by controlling the segmental duration. Furthermore, many of the children attempted to differentiate two vowels /æ/ and /ɚ/, both of which are new to them, from /ʌ/ and /ɑ/ by articulation. The degree of distinguishing the low / mid vowels was considered to increase over time due to the fact that the F-value of the ANOVA test in the low / mid vowel group was increasing.

6. ACKNOWLEDGEMENT

We are grateful to Ms. Arisa Furimoto and teachers in Doshisha elementary school for continuous support in speech data collection, and to Professor Mariko Sugahara for helpful discussions. This work was supported by JSPS KAKENHI Grant Number 17K02954.

7. REFERENCES

[1] Aoyama, K., Flege, J., Guion, S., Akahane-Yamada, R., Yamada, T. 2004. Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. *J. Phonetics* 23, 233–250.

[2] Aoyama, K., Guion, S., Flege, J., Yamada, T., Akahane-Yamada, R. 2008. The first years in an L2 speaking environment: A commparison of Japanese children and adults learning American English. *Intl. J. of Applied Linguistics* 46, 61–90.

[3] Baker, W., Trofimovich, P. 2005. Interaction of native- and second-language vowel system(s) in early and late bilinguals. *Language & Speech* 48, 1–27.

[4] Best, C. 1994. The emergence of native-language phonological influences in infants: A perceptual assimilation model. In: Goodman, C., Nusbaum, H., (eds), *The development of speech perception*. Cambridge: MIT Press 167–224.

[5] The CMU Pronouncing Dictionary. http://www.speech.cs.cmu.edu/cgi-bin/cmudict.

[6] Flege, J. 1995. Second-language speech learning: Theory, findings, and problems. In: Strange, W., (ed), *Speech perception and linguistic experience: Theoretical and methodological issues*. Timonium, MD: York Press 233–277.

[7] Ingram, J., Park, S. 1997. Cross-language vowel perception and production by Japanese and Korean learners of English. *J. Phonetics* 25, 343–370.

[8] Jenkins, J. 2000. *The Phonology of English as an International Language*. Oxford: Oxford University Press.

[9] Lambacher, S., Martens, W., Kakehi, K., Marasinghe, C., Molholt, G. 2005. The effects of identification training on the identification and production of american english vowels by native speakers of Japanese. *Applied Psycholinguistics* 26, 227–247.

[10] Lee, S., Potamianos, A., Narayanan, S. 1999. Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *J. Acoust. Soc. Am.* 105(3), 1455–1468.

[11] Oh, G., Guion, S., Aoyama, K., Flege, J., Akahane-Yamada, R., Yamada, T. 2011. A one-year longitudinal study of English and Japanese vowel production by Japanese adults and children in an English-speaking setting. *J. Phonetics* 39, 156–167.

[12] Tsukada, K. 1999. *An acoustic phonetic analysis of Japanese-accented English*. Sydney, Australia: Doctoral dissertation in Macquarie University.