



Acoustic properties of shared vowels in bilingual Mandarin-English children

Jing Yang^{1,2}, Robert Allen Fox¹

¹Department of Speech and Hearing Science, Ohio State University, USA

²Department of Communication Sciences and Disorders, University of Central Arkansas, USA
yang.1198@osu.edu, fox.2@osu.edu

Abstract

This study investigates L1-L2 interactions in relatively young Mandarin (L1) - English (L2) bilingual children through comparing their static and dynamic vowel acoustic features with those of age-matched corresponding monolingual children. Two groups of sequential bilingual children aged 5-6 years (one with low proficiency in English and the other with high proficiency in English) were recorded producing a set of words containing the shared vowels /a/, /i/, /u/ in both languages. Age-matched monolingual children only produced the words in their native language. F1 and F2 values were measured at 5 equidistant time locations. It is found that both groups of bilingual children showed distinctive vowel dispersion patterns and dynamic spectral change patterns from those of monolingual children. Low proficiency bilingual children showed an assimilatory process of L1 on L2 and high proficiency bilingual children showed an assimilatory process of L2 on L1.

Index Terms: acoustic characteristics, shared vowels, Mandarin-English bilingual children

1. Introduction

The interdependence of two language systems in bilingual speakers has been widely acknowledged in previous studies [1, 2, 3, 4]. In the subsystem of phonetics, it has been proposed that a bilingual's L1 and L2 would naturally influence each other because of coexistence of these two vowel systems in a common space [5]. Since then, a large number of studies examined the L1-L2 interaction effect. Generally, there exist two types of interaction effects: the effect of L1 on L2 and vice versa. In terms of the effect of L1 on L2, as Watson [6] pointed out, for sequential bilinguals who learn a L2 after complete or relatively complete acquisition of L1, their native language (L1) is used as the base to establish the L2 phonetic system. In this case, bilinguals are likely to assimilate the phonetically similar L2 sounds into established L1 categories at the beginning of L2 acquisition. However, not every single L2 sound will be equally assimilated to a L1 sound category. The extent to which the L2 sound is assimilated to the L1 sound category is primarily determined by the phonetic-acoustic similarity between the L2 and L1 sounds (the Speech Learning Model [5] and the Perceptual Assimilation model [7]). Following continuous exposure and immersion in L2, the influence of L1 on L2 is attenuated and separate L2 sound categories are eventually established.

Unlike the convergent findings of the effect of L1 on L2 in early stage L2 speakers, there are divergent opinions on the influence of L2 on L1. Some studies have reported a change in L1 as a result of L2 learning in both perception [8, 9] and production [10, 11, 12, 13, 14]. However, other studies have observed no change of L1 sounds as a function of L2 experience [15]. This study examined the production of vowels /i, ɪ, ε/ (similar vowels which occur in both German and English) and /æ/ (which occurs only in English). The

results showed that the production of native German vowels by experienced German speakers of English was not influenced by the long-term immersion in English (L2).

Second, among those studies which reported an influence of L2 on L1, two opposite processes have been observed. One type of mechanism is a dissimilatory process and the other type is an assimilatory process. The dissimilarity process takes place when speakers shift a phonetically similar L1 sound away from the L2 sound to make a contrast between the two sounds, establishing a new phonetic category in L2 [5, 16]. For example, one study [17] found that Quichua (L1) vowels systematically raised and moved away from the similar L2 vowels in Quichua-Spanish bilinguals who have developed distinct vowel categories for the L2 (Spanish). However, researchers have also found instances of a shift in the L1 sound towards an acoustically similar L2 sound, which represents an assimilatory process [11, 12]. One study found the forward movement of French (L1) /u/ as a result of influence from English (L2) /u/ which is located in a more fronted position relative to French /u/ [11]. Another study found that only six weeks' intensive immersion in Korean (L2) resulted in noticeable assimilatory modification of most English (L1) sounds in native English speakers [18].

While the underlying driving forces of dissimilatory and assimilatory process of L1 sounds as a function of L2 immersion remains unknown, researchers have shown that the magnitude and direction of L1-L2 interaction in bilinguals were highly correlated with the starting age of L2 learning and the amount of L2 experience [19, 20]. However, most existing studies have focused on adults or relatively older child L2 speakers, while relatively few examined young children. The present study, therefore, aims to further our understanding of L1-L2 interaction in phonetic features in young bilingual Mandarin-English children. In particular, both static and dynamic acoustic features in bilingual children were compared with those of age-matched monolingual children.

2. Methods

2.1. Speakers

The speakers included 15 sequential Mandarin-English bilingual children aged 5 to 6 years old, 15 age-matched Mandarin monolingual children and 9 age-matched English monolingual children. The bilingual Mandarin-English children were divided into two groups (Bi-low and Bi-high) based on their proficiency in English. The Bi-high group were born and raised in the US (central Ohio region). They were immersed in a near-monolingual (Mandarin) environment until they went to English daycare or kindergarten at about 3 years of age. By the time of recording, these children had had extensive experience with English for about 3 years. The Bi-low group were born and raised in China who had lived in the U.S. (central Ohio region) for less than 6 months. In terms of parents' dialect background, all bilingual children had at least one parent from northern dialect regions and Mandarin is the

daily-life language used in all bilingual children's families. The 15 age-matched monolingual Mandarin speakers were native Mandarin speakers born and raised in the Beijing area. The 9 monolingual English children were native English speakers born and raised in central Ohio region. All children were reported as having no speech and language disorders.

2.2. Stimuli

The recording material included a list of words containing the shared vowels /a/, /i/ and /u/ in both Mandarin and English. Because there was no direct counterpart in English for Mandarin /a/, two acoustically/phonetically similar vowels /a/ and /æ/ were selected to pair with Mandarin /a/. Specifically, the speech material includes 6 Mandarin disyllabic words containing the vowels /a/ (da4 xiang (elephant), da4 suan (garlic), /i/ (pi2 qiu (ball), bi2 zi (nose) and /u/ (tu4 zi (rabbit), pu2 tao (grape) and 8 English monosyllabic/disyllabic words containing the vowels /æ/ (cat, bat), /a/ (box, stop), /i/ (feet, geese) and /u/ (boot, goose). Each word was repeated twice. Bilingual children produced both Mandarin and English words while monolingual children only produced words in their native language. The selection of both Mandarin and English words was based on word familiarity, word frequency [21, 22] and picturability. However, the consonant environment (phonetic context) was not as strictly controlled. The third tone was excluded for Mandarin words to avoid the longer duration of vowels in syllables with tone 3.

2.3. Procedures

There were two recording sessions for each bilingual subject. Mandarin words were produced in the first session and English words were produced in the second session after a 15-20 minute break. The experimenter (a fluent bilingual Mandarin-English speaker) interacted with the speakers in Mandarin during Mandarin session and English during English session. For each monolingual speaker, only one recording session was conducted in their native language. In each recording session, a visual-auditory word repetition task was used to collect speech samples under control of a custom Matlab program. To have a better control of the stimulus presentation and to ensure the speakers produce the specific target words as expected, this present study used the word-repetition instead of the picture naming task [23]. During the recording period, each speaker was seated facing a laptop computer in a quiet room. Randomly ordered pictures containing target words were presented on the computer screen followed by audio prompts produced by a native adult speaker of each language. The participants were then asked to repeat each word immediately after the prompt. During these sessions, a Shure SM10A head-mounted microphone was situated approximately 1-inch from the subject's mouth. Speech samples were recorded directly onto a hard drive disk with a 16-bit quantization rate and 44.1 kHz sampling rate.

2.4. Acoustic measurements

2.4.1. Formant frequencies

All tokens were first down-sampled to 11.25 kHz to enable better visual examination of formant frequency values following spectrographic analysis. The program TF32 [24] was used to extract the frequency values of F1 and F2 at five equidistant temporal locations (20-35-50-65-80% point) in order to capture the dynamic spectral change of the vowel movement [25, 26]. The landmark locations of vowel onset

and offset were located by hand and determined primarily on the basis of the waveform, accompanying with visual check of the spectrogram [27]. Since the speakers across all three groups were similar in age, the effect of different vocal tract lengths on the formant frequencies was expected to be small. Thus, unnormalized formant frequency values were used for further calculation and analysis.

2.4.2. Trajectory length

Trajectory length (TL) is defined as the sum of the Euclidean distances (in Hz) between each two consecutive temporal points, (i.e. 20-35%, 35-50%, 50-65%, 65-80%). This measure provides an unsigned measure of the magnitude of vowel movement in the F1 x F2 acoustic plane over the course of vowel duration between the 20 to 80% points. It is calculated following formula [25] in which n stands for the vowel section between each two consecutive points:

$$TL = \sum_{n=1}^4 VSL_n \quad (1)$$

where the length of each vowel section (VSL) is calculated based on the formula:

$$VSL_n = \sqrt{(F1_{n+1} - F1_n)^2 + (F2_{n+1} - F2_n)^2} \quad (2)$$

3. Results

3.1. Midpoint F1 by F2 vowel space

The means and standard deviations of midpoint formant frequency values of the shared vowels in each group are shown in Figure 1. For the monolingual children, Mandarin (M) /i/ and English (E) /i/ were close to each other while M/u/ and E/u/ showed a great positional separation in the acoustic space. For the vowel pair of M/a/ and E/a/ and E/æ/, M/a/ was located in a center and lower position and separated from the two English counterparts. Obviously, we expect formant value differences as a function of vowel quality differences but this is not of particular interest in the present study. One-way ANOVA was conducted on each of the midpoint formants (F1 and F2) for each vowel pair. The results showed no significant difference between M/i/ and E/i/ but a significant higher F2 in E/u/ than M/u/ ($p < 0.001$). For the vowels of M/a/ and E/a/ and /æ/, Tukey HSD showed a significantly higher F1 in M/a/ than E/a/. It also showed significant differences in F2 for each between M/a/ and E/æ/, M/a/ and E/a/ and E/a/ and E/æ/.

For the Bi-low children, similar to monolingual children, M/i/ was close to E/i/ and showed no significant difference. E/u/ showed a significant higher F2 than M/u/ ($p < 0.001$). However, compared to monolingual children, E/u/ was located at a relative back position and was closer to M/u/ (the Euclidean distance between M/u/ and E/u/ was 616Hz while that in monolingual children was 1115Hz). The locations of M/a/ and E/a/ and E/æ/ showed similar pattern as that of monolingual children. The ANOVA tests also returned significant difference for both F1 ($p < 0.001$) and F2 ($p < 0.001$). However, the pairwise comparisons of F2 showed no difference between M/a/ and E/a/, which are different from monolingual children.

For the Bi-high children, M/i/ and E/i/ were highly overlapped and no significant differences were found. Similar to monolingual children, Bi-high children also showed a large and significant separation ($p < 0.001$ for F2) between M/u/ and

E/u/ in the acoustic space. However, for the M/a/ and E/a/ and E/æ/, it can be observed that M/a/ was considerably raised and located at a position with similar height as E/a/ and E/æ/. The pairwise comparison found no significant difference in F1 between M/a/ and E/æ/, nor between M/a/ and E/a/.

In general, the statistical analyses on the midpoint formant frequency values showed no difference between M/i/ and E/i/ for any group of children. It showed a significantly higher F2 in E/u/ than M/u/ across all three groups of children. However, for M/a/ and E/a/ and E/æ/, the statistical analyses revealed inconsistency among these groups of children. Both Bi-low and Bi-high children showed some differences from monolingual children on the structure of these three vowels.

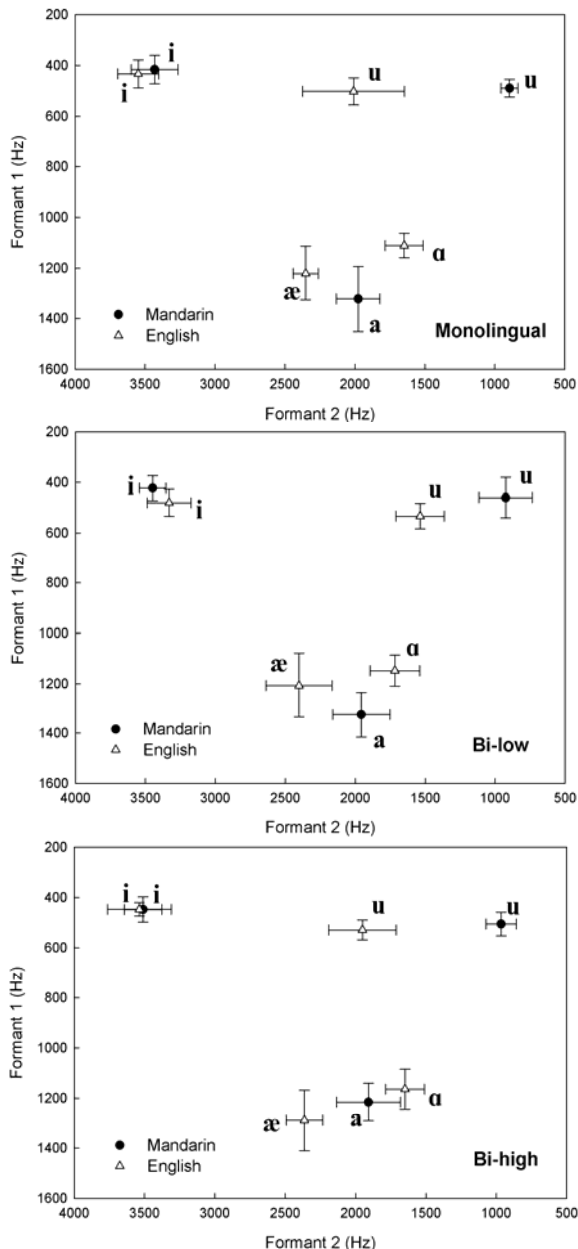


Figure 1: Means and standard deviations of midpoint formant frequency values for Mandarin and English shared vowels in each group of children.

3.2. Vowel dynamic spectral change

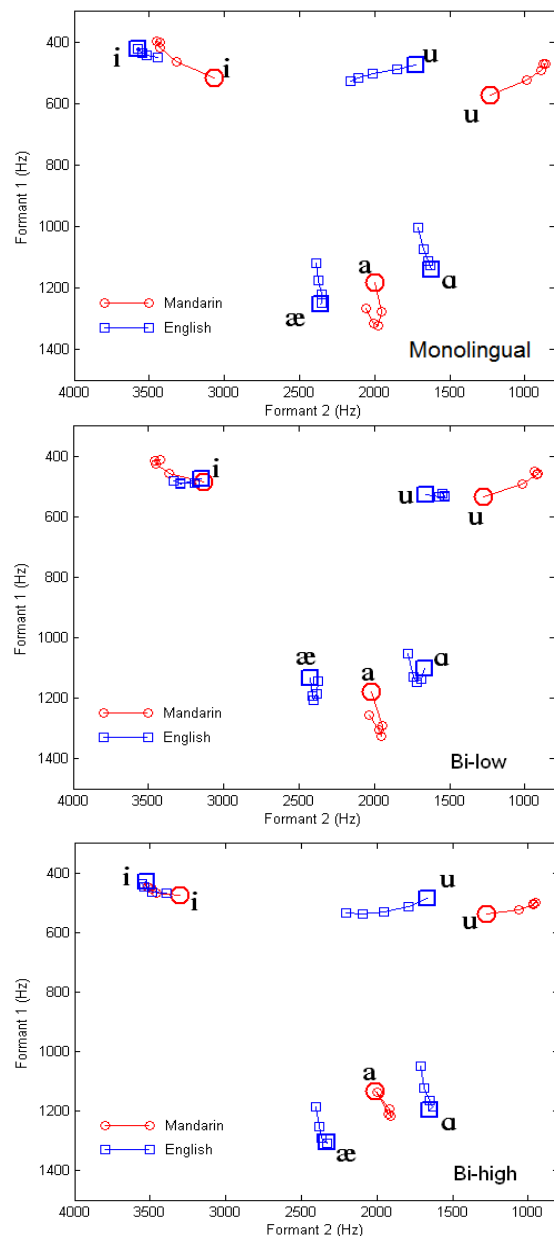


Figure 2: Vowel spectral change plotted on the basis of formant frequencies at five time locations (20-35-50-65-80%) over the course of vowel duration for shared vowels in each group. The larger size symbol represents the 80% point.

Figure 2 shows the formant trajectories (in the F1 x F2 plane) for these shared vowels in each group. These formant trajectories provide detailed information about the nature of vowel dynamic spectral change, which is not characterized in midpoint vowel dispersion patterns. For monolingual children, all three Mandarin vowels displayed opposite directions of formant movement than their English counterparts. For example, the trajectory of M/i/ moved to a relatively low and back position while the trajectory of E/i/ moved to a relatively high and front position from the 20 to 80% point. In addition, the magnitude of formant movement of M/i/ was larger than that of E/i/. These differences in the vowel spectral change of this vowel pair did not show in their midpoint formant

frequency values. In this case, although these two /i/s showed similar relative positions in the acoustic space, they were still different in the pattern of formant movement.

For the Bi-low children, while they maintained the formant dynamic pattern of their Mandarin (native) vowels, the directions of formant movement of their English vowels were quite different from those of monolingual children. In particular, the formant trajectories of their English vowels moved in the same direction as their Mandarin counterparts. This finding suggests that these children were assimilating the vowel spectral change of their native language to the new vowels in the second language even though some vowels are very similar in the relative positions such as M/i/ and E/i/.

For the Bi-high children, the direction of formant trajectories in both Mandarin and English vowels conformed to those of monolingual children. However, their Mandarin vowels still showed differences from monolingual children in either the magnitude of the trajectories or the relative positions of the vowels. Specifically, their M/i/ showed much less formant movement than that in monolingual children. These findings indicate that Bi-high children's Mandarin vowels were affected by their English and thus showed different formant movement pattern from monolingual children.

3.3. Trajectory length

Table 2 shows the means (and standard errors) of the trajectory lengths for each of the vowels for all four groups of speakers. Several points are important to note. The mean TLs of all four English vowels for the Bi-high group are closer to those of the monolingual English group than are the TLs for the Bi-low speakers. This shows assimilation of L1 to L2 for the Bi-high group. Similarly, the TLs of two Mandarin vowels, M/i/ and M/a/ for the Bi-low speakers are closer to the TLs of the monolingual Mandarin speakers than are those for the Bi-high group--this demonstrates an effect of L2 on L1 by the Bi-high speakers. One-way ANOVAs were used to examine the difference among monolingual and two bilingual groups for each shared vowel. For the three Mandarin vowels, the results revealed a significant difference among monolingual and two bilingual groups on M/i/ ($p=0.002$). In particular, Bi-high speakers produced significantly shorter TL than did monolingual Mandarin and Bi-low children. For the four English vowels, the results revealed a significant difference among monolingual English and two bilingual groups on the vowel /a/ ($p=0.007$), /i/ ($p<0.001$), and /æ/ ($p=0.022$). In particular, Bi-low children produced significantly longer TLs for these three vowels than monolingual English and Bi-high children. In future studies, the effect of vowel quality on trajectory lengths in each group will be examined.

Table 2. Means and standard errors of trajectory lengths for Mandarin and English shared vowels in each group of children (in Hz).

Vowel	Monolingual English	Monolingual Mandarin	Bi-low	Bi-high
E/i/	254.4 (85)		498.8 (52)	300.4 (33)
E/u/	517.0 (108)		325.1 (37)	595.3 (72)
E/æ/	247.1 (19)		352.3 (31)	274.4 (25)
E/a/	240.5 (10)		317.9 (17)	294.5 (21)
M/i/		514.0 (27)	447.0 (28)	329.8 (45)
M/u/		493.5 (26)	510.2 (40)	421.8 (46)
M/a/		412.3 (31)	397.7 (12)	342.1 (19)

4. Summary and discussion

This study examined the acoustic properties of shared vowels produced by relatively young bilingual Mandarin-English children and corresponding monolingual children. Of particular interest was how L1-L2 interactions affect the phonetic characteristics in bilingual children. The results showed an active bi-directional interaction in young bilingual children. In particular, at the early stage of L2 acquisition (as shown in Bi-low children), bilingual children maintained their L1 acoustic features and transferred their L1 features to the new phonetic system. Once they have established the new sound system (as shown in Bi-high children), the bilingual children modified their native sounds due to the influence of the new language.

As shown in the vowel scatter plots of midpoint formant frequency values, both Bi-low and Bi-high children showed different structures from monolingual children. Monolingual English and monolingual Mandarin children produced similar /i/ but different /u/ and /a/ in the acoustic space. Compared to monolingual children, Bi-low children's English /u/ was located at a further back position and moved closer to Mandarin /u/. This finding evidenced the assimilatory process. That is, Bi-low children assimilated their L2 to its phonetically similar L1 sound. In contrast, Bi-high children produced their English vowels in a monolingual-like manner; however, their production of Mandarin vowels was different from the monolingual children of that language. In particular, they raised Mandarin /a/ to a higher position, which caused the Mandarin /a/ become closer to English /a/ and /æ/. This again, evidenced the assimilation process of L1 vowels to the similar L2 vowel categories.

The assimilatory process for both L1-L2 interaction mechanisms (the effect of L1 on L2 and the effect of L2 on L1) can be further observed in the vowel dynamic spectral patterns. Although Mandarin /i/ and English /i/ showed no positional difference in the midpoint static features, they presented distinctive vowel dynamic features in terms of both the direction and magnitude of formant movement. In Bi-low children, the formant movement direction of English /i/ and English /u/ was consistent with their Mandarin counterparts while opposite to those of monolingual English children. This provides further evidence of the assimilatory influence of L1 on L2 (as do the TL difference between Bi-low and Bi-high speakers for Mandarin vowels). In Bi-high children, although the directions of formant trajectory of the two /i/s were consistent with those of mono children, the magnitude of formant movement of the Mandarin /i/ was much shorter than that of monolingual Mandarin children and similar to that of English /i/. This also provides additional evidence to demonstrate the assimilatory influence of L2 to L1. In addition, these findings also demonstrated the importance of the vowel dynamics in understanding the nature of vowel acoustic features.

In sum, the detailed comparison of acoustic vowel characteristics between bilingual and monolingual children revealed that young children do demonstrate bi-directional assimilatory process of L1-L2 interaction. However, the strength and size of the bi-directional effect are determined by different stage of L2 learning in the bilingual children.

5. References

- [1] Grosjean, F. "Neurolinguists, beware! The bilingual is not two monolinguals in one person", *Brain and Language*, 36, 3-15, 1989.
- [2] Keshavarz, M., and Ingram, D. "The early phonological development of a Farsi-English bilingual child", *International Journal of Bilingualism*, 6, 255-269, 2002.
- [3] Paradis, J., and Genesee, F. "Syntactic acquisition in bilingual children: Autonomous or interdependent?" *Studies in Second Language Acquisition*, 18, 1-25, 1996.
- [4] Paradis, J. "Do bilingual two-year-olds have separate phonological systems?" *International Journal of Bilingualism*, 5, 19-38, 2001.
- [5] Flege, J. E. "Second language speech learning: Theory, findings and problems", In W. Strange [Eds.], *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-273). Timonium, MD: York Press, 1995.
- [6] Watson, I. "Phonological processing in two languages", In E. Bialystock [Ed.], *Language processing in bilingual children* (pp. 25-48). Cambridge, United Kingdom: Cambridge University Press, 1991.
- [7] Best, Catherine T. "A direct realist view of cross-language speech perception" In Winifred Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, (pp.171 - 204). Timonium, MD: York Press. 1995.
- [8] Caramazza, A., Yeni-Komshian, G. H. Zurif, E. B., and Carbone, E. "The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals", *Journal of the Acoustical Society of America*, 54, 421-428, 1973.
- [9] Hazan, V. L., and Boulakia, G. "Perception and production of a voicing contrast by French-English bilinguals", *Language and Speech*, 36, 17-38, 1993.
- [10] Flege, J. and Hillenbrand, J. "Limits on pronunciation accuracy in adult foreign language speech production", *Journal of the Acoustical Society of America*, 76, 708-721, 1984.
- [11] Flege, J. E. "The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification", *Journal of Phonetics*, 15(1), 47-65, 1987.
- [12] Harada, T. "L2 influence on L1 speech in the production of VOT", In: M.-J. Sole, D. Recasens, and J. Romero [Eds.], *Proceedings of the 15th international congress of phonetic sciences* (pp.1085-1088). Barcelona, Spain: Causal Productions, 2003.
- [13] Kang, K.-H., and Guion, S. G. "Phonological systems in bilinguals: Age of learning effects on the stop consonant systems of Korean-English bilinguals", *Journal of the Acoustical Society of America*, 119(3), 1672-1683, 2006.
- [14] Mack, M. "Phonetic transfer in a French-English bilingual child", In: P. H. Nelde [Ed.], *Language attitudes and language conflict* (pp. 107-124). Bonn, Germany: Dümmler, 1990.
- [15] Bohn, O.-S., and Flege, J. "The production of new and similar vowels by adult German learners of English", *Studies in Second Language Acquisition*, 14, 131-158, 1992.
- [16] Flege, J., Schirru, C., and MacKay, I. "Interaction between the native and second language phonetic subsystems", *Speech Communication*, 40, 467-491, 2003.
- [17] Guion, S. "The vowel systems of Quichua-Spanish bilinguals: An investigation into age of acquisition effects on the mutual influence of the first and second languages", *Phonetica*, 60, 98-128, 2003.
- [18] Chang, C. B. "Rapid and multifaceted effects of second-language learning on first-language speech production", *Journal of Phonetics*, 40, 249-268, 2012.
- [19] Baker, W., Trofimovich, P., Flege, J., Mack, M., and Halter, R. "Child-adult differences in second-language phonological learning: The role of cross-language similarity", *Language & Speech*, 51, 316-341, 2008.
- [20] Oh, G. E., Guion-Anderson, S., Aoyama, K., Flege, J. E., Akahane-Yamada, R., and Yamada, T. "A one-year longitudinal study of English and Japanese vowel production by Japanese adults and children in an English-speaking setting", *Journal of Phonetics*, 39, 156-167, 2011.
- [21] Thorndike, E.L. and Lorge, I. "The teacher's word book of 30,000 words", New York: Bureau of Publications, Teachers college, Columbia University, 1963.
- [22] Liu, S., Han, D., Zhang, N., Wu, X., Sheng, Y., Mo, L., Yang, Y., and Kong, Y. "The development of database of everyday speech material for pre-school Mandarin children", *Journal of Audiology and Speech Pathology*, 16 (2), 2008.
- [23] Edwards, J. and Beckman, M. E., "Methodological questions in studying consonant acquisition", *Clinical Linguistics and Phonetics*, 22(12):939-958, 2008.
- [24] Milenkovic, P., TF32 software program. University of Wisconsin, Madison, 2003.
- [25] Fox, R. A. and Jacewicz, E. "Cross-dialectal variation in formant dynamics of American English vowels", *Journal of the Acoustical Society of America*, 126, 2603-2618, 2009.
- [26] Jacewicz, E., Fox, R. A., and Salmons, J. "Regional dialect variation in the vowel systems of typically developing children", *Journal of Speech, Language and Hearing Research*, 54, 448-470, 2011.
- [27] Kent, R. D. and Read, C., *The acoustic analysis of speech*, San Diego, CA: Singular Publishing, 1992.