P07

F0, long-term formants and LTAS in Korean-English Bilinguals

Sylvia Cho, Murray J. Munro (Simon Fraser University)
sylvia_cho@sfu.ca, mjmunro@sfu.ca

1. Introduction

Voice quality measures such as speaking fundamental frequency (SFF), long-term formants (LTF), and long-term average speech spectra (LTAS) have been useful for speaker identification of monolingual speech samples (Nolan & Grigoras, 2005). However, the existing studies on SFF, LTFs, and LTAS either lack bilingual data or have reported obscure or confusing results. It is thus largely unknown whether the acoustic properties of speech samples from a bilingual speaker differ according to the language being spoken. Such comparisons have important implications for the practice of forensic speaker identification, as bilinguals may or may not behave like their monolingual counterparts in their productions. The current study therefore aims to contribute to this line of work by (1) extending SFF comparisons to Korean-English bilinguals, (2) conducting a comparison of bilingual LTFs that employs LTF histograms (as well as mean values), and (3) confirming whether LTAS differences between monolingual speakers of Korean and English are also found in Korean-English bilinguals.

2. Production Experiment

2.1. Methods

2.1.1 Participants

Ten Korean-English bilingual speakers (5M 5F) were recruited at Simon Fraser University. All subjects spoke standard Seoul Korean and Western Canadian English. Their age ranged from 18 to 36 (Mean age: 25), and their age of arrival (AOA) in Canada ranged from 0 to 12 years (Mean AOA: 5.8). Each subject was screened prior to the recording to ensure that they met the following criteria: (1) they were judged to be proficient Korean-English bilinguals by one of the authors who is proficient in both languages (2) they did not report any hearing or speech impairments; (3) they had no signs of sickness from colds or other respiratory problems on the day of the recording.

2.1.2 Procedures

A total of four speech samples were taken from each subject. The subjects were instructed to read a passage out loud (i.e., The North Wind and the Sun), at a comfortable pace in both Korean and English. They were also asked to produce extemporaneous narratives in both languages, in which they were instructed to describe a picture story (Derwing, Munro, Thomson, & Rossiter, 2009).

3. Results and Discussion

3.1. Speaking Fundamental Frequency (SFF)

Speaking fundamental frequency (SFF) was measured using the autocorrelation method in Praat (Boersma & Weenik, 2016). Paired-sample *t*-tests[1] were then carried out to assess the existence of a language effect on SFF. These results are shown in Figures 1 and 2.

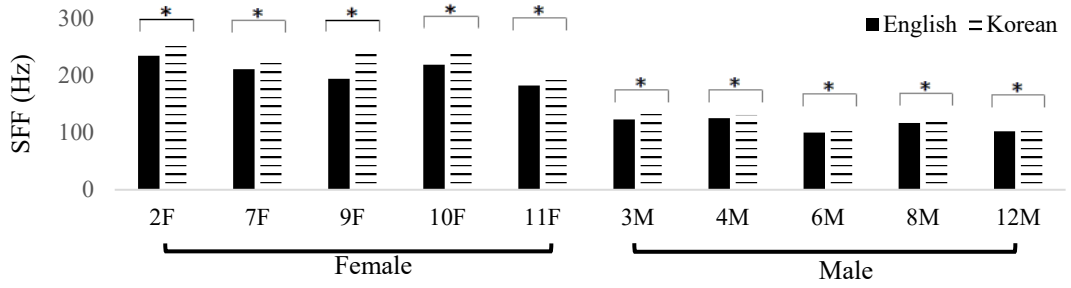

Figure 1: Average SFF (in Hz) for passage reading in Korean and English

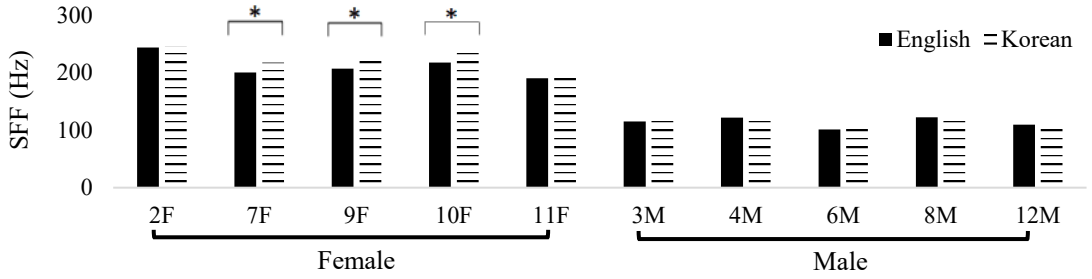

Figure 2: Average SFF (in Hz) for extemporaneous speech in Korean and English (* $p < 0.05$)

As displayed in Figures 1 and 2, significant SFF differences were found between English and Korean in the passage reading, but this language effect was not as robust in the extemporaneous data (i.e., only 3 out of 10 speakers adjust their SFF significantly). With the passage reading, there was a mean F0 difference of 15 to 45 Hz between the languages for the female subjects, and for the male subjects, there was a difference of 5 to 20 Hz. In the extemporaneous speech task, there was a maximum of 15 Hz difference in the females, and up to a 5 Hz difference in the males. This indicates an effect of speaking style on SFF, and that casual or spontaneous speech is produced with smaller between-language differences. In terms of the direction of the SFF change, there was either no statistical difference between the bilinguals' production of Korean and English, or English was produced with a lower SFF. Regarding this observed pattern, it is possible that subjects produced English with a lower

[1] Each speech sample was parsed into 20 stretches of equal duration. The mean F0 value for each stretch was obtained, and these values (i.e., 20 data points for Korean; 20 data points for English) were used in the paired-sample *t*-tests.

SFF due to a language-dependent tendency. That is, English has been associated with a relatively low SFF, compared to that of other languages (Hanley & Snidecor, 1967). This tendency seems to be more readily seen in read speech, as opposed to extemporaneous speech, which better approximates natural speaking conditions.

3.2. Long-term Formants (LTF)

LTFs were computed for speech samples of equal duration (i.e., first 30 seconds of speech sample) using formant frequencies chosen by the automatic formant tracker in Praat, which uses linear predictive coding. In terms of mean formant values, there were very small within-subject differences. This study finds that the shapes of the formant frequency distributions are more indicative of speaker identity than are the means. For example, the histograms in Figures 3 and 4 display the distributions of formant value ranges (displayed in 100 Hz bins) for two speakers that have very similar mean formant values (i.e., F1: 400 Hz, F2: 16-1800Hz, F3: 2800Hz). While their average formant frequencies are similar, Subject 4M, has a more prominent F1 peak, and his F3 peak is skewed more to the left.

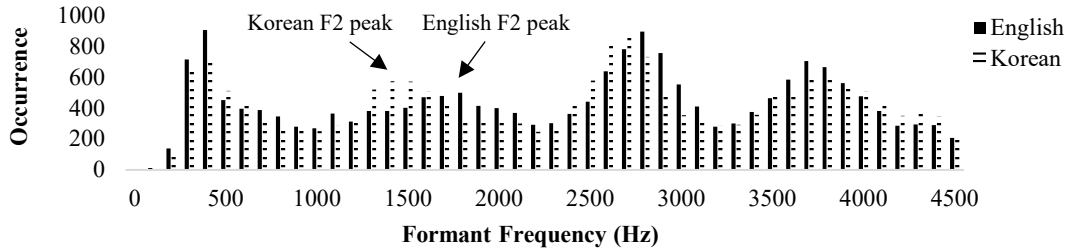

Figure 3: Read Speech LTF for Subject 3M

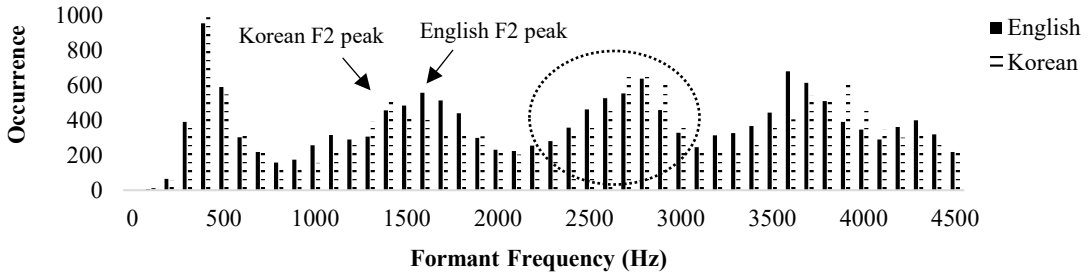

Figure 4: Read Speech LTF for Subject 4M

As shown in the histograms above, there is also a small effect of language across bilingual LTFs. That is, Korean productions are often associated with a lower F2 peak. This is consistent with acoustic studies that report slightly lower F2 values for Korean vowels in comparison to English vowels (Yang,

1996). A strong effect of speech style was not found with bilingual LTFs. Altogether, the current study finds that LTFs are mostly speaker-specific with small language-specific effects.

3.3. Long-term Average Speech Spectra (LTAS)

There appear to be no existing studies on bilingual LTAS, but a study that compared LTAS of twelve languages[2] found that most of the variations among languages are smaller than 3 dB in magnitude and occur within the range of 200 to 6300 Hz (Bryne et al., 1994). It was thus suggested that a "universal" LTAS can be applied across languages. A more recent study (Noh & Lee, 2012) compared Korean LTAS to that of English, and found that the LTAS of Korean speakers show significantly lower intensity levels in frequencies above 2000 Hz, except at 4000 Hz. The lower intensity levels found in Korean are thought to stem from lower occurrence of phonemes with high frequency components (i.e., fewer fricatives in Korean). This suggests that for languages such as Korean, the universal LTAS is rendered inappropriate. The current study examines whether the trends found for monolingual English and Korean speakers also hold true in bilingual speakers.

LTAS were computed using Praat to observe overall spectral properties of the continuous speech samples. Each speaker's samples were then normalized for peak intensity (dB) to permit straightforward comparisons. This study finds that most speech spectra show maximum intensity around 500 Hz, and a decrease is seen in the higher frequencies. Similar to the monolinguals in Noh and Lee's study, many bilingual speakers (6 out of 10) showed significantly lower intensity levels in frequencies above 2000 Hz (except at 4000 Hz) in their production of Korean (see Figure 5). There was a 3 to 5 dB difference from 2000 to 4000 Hz, and a difference of 5 to 10 dB in the higher frequencies.

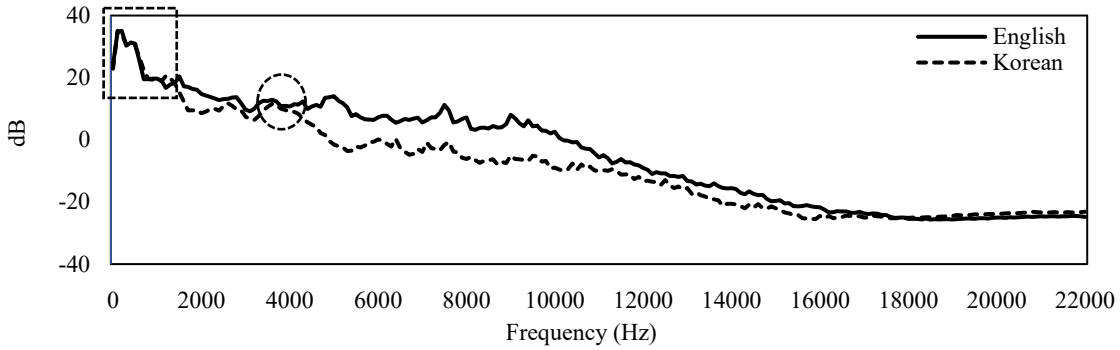

Figure 5: Extemporaneous Speech LTAS for Subject 4M

It should be noted that this pattern was not observed across all speakers. Some bilinguals (4 out of 10) had very similar intensity levels in the two languages up to 4000 Hz (see Figure 6). This suggests that

[2] English (several dialects), Swedish, Danish, German, French, Japanese, Cantonese, Mandarin, Russian, Welsh, Singhalese, Vietnamese

there are some within-speaker characteristics that influence bilingual LTAS – which is different from what is found across monolingual speakers (i.e., between speakers). The reasons for such a distribution are not clear at the moment, but it seems that the bilingual speakers that were impressionistically more dominant in one language (i.e., better at either Korean or English) demonstrated such patterns.

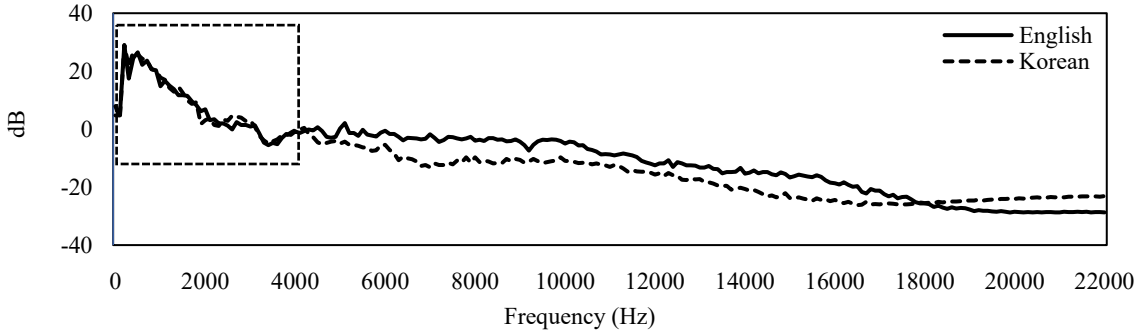

Figure 6: Extemporaneous Speech LTAS for Subject 2F

Furthermore, this study found nearly identical LTAS for bilingual speakers that had clearly lower proficiency in one of the languages. Figure 7 shows the LTAS of a speaker that was omitted from the scope of this study due to his lack of native-like proficiency in English (late-onset bilingual, AOA: 18 years old).

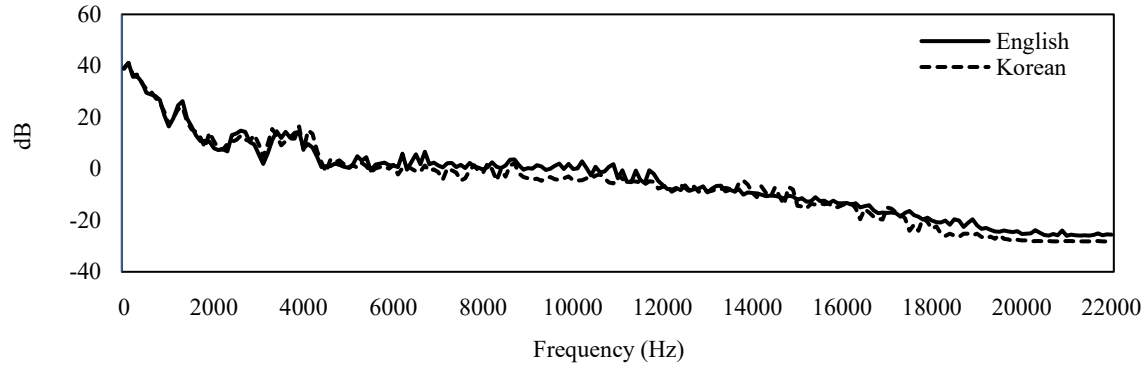

Figure 7: Extemporaneous Speech LTAS for Subject 5M

As can be observed from the figure above, energy differences seen across other Korean-English bilinguals were not found for Subject 5M, namely a subject who is not very proficient in one of the languages. It is plausible that such speakers produce the less-proficient language in much the same manner as their native language. Such findings suggest that LTAS analysis can be somewhat useful in speaker identification cases involving multilingual samples, given that the speaker lacks proficiency in one of the languages spoken. Lastly, there was little indication of an effect of speech style on LTAS distributions.

4. Conclusion

This study investigated the usefulness of SFF, LTF, and LTAS analyses in bilingual speaker identification. Among the different acoustic measurements, formant frequency estimates across long speech samples were found to be most reliable and speaker-specific. Speaking style was also found to be important; stronger SFF correlations were found between languages in extemporaneous speech. Such results suggest that multilingual samples may serve some useful purpose in speaker identification cases that involve bilingual speakers. Future research could further expand on the results obtained from this project by applying the methods to other bilingual populations, such as Japanese-English bilinguals. One previous study finds that SFF is consistently higher in Japanese for Japanese-English bilinguals (Graham, 2015), and it might be the case that, like Korean-English bilinguals, there is some effect of language in the bilingual LTFs given the small vowel inventory of Japanese.

References

Boersma, P., & Weenink, D. (2016). Praat: doing phonetics by computer (Version 6.0.23) [Computer program]. Retrieved from http://www.praat.org/

Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wilbraham, K., Cox. R., ... Ludvigsen, R. (1994). An international comparison of long-term average speech spectra. T*he Journal of the Acoustical Society of America, 96*(4), 2108-2120.

Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, *31(4)*, 533-557.

Graham, C. (2015). Fundamental Frequency Range in Japanese and English: The Case of Simultaneous Bilinguals. *Phonetica, 71*(4), 271-295.

Hanley, T. D., & Snidecor, J. C. (1967). Some acoustic similarities among languages. *Phonetica, 17*, 141-148.

Heeren, W., Van der Vloed, D., & Vermeulen, J. (2014). Exploring long-term formants in bilingual speakers. *Proceedings from the 2014 Conference of the International Association for Forensic Phonetics and Acoustics*. Zurich, Switzerland.

Krebs, P., & Braun, A. (2015). Long Term Formant measurements in bilingual speakers. *Proceedings from the 2015 Conference of the International Association for Forensic Phonetics and Acoustics*. Leiden, Netherlands.

Noh, H., & Lee, D. (2012). Cross-Language Identification of Long-Term Average Speech Spectra in Korean and English. *Ear and Hearing, 33*(3), 441-443.

Nolan, F., & Grigoras, C. (2005). A case for formant analysis in forensic speaker identification. *International Journal of Speech, Language and the Law, 12*(2), 143-173.

Yang, B. (1996). A comparative study of American English and Korean vowels produced by male and female speakers. *Journal of Phonetics, 24*(2), 245-261.