JASA EXPRESS LETTERS

https://doi.org/10.1121/10.0000498

Cross-linguistic *f*0 differences in bilingual speakers of English and Korean

Andrew Cheng[a)]
University of California, Berkeley, Berkeley, California 94720, USA
andrewcheng@berkeley.edu

Abstract: Languages may differ in fundamental frequency of voicing (*f*0), even when they are spoken by a bilingual individual. However, little is known in bilingual/L2 acquisition research about simultaneous bilinguals. With the expectation that speakers who acquired two languages early use *f*0 differently for each language, this study measured *f*0 in English–Korean early bilinguals' natural speech. The *f*0 level was higher for Korean than English, regardless of gender, age, or generational status (early and late bilinguals did not differ). The *f*0 span showed a language-gender interaction: males' span was larger in Korean, while females' span was larger in English. This study demonstrates that languages differ in *f*0 independent of speaker anatomy and suggests that children may acquire these differences in early childhood.

[Editor: Martin Cooke] Pages: EL67–EL73

Received: 29 August 2019 **Accepted:** 19 December 2019 **Published Online:** 3 February 2020

1. Introduction

Studies show that some languages fundamentally differ in average vocal *f*0 measurements, even when speaker differences are accounted for. Some common measurements include *f*0 level, which is similar to sustained average *f*0, and *f*0 span, which is the range between the high and low ends of a speaker's *f*0 range. For example, Mennen *et al.* (2012) compared female speakers of German and English and found that the English speakers had a higher *f*0 level and a wider *f*0 span.

However, in a comparison of English and Mandarin, Keating and Kuo (2012) found virtually similar *f*0 ranges for both groups of speakers, while other aspects of the *f*0 profile (including maximum, minimum, range, standard deviation, and multiple means) differed only unsubstantially. The studies above compare similar populations of monolingual speakers. However, a more sensitive measure might emerge from an investigation of how bilingual speakers produce their two languages. Bilingual speech controls for speaker-inherent differences (i.e., neutralizing the effect of anatomical differences on fundamental frequency) and places the languages in contrast with one another in the experience of the speaker. For example, Altenberg and Ferrand (2006) found that Russian–English bilingual women spoke Russian with a higher mean *f*0 than English, but Cantonese–English bilingual women showed no significant difference between languages, and all three groups were comparable across languages.

Graham (2014) analyzed Japanese-English bilinguals, measuring *f*0 at relevant points in a standard Japanese or English utterance, given the known patterns of each language's intonational prosody, and found statistically significant cross-language differences: Japanese has a higher level and a wider span than English. Graham (2014) concludes that both sociophonetic and phonological factors (for *f*0 span) must be at play here, allowing that bilingual speakers of different genders may speak their languages differently. Indeed, a recent study of bilingual range, Ordin and Mennen (2017), found that female speakers of Welsh and English had systematically wider *f*0 ranges in Welsh compared to English, although male speakers showed no regular patterning.

Finally, Lee and Van Lancker Sidtis (2017) examined Mandarin–English and Korean–English bilingual speakers performing a variety of speech tasks: reading a passage, describing pictures, and giving a spontaneous monologue. They found a higher *f*0 level in Korean compared to English across all tasks, and higher *f*0 variability in Korean compared to English in the monologue. This study's participants were all female speakers (mean age: 25 years) who had immigrated from South Korea to the United States at variable times [mean AOA (age of arrival): 13 ± 7 years] and had been living in the United States for an average of 10.5 years

[a)]Author to whom correspondence should be addressed.

(±4). This means that in terms of bilingual acquisition, the participants were most likely all sequential bilinguals, or speakers who had fully acquired proficient Korean prior to immersion in an English-dominant environment.

Simultaneous bilinguals are speakers who acquire two languages at approximately the same time, either from birth or prior to the age of 3 (Paradis, 2007). For many second generation Korean Americans, their language input from early childhood onward is a mix of Korean and English, and they come to be proficient in both languages natively, rather than being a native speaker of one language and then acquiring another. A comparison of sequential bilinguals to simultaneous bilinguals would reveal whether the kind of early childhood input given to simultaneous bilinguals affects their use of vocal *f*0 in a way that the input of speakers who acquire one language much later than the other does not. Studies have shown the effects of age of acquisition on L1 and L2 vowel production (Baker and Trofimovich, 2005) and native-like consonant VOT (Voice Onset Time) or the development of separate phonological categories for consonants (Kang and Guion, 2006; Lee and Iverson, 2012), for example.

The current study tests vocal fundamental frequency (*f*0) level and span of both languages of English–Korean bilingual speakers.[1] Building on the findings of Lee and Van Lancker Sidtis (2017), the expectation is that the two languages will differ in the direction of Korean having a higher level and a wider span, even when speakers are engaging in natural conversational speech, instead of participating in facilitated speech tasks.

2. Methods

2.1 Sociolinguistic interview

Subjects recruited for the study participated in a bilingual sociolinguistic interview. The structure of the bilingual interview was as follows: interviewees were asked to give a short self-introduction in Korean, which was followed by a casual interview conducted in Korean, centered on the interviewee's family, life, and hobbies. Then, the interviewee was asked to read several sentences in written Korean (designed for an accompanying study). Next, the interviewee silently read and signed some documents written in English, which served as a buffer between interview sections and languages. Finally, the interviewer began a second interview conducted in English, centered on the interviewee's opinions and experiences regarding Korean language, culture, and identity.

The Korean portion of the interview generally lasted from 5 to 15 min. The Korean reading sections generally lasted about 5 min, but were excluded from the present analysis, so as to only compare spontaneous speech in both languages. The English portion of the interview lasted from 20 to 45 min. The total interview time was 1 h. Interviewees were not prohibited from code-switching, so interviewees sometimes uttered English words during the Korean portion, and vice versa.

2.2 Subjects

The results from 15 interviewees are reported. They all identified as Korean American. Ten identified as cisgender female, and five as cisgender male. The age range was 18–29 years old (mean age: 21.1 ± 3.1).

The Korean American community, a "first generation" immigrant is a Korean national who immigrated to the United States as an adult and became a naturalized citizen; "second generation" Korean Americans are their children, born and raised in the United States. In addition, "1.5 generation" Korean Americans were born in Korea but immigrated with their families to the United States at a young age (Park, 1999). Of the 15 subjects, eight identified as second generation (mean age: 21.75 ± 3.3 years), and seven as 1.5 generation (mean age: 20.3 ± 2.8 years; mean age of arrival: 7.8 ± 2.8 years). All of the 1.5 generation interviewees were born in Seoul or the Seoul metropolitan area, with the exception of one interviewee, a female, who was born in Busan. Every interviewee had at least one parent who was born and raised in Seoul. Basic demographic information about the 15 interviewees is provided in Table 1.

All the interviewees involved in this study were able to speak proficient conversational Korean. Two independent raters were asked to score a sample of speech from each subject's Korean interview on two 5-point Likert Scales for strength of accent in Korean and level of proficiency in Korean. Only the speakers who scored at a 3 or above on both scales were included in the analysis.[2]

In terms of language acquisition, all of the 1.5 generation interviewees reported learning Korean first, then acquiring English upon immigrating to the United States. Thus, they could also be characterized as early sequential bilinguals. (Note, however, that most Korean children are exposed to minimal English even in South Korea, as it is a compulsory subject in the country's public education system.) Second generation interviewees were simultaneous bilinguals, as most reported learning Korean at home as their first language, then acquiring English either at

https://doi.org/10.1121/10.0000498

home, in their neighborhoods, or at school once they reached schooling age. Overall, the crucial difference between this population and the population studied in Lee and Van Lancker Sidtis (2017) is that the current study's subjects acquired English at a much earlier age on average, and can be split into two categories of bilingual acquisition: simultaneous bilingualism (all of the second generation interviewees) or early sequential bilingualism (the 1.5 generation interviewees).

2.3 Interview transcription and data analysis

The interviewers were a team of four trained undergraduate research assistants (RAs) , as well as the author, who are all fluent in Korean. Two of the RAs are Korean American (one 1.5 generation, one second generation); the other two, and the author, are of Asian descent, but not Korean. The phonetic data from the interviewers' speech were also recorded during the interviews, but has not been analyzed. The interviews all took place in a specially outfitted laboratory designed to resemble a living room, to ensure interviewee comfort and minimize some effects of "laboratory speech," while also contributing to "stylistic diversity" in speech research, as in Wagner *et al.* (2015). Interviewers and interviewees wore lapel microphones and their speech was digitally recorded onto a computer.

By-hand transcriptions were force-aligned to the audio recording using two modified versions of the Penn Forced Aligner (Yuan and Liberman, 2008), one for English and one for Korean. Formant and *f*0 data were extracted using the Inverse Filter Control method (Ueda *et al.*, 2007). *F*0 measurements taken every 5 ms were averaged across the duration of each vowel. Then, the average *f*0 was calculated per word, based on the average *f*0 of the stressed vowel of an English word, and the average *f*0 of every vowel of a Korean word (*f*0 level, per Mennen *et al.*, 2012). Then, all Hertz measurements were converted to semitones with a base of 100 Hz.

To determine the *f*0 span, the coefficient of variance (ratio of standard deviation to mean *f*0 in Hertz, then converted to semitones) as well as four range measurements were calculated: IQR (or the middle 50% of the *f*0 measurements per speaker), 80%, 90%, and 98% ranges. Then, using percentile calculation functions, the middle 80%, the middle 90%, and the middle 98% ranges were calculated ("rg80," "rg90," "rg98"). A greater value for any of these ranges indicates a wider *f*0 span.

3. Results

3.1 Mean f0

Figure 1 demonstrates the results for one interviewee, Subject 20. It is clear that while *f*0 varies widely over the duration of the interview, the Korean and English sections of the interview also differ. The orange dots, representing *f*0 measurements from Korean words, cluster at a higher *f*0 value in semitones, compared to the green dots, which represent *f*0 measurements from English words. (Code-switched words, such as English words occurring during the Korean interview, were excluded from analysis.)

This pattern, which is evident in one subject, also held when all the interviewees were pooled by gender. Figure 2 illustrates overall *f*0 measurements from all interviewees, separated by gender and language. Despite a considerable overlap, *f*0 measurements from Korean words were greater on average than *f*0 measurements from English words by up to five semitones.

To test for the significance of the observed difference, two linear mixed effects models were fit on the word-level *f*0 measurements and per-subject *f*0 means (in semitones), with fixed effects of gender, generation, and language spoken, and random effects for subject, according to the formula: *f*0_wd_st $\sim$ language * Gender + Generation + (1 | subject). The results are illustrated in Table 2.

Results from the model show that the interviewees' Korean was significantly higher than their English, regardless of gender or generational status. There was no effect found for generation; that is to say, second generation and 1.5-generation Korean Americans did not behave differently in terms of *f*0. However, a small interaction between language and gender indicates that male and female speakers had somewhat different language effects.

Table 1. Fifteen interviewees' demographic information.

Gender	Second generation	1.5 generation
Female	$N = 5$ (born in USA; AOA = 0)	$n = 4$ (born in Seoul; AOA = 3, 8, 10, 10) $n = 1$ (born in Busan; AOA = 5)
Male	$N = 3$ (born in USA; AOA = 0)	$n = 2$ (born in Seoul; AOA = 9, 10)
Sum	mean age = 21.75 ± 3.3	mean age = 20.3 ± 2.8; mean AOA 7.8 ± 2.8

https://doi.org/10.1121/10.0000498

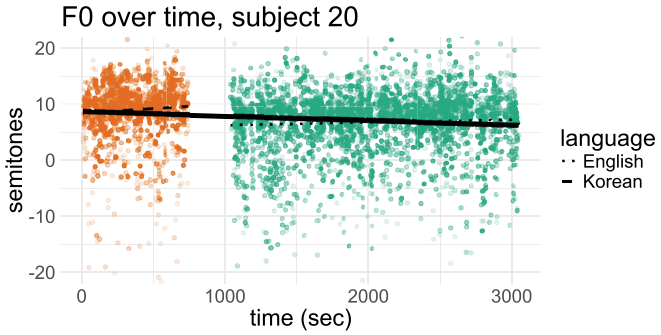

Fig. 1. (Color online) Each dot represents the *f*0 of a vowel during the bilingual interview, with Korean speech (on the left) occurring first, followed by English speech (on the right), after a short period of silence. The regression lines were calculated for each language separately (dashed, dotted) as well as pooled (solid), using lm smoothing in R. This figure represents one subject, a 19-year-old second generation female.

3.2 F0 variance

As for the *f*0 span, the hypothesis was that in these bilingual speakers, Korean speech would have a wider span, indicated by greater variance. Figure 3 shows the coefficient of variance for gender groups, by language. As with *f*0 mean, the two generational groups did not differ, so the two groups are pooled for the visualization. Variance was greater in Korean than in English for the male speakers, but lower in Korean than in English for the female speakers.

To test for the significance of the observed difference, a linear mixed effects model was fit on the word-level *f*0 variance measurements ("cv"), with fixed effects of gender, generation, and language spoken, and random effects for subject, according to the formula: cv ~ language * Gender + Generation + (1 | subject). The results are illustrated in Table 2.

Results from the model show that interviewees' *f*0 variance did not significantly differ depending on the language being spoken alone, but on the language as well as speaker gender.

3.3 IQR and other ranges

A wider span is also indicated by higher values for ranges such as IQR. Figure 3 illustrates the data organized by language, gender, and range type. As with *f*0 mean and variance, the two generational groups did not differ, so that is not plotted. It is clear that male ranges are lower than female ranges in general. In addition, there is a visible interaction effect with language. While male speakers had larger ranges in Korean and smaller ranges in English, female speakers had slightly smaller ranges in Korean compared to English.

A linear mixed effects model was fit for each of the ranges in order to test for the effects of language, generation, and gender on the *f*0 span, according to the formula: rg(50,80,90,98) ~language * Gender + Generation + (1 | subject). The results are illustrated in Table 3.

Results from the model show that interviewees' *f*0 span did not significantly differ depending on the language being spoken. Speakers' generational status was also not a factor. Speaker gender was found to be significant—male speakers had a narrower *f*0 span than female speakers—but more importantly, the interaction effect of language and gender was confirmed. Figure 4 demonstrates this by showing each of the 19 subjects individually, ordered by female speakers first, followed by male speakers. From the English ranges to the Korean ranges, most of the female speakers have a decrease in range or remain level, whereas most of the male speakers have an increase in range.

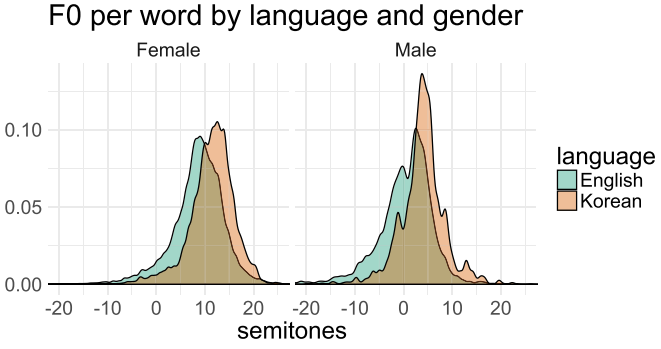

Fig. 2. (Color online) Distribution plots of *f*0 measurements for all interviewees, separated by gender and language. On average, male-identified interviewees had a lower *f*0, and English *f*0 measurements were lower than Korean *f*0 measurements.

https://doi.org/10.1121/10.0000498

Table 2. Two linear mixed effects models were fit on the data to determine the effects of language, speaker gender, and speaker generation on mean *f*0 of a particular word in semitones (*f*0_wd_st) and mean *f*0 per subject in semitones (*f*0_mean_sbj_st). One linear mixed effects model was fit on the data to determine the same effects on *f*0 variance (cv). The table displays coefficients, standard errors, and *p*-value indicators for each fixed effect. Note: ***$p < 0.01$.

	Dependent variable:		
	*f*0_wd_st	*f*0_mean_subj_st	cv
langKrn	2.333***	2.222***	−2.565***
	(0.019)	(0.329)	(0.702)
GenderM	−8.923***	−8.863***	1.322
	(1.433)	(1.264)	(2.393)
GenerationG2	0.759	0.285	−1.506
	(1.354)	(1.164)	(2.187)
langKrn:GenderM	−0.154***	0.881	4.030***
	(0.050)	(0.570)	(1.217)
Constant	8.296***	9.181***	−22.841***
	(1.067)	(0.931)	(1.758)
Observations	480 737	30	30
Log Likelihood	−1 459 200	−50.238	−67.658
Akaike Inf. Crit.	2 918 413	114.476	149.316
Bayesian Inf. Crit.	2 918 491	124.285	159.125

4. Discussion and conclusion

The hypothesis that in bilingual speakers, Korean speech would have a higher *f*0 level was demonstrated to be correct. However, the hypothesis that Korean speech would also have a wider *f*0 span was not supported; instead, whether Korean had a wider *f*0 span depended on speaker gender.

Although the difference is basic and significant, the present data cannot tell us what the cause of the difference is, and one can only speculate. There are four hypotheses that warrant further investigation.

First, the difference could be purely phonetic. As suggested in Lee and Van Lancker Sidtis (2017), the occurrence of fortis and aspirated consonants in Korean might push up the average Korean *f*0 in comparison to English. However, the *f*0 raising effect only occurs at the beginnings of accentual phrases (Cho and Jun, 2000), so the relative rarity of this phenomenon makes it an unlikely contributor to the greater overall values of Korean *f*0. Second, Korean and English may differ in prosodic structure such that Korean *f*0 is higher, an idea supported by studies that specifically looked at *f*0 at crucial intonational points of sentence-long utterances (Mennen *et al.*, 2012; Graham, 2014). Third, the difference could be socio-indexical, in that Korean Americans speak Korean with a higher *f*0 due to social expectations (Loveday, 1981; Ohara, 1999; Brown *et al.*, 2014). However, the results of this study are somewhat inconsistent with the previous literature. Finally, it is also unlikely that the use of a second language played a role. Although Lee and

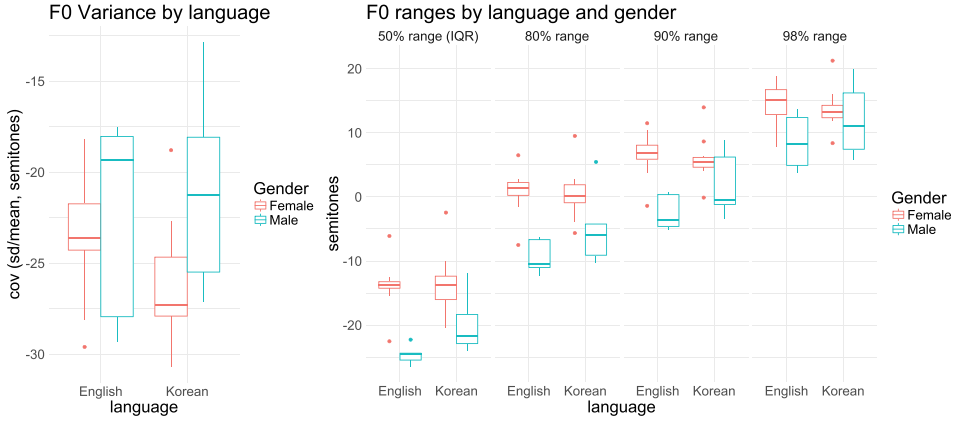

Fig. 3. (Color online) Means and distributions of *f*0 variance (coefficient of variance) for gender and language (left) and *f*0 range (IQR, 80%, 90%, and 98% range) for gender and language (right). No effect was found for gender or generation, but an interaction effect was found with gender and language.

Table 3. A linear mixed effects model was fit on the data to determine the effects of language, speaker gender, and speaker generation on *f*0 IQR, 80% range, 90% range, and 98% range per subject, in semitones. Note: $p < 0.1$; $p < 0.05$; $p < 0.01$.

	Dependent variable:			
	rg50 (IQR)	rg80	rg90	rg98
langKrn	0.598	−0.365	−0.615	−0.813
	(1.321)	(1.343)	(1.012)	(0.862)
GenderM	−10.647***	−9.953***	−8.837***	−5.921***
	(2.350)	(2.289)	(2.107)	(2.274)
GenerationG2	−0.536	−1.853	−1.374	−0.502
	(1.939)	(1.863)	(1.810)	(2.029)
langKrn:GenderM	4.226*	4.905**	5.039***	4.312***
	(2.287)	(2.325)	(1.752)	(1.493)
Constant	−13.638***	1.718	7.210***	14.809***
	(1.664)	(1.613)	(1.513)	(1.655)
Observations	30	30	30	30
Log Likelihood	−74.420	−74.155	−70.132	−69.414
Akaike Inf. Crit.	162.840	162.309	154.263	152.827
Bayesian Inf. Crit.	172.648	172.118	164.072	162.636

Van Lancker Sidtis (2017) propose that speaking a foreign language may raise *f*0, the speakers in the current study acquired English at an early age and would not consider English to be a foreign or second language; neither would they be considered L2 Korean learners. No significant effects of age of acquisition or Korean proficiency were found to affect the observed phenomenon of higher *f*0 in Korean.

4.1 Conclusion

The present study has examined the vocal *f*0 level and span in the natural speech of bilingual speakers of Korean and English and found a significant difference in the *f*0 level. One can be sure that anatomical differences play no role in this robust cross-linguistic difference. The *f*0 span was found to be dependent on speaker gender as well as language spoken. Finally, the study compared simultaneous bilinguals (second generation Korean Americans) to sequential bilinguals (1.5 generation Korean Americans) and found no significant difference between the two groups in the *f*0 level and span.[3]

With respect to other acoustic and phonetic research, the results of this study indicate that it is important to account for the effects of bilingualism (or a bilingual mode of speaking) when conducting research with natural speech. A bilingual speaker may have different average *f*0 values for each of the languages they speak, depending on what the languages are, or which language was spoken in childhood.

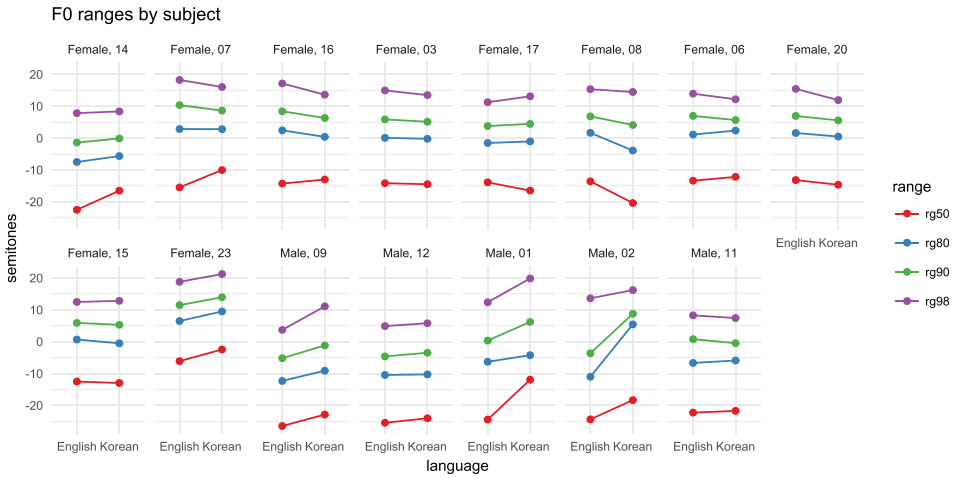

Fig. 4. (Color online) Individual range measurements for each speaker: English on the left and Korean on the right, with a comparison line in between. The widest range (98%) is at the top, followed by 90%, 80%, and 50% (IQR). The general trend is for female speakers to have a smaller range in Korean, while male speakers have a larger range in Korean.

https://doi.org/10.1121/10.0000498

Acknowledgments

The author would like to thank the following people for their help and guidance with this project: Cindy Jin, Francis Zheng, Ung Bee Anna Park, Mingde Chong, Esther Yom, Younie Park, Ronald Sprouse, Keith Johnson, and two anonymous reviewers.

References and links

[1]In this paper, I use *f*0 instead of pitch, which refers to the perceptual dimension of frequency.

[2]I am grateful to an anonymous reviewer who noted that if early bilinguals experienced any language attrition and were no longer proficient in Korean but were aware of this, then a type of stress that results from speaking a non-dominant language may have raised their *f*0 (Giddens *et al.*, 2013). It remains possible that this phenomenon may have affected these speakers, as bilingual speakers who are more dominant in the ambient language of English.

[3]See supplementary material at https://doi.org/10.1121/10.0000498 for additional notes on research methodology and their bearing on the conclusions.

Altenberg, E. P., and Ferrand, C. T. (**2006**). "Fundamental frequency in monolingual English, bilingual English/Russian, and bilingual English Cantonese young adult women," J. Voice **20**(1), 89–96.

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

Brown, L., Winter, B., Idemaru, K., and Grawunder, S. (**2014**). "Phonetics and politeness: Perceiving Korean honorific and non-honorific speech through phonetic cues," J. Pragmatics **66**, 45–60.

Cho, T., and Jun, S.-A. (**2000**). "Domain-initial strengthening as enhancement of laryngeal features: Aerodynamic evidence from Korean," UCLA Working Papers in Phonetics, pp. 57–70.

Giddens, C. L., Barron, K. W., Byrd-Craven, J., Clark, K. F., and Winter, S. A. (**2013**). "Vocal indices of stress: A review," J. Voice **27**(3), 390.

Graham, C. R. (**2014**). "Fundamental frequency range in Japanese and English: The case of simultaneous bilinguals," Phonetica **71**, 271–295.

Kang, K.-H., and Guion, S. G. (**2006**). "Phonological systems in bilinguals: Age of learning effects on the stop consonant systems of Korean-English bilinguals," J. Acoust. Soc. Am. **119**(3), 1672–1683.

Keating, P., and Kuo, G. (**2012**). "Comparison of speaking fundamental frequency in English and Mandarin," J. Acoust. Soc. Am. **132**(2), 1050–1060.

Lee, B., and Van Lancker Sidtis, D. (**2017**). "The bilingual voice: Vocal characteristics when speaking two languages across speech tasks," Speech, Lang. Hear. **20**, 174–185.

Lee, S. A. S., and Iverson, G. K. (**2012**). "Stop consonant productions of Korean–English bilingual children," Bilingualism: Lang. Cognit. **15**(2), 275–287.

Loveday, L. (**1981**). "Pitch, politeness and sexual role: An exploratory investigation into the pitch correlates of English and Japanese politeness formulae," Lang. Speech **24**(1), 71–89.

Mennen, I., Schaeffler, F., and Docherty, G. (**2012**). "Cross-language differences in fundamental frequency range: A comparison of English and German," J. Acoust. Soc. Am. **131**(3), 2249–2260.

Ohara, Y. (**1999**). "Performing gender through voice pitch: A cross-cultural analysis of Japanese and American English," in *Perceiving and Performing Gender*, edited by U. Pasero and F. Braun (Westdeutscher Verlag, Opladen/Wiesbaden), pp. 105–116.

Ordin, M., and Mennen, I. (**2017**). "Cross-linguistic differences in bilinguals' fundamental frequency ranges," J. Speech, Lang., Hear. Res. **60**(6), 1493–1506.

Paradis, J. (**2007**). "Early bilingual and multilingual acquisition," Handbook Multiling. Multiling. Commun. **5**, 15–44.

Park, K. (**1999**). "'I Really Do Feel I'm 1.5!': The construction of self and community by young Korean Americans," Amerasia J. **25**(1), 139–163.

Ueda, Y., Hamakawa, T., Sakata, T., Hario, S., and Watanabe, A. (**2007**). "A real-time formant tracker based on the inverse filter control method," Acoust. Sci. Technol. **28**(4), 271–274.

Wagner, P., Trouvain, J., and Zimmerer, F. (**2015**). "In defense of stylistic diversity in speech research," J. Phonetics **48**, 1–12.

Yuan, J., and Liberman, M. (**2008**). "Speaker identification on the scotus corpus," J. Acoust. Soc. Am. **123**(5), 9687–9690.