

Effects of Early Bilingualism on Voice Quality and Voice Perception

Sylvia Cho

Simon Fraser University

This study examines how early multilingual experience shapes voice quality production and perception, focusing on Korean heritage speakers in Vancouver, Canada. Heritage speakers acquire a minority language early in childhood but typically become dominant in the societal language, leading to long-term reorganization of their phonetic systems. While some experience substantial attrition, others maintain high proficiency, resulting in wide variability in heritage-language ability. Previous research on heritage language development has focused primarily on segmental contrasts and a limited set of voice quality measures (e.g., fundamental frequency [1]). Much less is known about how voice quality develops in multilingual speakers, particularly when phonation is not lexically contrastive. Although recent studies on Cantonese and Mandarin heritage speakers suggest relative cross-language stability in individual voice quality [2, 3], the perceptual consequences of within-speaker, language-dependent shifts in acoustic voice space remain unclear.

Production data were collected from 28 Korean heritage speakers producing Korean and English speech in read and extemporaneous styles. Acoustic analyses targeted fundamental frequency (F0), long-term average spectra (LTAS), and long-term formant distributions (LTF). Building on prior work [2–5], 24 source- and filter-based measures were extracted to examine structured covariation among perceptually salient dimensions of voice quality.

Results reveal robust within-speaker, cross-language differences in voice quality persisting into adulthood. Speakers produced Korean with a higher F0 mean and greater F0 variability, with stronger effects in read speech and among female talkers. F0 range was often greater in extemporaneous speech, suggesting that early-acquired pitch patterns remain available but are modulated by task and sociophonetic factors. Long-term formant analyses showed language-dependent shifts in F2 distributions (higher in English), consistent with persistent vocal-tract configuration differences (e.g., /u/ fronting in English). LTAS measures further indicated systematic language effects in spectral energy and tilt, with more high-frequency energy in English speech (see Figure 1), pointing to distinct long-term voice settings linked to early language experience. Source- and filter-based analyses showed that although between-speaker differences accounted for most variance, consistent language-related shifts emerged along dimensions associated with F0 and breathiness.

A complementary perception experiment assessed cross-language speaker recognition using time-reversed speech. In a speaker identification task completed by 276 listeners (heritage, bilingual, and monolingual), ordinal regression analyses revealed a main effect of language, with listeners more likely to judge a talker as a different person across language shifts (see Figure 2). All groups performed similarly but bilingual and Korean heritage listeners showed a small advantage in speaker discrimination, supporting a bilingual benefit in voice perception. Listeners were also more accurate overall when speech style matched across samples. These results demonstrate sensitivity to long-term voice characteristics shaped by linguistic experience and show that acoustic differences emerging from linguistic structure influence the perceptual processing of indexical, speaker-identity-relevant information.

These findings show that voice quality reflects both developmental stability and adaptive flexibility in multilingual speakers. Early language exposure leaves lasting imprints on global voice structure that remain perceptible to listeners decades later, shaping not only how voices are produced but also how they are perceived. Importantly, results demonstrate that early linguistic experience influences the shapes not only phonological categories but also long-term, indexical properties of the voice and their perceptual interpretation.

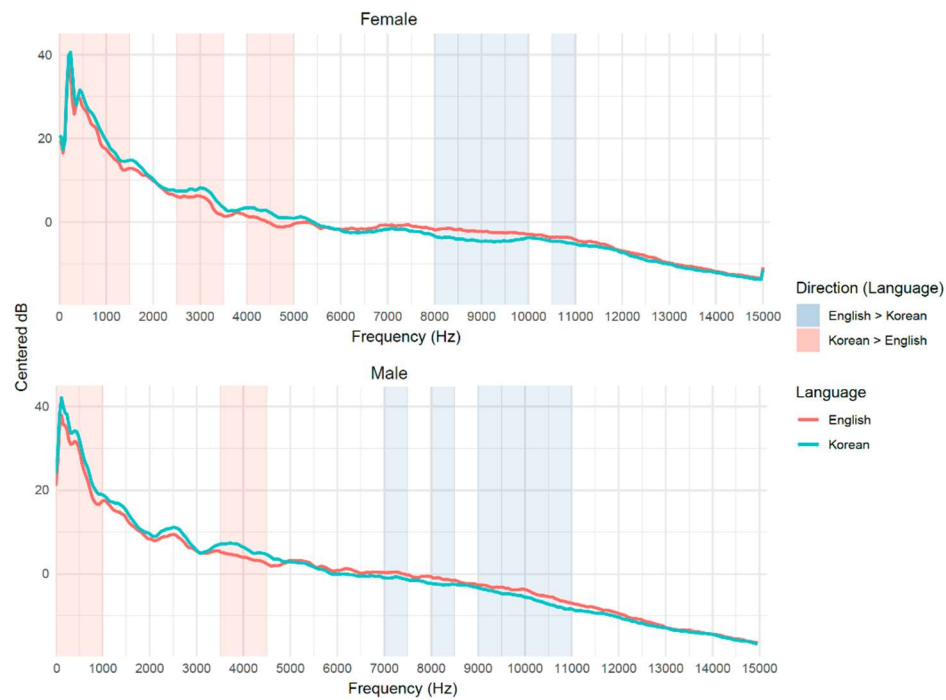


Fig. 1. LTAS for English and Korean Speech by Talker Sex

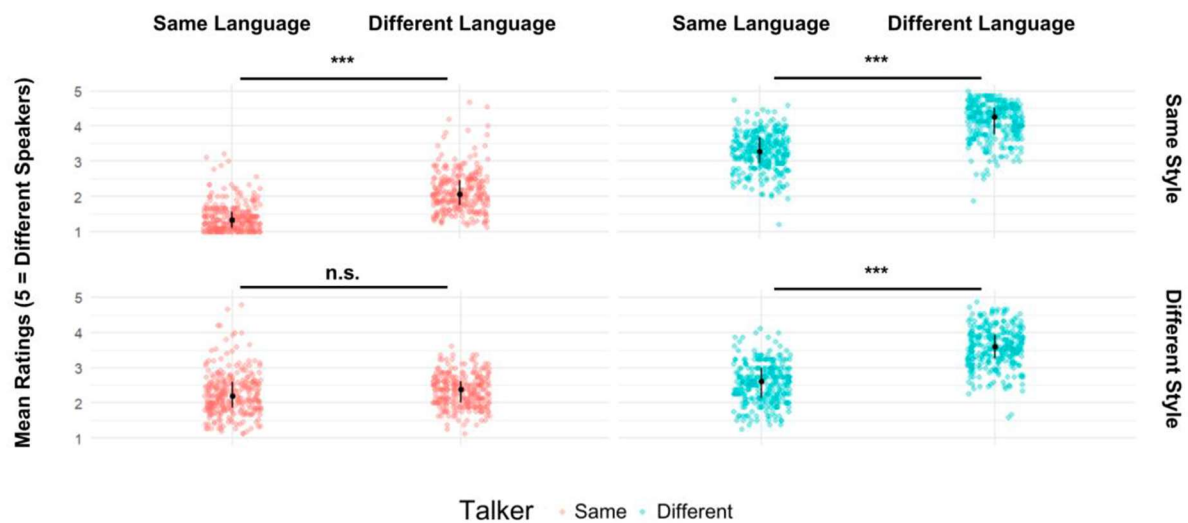


Figure 2. Jitter Plot Showing the Interaction of Talker, Language, and Style on Voice Perception Ratings, Highlighting Language Comparisons Based on Subject-Level Means, Estimated Marginal Means, and 95% Confidence Intervals. (***) = $p < .001$, (**) = $p < .01$, (*) = $p < .05$)

References

- [1] Lee, B., & Sidtis, D. V. L. (2017). The bilingual voice: Vocal characteristics when speaking two languages across speech tasks. *Speech, Language and Hearing*, 20(3), 174-185.
- [2] Johnson, K. A., & Babel, M. (2023). The structure of acoustic voice variation in bilingual speech. *The Journal of the Acoustical Society of America*, 153(6), 3221-3221.
- [3] Liu, S., & Babel, M. (2025, December). Acoustic consistency of Mandarin-English bilingual voices: Evidence from the Mandarin-English Language Interview (MELI) Corpus [Conference presentation, Abstract 3pSC18]. *190th Meeting of the Acoustical Society of America*, Honolulu, HI, United States.
- [4] Lee, Y., Keating, P., & Kreiman, J. (2019). Acoustic voice variation within and between speakers. *The Journal of the Acoustical Society of America*, 146(3), 1568-1579.
- [5] Kreiman, J., & Lee, Y. (2025). Biological, linguistic, and individual factors govern voice quality. *The Journal of the Acoustical Society of America*, 157(1), 482-492.