Crosslinguistic influence and language-specific strategies in bilingual clear speech

Ye-Jee Jung, Olga Dmitrieva

Purdue University
jung292@purdue.edu, odmitrie@purdue.edu

ABSTRACT

The current study investigated how late Korean-English bilinguals produced clear speech in their L1 (Korean) and L2 (English). Specifically, we examined the acoustic enhancement of laryngeal contrasts in English and Korean clear speech. Thirty Korean-English bilinguals produced casual and clear speech in each of the two languages by reading Korean and English word-lists. In addition, monolinguals of the two languages (20 in each group) produced casual and clear speech under the same experimental conditions, for comparison.

The results showed that bilinguals used onset f0 to enhance the laryngeal categories in a distinctly language-specific way in English and Korean. The use of VOT as a correlate of laryngeal categories in bilingual clear speech suggested mutual crosslinguistic influence between English and Korean. Therefore, the acoustic structure of bilingual clear speech resembled the monolingual counterparts but also carried unique acoustic characteristics.

Keywords: Korean, English, bilingual speech, clear speech, crosslinguistic influence

1. INTRODUCTION

The H & H theory [12] postulates that speech is produced on a continuum between hypo-articulation (less articulatory efforts) and hyper-articulation (more articulatory efforts), depending on the communicative goals and settings. Clear speech is one form of hyper-articulated speech, and it is a distinctive speaking style directed at increasing speech intelligibility in communicatively compromised settings (e.g., the presence of background noise). Increased acoustic difference between phonological categories contributes to speech intelligibility and is typically found among the features of clear speech. Because languages differ with respect to the types and acoustic implementation of phonological categories, clear speech strategies are by necessity at least partly language-specific.

For example, in English, word-initial stop voicing distinction is realized primarily via the voice onset time (VOT) difference: longer VOT for voiceless stops, shorter VOT for voiced stops. This VOT difference is further enhanced in clear speech, typically via the asymmetrical lengthening of voiceless VOT [13]. In Korean, the three-way contrast between lenis, aspirated, and fortis stops is realized via VOT and/or onset f0. One or the other cue becomes more important depending on the specific binary contrast. For example, the lenis-aspirated contrast in modern Seoul Korean is believed to rely primarily on the onset f0 difference. Correspondingly, this cue is further enhanced in clear speech, where lenis and aspirated stops are even more distanced via onset f0 [9].

While these speech behaviours have been relatively well researched in monolingual speech, considerably less is known about clear speech produced by bilingual second language speakers. There are at least two main questions that can be asked with respect to bilingual clear speech. First, to what extent do bilingual speakers maintain language-specific clear speech strategies in both of their languages. And second, to what extent is clear speech subject to bidirectional crosslinguistic influence in bilingual speakers predicted by the Speech Learning Model (SLM/SLM-r) [5, 6]. In order to address these questions, the present study examined how laryngeal contrasts were realized in casual speech and enhanced in Korean and English clear speech produced by Korean (L1) – English (L2) bilinguals, in comparison to Korean and English monolingual speakers.

To our knowledge, previous research on bilingual clear speech has been limited. For example, native Croatian speakers seemingly transferred their L1 clear speech strategies to their L2 (English) [14]. Particularly, they did not lengthen the VOT of English /p/ in clear speech, instead tending towards the enhancement of prevoicing for English /b/. In another study, native Finnish speakers increased the VOT of the English /p/ in English clear speech, enhancing the acoustic distance between the English /p/ and /b/, similarly to native speakers of English [8]. Therefore, existing research suggests that bilingual clear speech can be both language-appropriate and subject to crosslinguistic influence.

Accordingly, we hypothesize that Korean bilinguals can produce English and Korean clear speech where the laryngeal categories are enhanced in a language-specific manner, at least to some extent. At the same time, we also expect to see some evidence of crosslinguistic influence, such that English laryngeal categories are enhanced in a somewhat Korean-like fashion (e.g. with a significantly greater reliance on onset f0 and weaker reliance on VOT than in monolingual English speech), and vice versa (e.g. Korean laryngeal categories' enhancement primarily via VOT).

2. METHODS

2.1. Participants

Three speaker groups participated in the present study: 20 native American English monolinguals (4 male, mean age = 24.95_{yrs}, SD = 9.30), 30 late Korean-English bilinguals residing in the United States at the time of recording (20 male, mean age = 29.73_{yrs}, SD = 3.48), and 20 Korean monolinguals living in South Korea (8 male, mean age = 27.40_{yrs}, SD = 5.52). Korean and English monolinguals served as control groups for each language of the bilingual speakers.

2.2. Procedures

Participants performed a word-list reading task, with words displayed one by one on the computer screen in random order. Each word remained on the screen for 1.8 seconds. Korean words were presented in Korean orthography. In each speaking style, three repetitions of each word were recorded for each participant. For casual speech, participants were instructed to read each target word casually, as if they were talking to their friends or family members. For clear speech, they were instructed to read each word clearly, as if their interlocutors were hearing-impaired or elderly. Casual speech recording preceded clear speech with a short self-timed break provided between the two recording sessions. Recordings were performed in a sound-attenuated booth in a speech laboratory or in a quiet classroom using professional quality recording equipment.

2.3. Stimuli

For English, six minimal pairs that differed in voicing of the word-initial alveolar stops were used (e.g. *tab* vs. *dab*). For Korean, we used six triplets/near triplets differing in the laryngeal category of the word-initial stops adopted from [9], for example, *t^{h}ant^{h}anhata* 'to be solid' (aspirated), *tantanhata* 'to be hard' (lenis), and *t*ant*anhata* 'to be hard/strong (fortis). Additional distractor items (16 English and 18 Korean) with similar syllabic structure were included in each word-list, e.g. *pan* and *manmanhata*. As a result, 72 English target words (168 including fillers) were and 108 Korean target words (216 including fillers) were recorded per participant per language.

2.4. Acoustic measurements

For each target word, VOT (in milliseconds) of the word-initial stops and onset f0 (in hertz) at the onset of the vowel following the word-initial stop vowel were annotated and measured using custom scripts in Praat [3]. To control for individual variability, including anatomic differences, in f0 production, raw onset f0 values were normalized into semitones (ST) using the formula in [4].

2.5. Statistical analysis

A series of linear mixed-effect models was implemented in R using the *lme4* package [1]. VOT and onset f0 were the dependent variables analysed in separate models. The following sum-coded fixed effects were used for English models: Speaker Group (English monolinguals and Korean-English bilinguals), Speaking Style (casual and clear), and Stop Type (voiced and voiceless) and all two- and three-way interactions. By-participant intercepts and random slopes for Speaking Style and by-item random intercepts were also included. For Korean data, a structurally identical model was used, except Stop Type in this model had three levels (aspirated, lenis, and fortis). Therefore, for the Stop Type effect and every interaction including the effect, the *emmeans* package [11] was implemented to perform pairwise post-hoc analyses. Statistical significance of test and interactions was established with ANOVA type III tests implemented using the *car* package [7]. In reporting the interactions, we focus on those that included Stop Type. For significant effects and interactions, we calculated partial eta squared (η_p^2) using the *effectsize* package as an estimate of effect size [2].

3. RESULTS

3.1. English clear speech

The VOT results for each speaker group and each speaking style are illustrated in Figure 1. The significant effects of Speaking Style and Stop Type were observed: VOT was longer in clear speech ($\chi2$ (1) = 51.43, p < .001, η_p^2 = 0.52) and in voiceless stops ($\chi2$ (1) = 6040.74, p < .001, η_p^2 = 1.00).

All two-way and three-way interactions including Stop Type were also significant. Most importantly, the English group made a greater VOT distinction between two voicing categories than the bilingual group (Stop Type by Group interaction, $\chi2$ (1) = 58.16, p < .001, η_p^2 = 0.02). The VOT distinction between voiced and voiceless stops was greater in clear speech than in casual speech (Stop Type by Speaking Style interaction, $\chi2$ (1) = 362.76, p < .001,

η_p^2 = 0.10). As Figure 1 demonstrates, the enhancement was achieved via the asymmetrical lengthening of voiceless stops' VOT. On the other hand, VOT of voiced stops remained stable across speaking styles in both speaker groups. Finally, the magnitude of clear speech voicing enhancement via VOT was greater for the English group compared to the bilingual group (Stop Type by Group by Speaking Style interaction, $\chi2$ (1) = 101.71, p < .001, η_p^2 = 0.03).

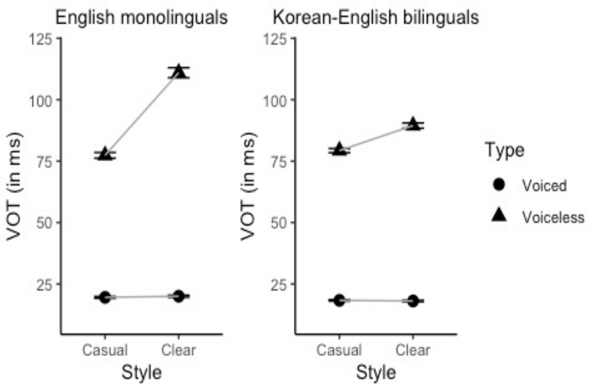

Figure 1: VOT of English stops produced by native English monolinguals (left) and Korean-English bilinguals (right) in each speaking style.

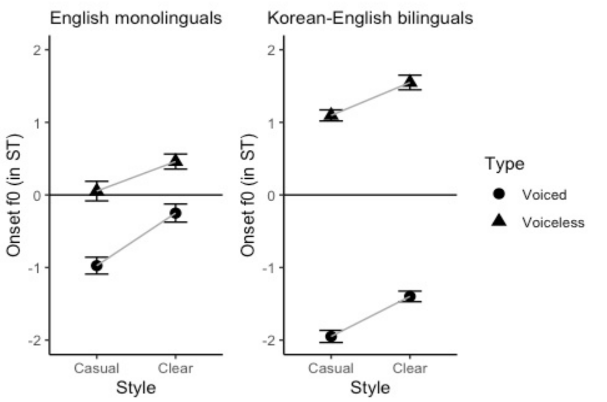

Figure 2: Onset f0 of English stops produced by native English monolinguals (left) and Korean-English bilinguals (right) in each speaking style.

Figure 2 demonstrates the onset f0 distinction between the two stop types in each speaker group and each speaking style. Speaking Style and Stop Type effects were significant: onset f0 was higher in clear speech ($\chi2$ (1) = 56.53, p < .001, η_p^2 = 0.02) and after voiceless stops ($\chi2$ (1) = 254.01, p < .001, η_p^2 = 0.96).

The significant interaction between Speaker Group and Stop Type was attested due to the fact that bilingual speakers made a greater onset f0 distinction between voiced and voiceless stops than English monolinguals ($\chi2$ (1) = 224.44, p < .001, η_p^2 = 0.06), as shown in Figure 2. However, no other interactions were significant, indicating that onset f0 difference between voiced and voiceless stops was not enhanced in clear speech, compared to casual speech, in either of the speaker groups.

3.2. Korean clear speech

Figure 3 shows the VOT results for Korean casual and clear speech for each speaker group. The effect of Stop Type was significant ($\chi2$ (2) = 239.74, p < .001, η_p^2 = 0.98). Post-hoc tests indicated that the VOT of aspirated and lenis stops was significantly longer than the VOT of fortis stops [aspirated-fortis: β = 52.43, p < .001; fortis-lenis: β = -49.48, p < .001]. Lenis and aspirated stops did not differ from each other in terms of VOT. Average VOT was also longer in clear speech than in casual speech ($\chi2$ (1) = 15.12, p < .001, η_p^2 = 0.24).

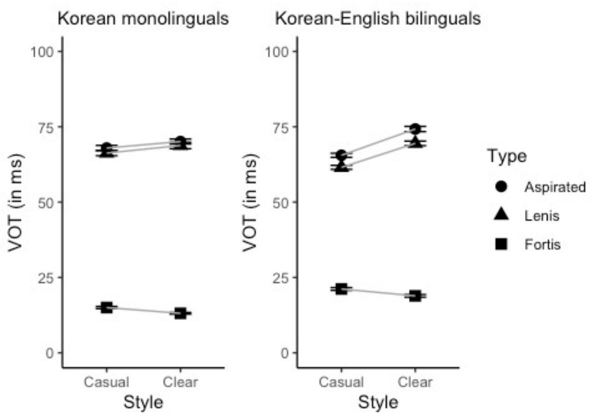

Figure 3: VOT of Korean stops produced by Korean monolinguals (left) and Korean-English bilinguals (right) in each speaking style.

A number of significant interactions indicated variability in the way each speaker group realized and enhanced laryngeal categories. First, aspirated and lenis VOT was lengthened in clear speech while fortis VOT was shortened in clear speech (Stop Type by Speaking Style interaction, $\chi2$ (2) = 106.75, p < .001, η_p^2 = 0.02). Second, bilinguals made a smaller VOT difference between laryngeal categories than monolingual speakers (Group by Stop Type interaction, $\chi2$ (2) = 98.02, p < .001, η_p^2 = 0.02). Finally, the enhancement of VOT differences between the laryngeal categories was more pronounced in bilingual than in monolingual clear speech (Stop Type by Speaking Style by Group interaction, $\chi2$ (2) = 19.98, p < .001, η_p^2 = 0.00).

As Figure 4 illustrates, Korean stop types were well-dispersed in terms of onset f0 for both speaker groups. The main effects of Stop Type ($\chi2$ (2) = 2139.26, p < .001, η_p^2 = 1.00) and Speaking Style ($\chi2$

(1) = 214.56, $p < .001$, $\eta_p^2 = 0.04$) were significant. Post-hoc tests indicated that onset f0 of aspirated stops was higher than that of both lenis and fortis stops (aspirated-fortis: $\beta = 2.63$, $p < .001$; aspirate-lenis: $\beta = 4.63$, $p < .001$) while lenis f0 was lower than fortis f0 ($\beta = 2.00$, $p < .001$). Average onset f0 was also higher in clear than in casual speech.

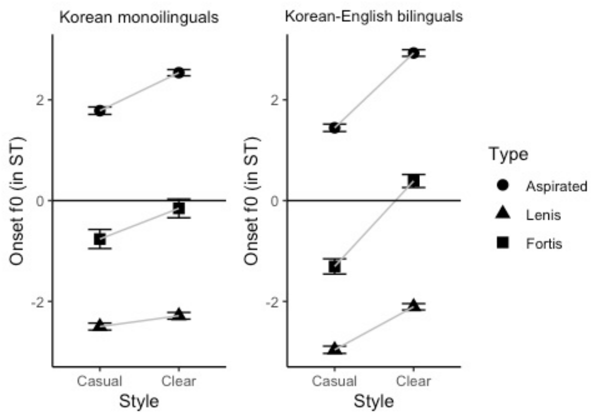

Figure 4: Onset f0 of Korean stops produced by Korean monolinguals (left) and Korean-English bilinguals (right) in each speaking style.

There was a significant two-way interaction between Speaking Style and Stop Type ($\chi2$ (2) = 19.77) = 9,87, $p < .001$, $\eta_p^2 = 0.00$). Onset f0 of every stop type was raised in Korean clear speech, but the magnitude of the increase was greater for aspirated and fortis stops than for lenis stops, resulting in an enhanced f0 difference between lenis and other types of stops in clear speech.

4. DISCUSSION AND CONCLUSIONS

The presented study investigated native and non-native clear speech produced by late Korean-English bilinguals, focusing on language-specific strategies and crosslinguistic influence. The results, first and foremost, indicated that the experiment was successful in eliciting clear speech. As could be expected based on previous research [9, 10, 13], average VOT was lengthened and average f0 was raised in both Korean and English clear speech across speaker groups.

Even more importantly, for both cues, this increase was often asymmetric: VOT was lengthened for voiceless but not voiced stops in English and for lenis and aspirated stops but not fortis stops in Korean. Similarly, onset f0 was raised for aspirated and fortis stops to a greater extent compared with lenis stops in Korean. As a result of these manoeuvres, the difference between contrastive phonological categories in both languages was acoustically enhanced in clear speech.

Finally, as predicted, we observed that bilingual clear speech was both language-specific/language-appropriate and subject to crosslinguistic influence.

We can deduce language-specificity by observing what bilinguals did differently across their two languages, especially if the difference is in line with the corresponding monolingual behaviours. The clearest example is the use of onset f0 to enhance laryngeal contrasts in clear speech. Bilinguals did not increase onset f0 difference between voiced and voiceless English stops in clear speech, similarly to native speakers of English. Yet bilinguals did increase onset f0 difference between Korean laryngeal categories, just as monolingual speakers of Korean did. These findings suggest a great sensitivity on the part of bilingual speakers to the way the same cue is used in different languages and awareness of the potential utility of this cue in increasing speech intelligibility in each language.

Crosslinguistic influence can be established by observing what bilinguals did differently from monolingual speakers of the same language, especially if this difference is in line with the behaviour of monolinguals of another language. At least two findings to that effect can be noted. First, bilingual speakers realized a significantly greater onset f0 difference between voiced and voiceless English stops than English monolinguals (although this difference was not further enhanced in clear speech). This behaviour is explainable by the fact that a subset of Korean laryngeal distinctions is cued by rather sizeable f0 differences.

Second, in clear speech, bilinguals produced *less* enhancement of VOT differences between English voicing categories than English monolinguals, but *more* enhancement of VOT differences between Korean laryngeal categories than Korean monolinguals. As such, bilinguals' clear speech enhancement behaviour with respect to VOT was 'intermediate' between the corresponding monolingual patterns. Such 'intermediate' patterns are often found in the phonetics of bilingual speech. This finding is explainable by the fact that VOT plays a more important role in cueing and enhancing laryngeal categories in English than in Korean.

It remains to be understood why onset f0 was used in a language-specific fashion while the use of VOT was subject to crosslinguistic influence. Notwithstanding, bilingual results indicate remarkable sensitivity to crosslinguistic differences, as well as sophistication and flexibility in navigating these differences in adaptive speech behaviours.

5. REFERENCES

[1] Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1-48.
[2] Ben-Shachar, M. S., Lüdecke, D., & Makowski, D. (2020). effectsize: Estimation of effect size indices and standardized parameters. *Journal of Open Source Software, 5*(56), 2815.
[3] Boersma, P. & Weenink, D. (2022). Praat: doing phonetics by computer [Computer program]. Version 6.3.03
[4] Dmitrieva, O., Llanos, F., Shultz, A. A., & Francis, A. L. (2015). Phonological status, not voice onset
time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics*, *49*, 77-95.
[5] Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research, 92*, 233-277.
[6] Flege, J., & Bohn, O. (2021). The Revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), *Second Language Speech Learning: Theoretical and Empirical Progress* (pp. 3-83). Cambridge: Cambridge University Press.
[7] Fox J, Weisberg S. (2019). An R Companion to Applied Regression, Third edition. Thousand Oaks, CA: Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/.
[8] Granlund, S., Hazan, V., & Baker, R. (2012). An acoustic–phonetic comparison of the clear speaking styles of Finnish–English late bilinguals. *Journal of Phonetics, 40*(3), 509-520.
[9] Kang, K. H., & Guion, S. G. (2008). Clear speech production of Korean stops: Changing phonetic targets and enhancement strategies. *The Journal of the Acoustical Society of America*, *124*(6), 3909–3917.
[10] Krause, J. C., & Braida, L. D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *The Journal of the Acoustical Society of America*, *115*(1), 362-378.
[11] Lenth, R. (2018). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.3.0. https://CRAN.R-project.org/package=emmeans.
[12] Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. *Speech production and speech modelling*, 403-439.
[13] Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech, Language, and Hearing Research, 29*(4), 434-446.
[14] Smiljanic, R., & Bradlow, A. R. (2009). Native and non-native clear speech production. *The Journal of the Acoustical Society of America*, *125*(4), 2753-2753.