THE EFFECTS AND INTERACTION OF AGE AND REGION ON THE ACOUSTIC REALIZATION OF THE KOREAN STOP CONTRAST

Jeffrey J. Holliday[1], Eun Jong Kong[2], Hyunjung Lee[3]

[1]University of Kansas, [2]Korea Aerospace University, [3]Incheon National University
[1]jjh@ku.edu, [2]ekong@kau.ac.kr, [3]hjlee123@inu.ac.kr

ABSTRACT

Previous studies exploring phonological contrasts undergoing sound change have found that the relationship between production and perception can depend on factors such as region, age, and social setting. It remains unclear how this may unfold with multiple acoustic cues, and in stationary and mobile speakers. In this study, we measured VOT and F0 in the word-initial stops of 197 speakers of Seoul (standard) and Gyeongsang (non-standard regional variety) Korean in elementary school, high school, and university, along with Gyeongsang-to-Seoul university student transplants. In line with perception results from the same population, we found that the high school students' productions in both regions were the most regionally distinctive. While the Gyeongsang university students were no different from the high school students, the acoustic cues used by the transplants shifted toward a Seoul pattern. These results suggest that although production may be less flexible than perception, physical migration may modulate this effect.

Keywords: sound change, Korean stop production, acoustic cue weighting, mobile speakers

1. INTRODUCTION

Studies on phonological contrast undergoing a sound change have produced mixed findings regarding the relationship between perception and production. While many studies have found that perception and production are correlated to some degree, it has been reported that changes in perception precede changes in production [1, 2], that changes in production precede changes in perception [3, 4], and that perception leads at the beginning of the change but begins to lag behind production as the change nears completion [5, 6]. Sound changes spread not only over time, however, but also within a sociolinguistic context. This includes not only geographical space (e.g. from one region to another), but also speech communities and social networks within a given region (e.g. entering university after high school). In this context, we are interested in how sound changes spread into non-standard varieties. The present study therefore explored this question by investigating local and mobile speakers' use of acoustic cues to produce a contrast undergoing a sound change from below originating in the standard variety.

The sound change in the Korean three-way stop laryngeal contrast has been extensively studied in terms of both production and perception, both within Seoul and across other regions [7]. While Seoul Korean (SK) lenis and aspirated stops were historically cued by VOT (intermediate vs. long, respectively), over the past 60 years speakers have shifted to relying primarily on F0 (lenis < aspirated) in both production [8] and perception [9], effectively neutralizing VOT.

The potential for F0 to serve as the primary cue to the lenis-aspirated contrast has been questioned, however, for tonal varieties, such as Gyeongsang Korean (GK), spoken in the southeast of the country. Ostensibly because GK uses F0 to cue lexical pitch accent, studies have found that GK speakers still rely on VOT to cue the lenis-aspirated contrast [10]. But when the productions of elderly and teenage GK speakers have been directly compared, it was found that younger speakers did use F0 more [11], suggesting that the change observed in SK may be spreading to GK as well.

One recent study investigated the perception of this contrast by elementary school, high school, and university students in Seoul and Gyeongsang [12]. It was found that although the sound change is by now most likely complete in SK speakers this young, it is still in progress in GK. It was further found that the high school students in both regions exhibited perceptual cue weightings at the extreme ends of the change: SK high school students relied on VOT the least, and GK high school students relied on VOT the most. This result suggests that the high school students' perception was most tuned to the local variety, and the experience of attending university and interacting more with speakers from regions more or less advanced in the sound change could influence their perceptual cue weighting.

But given the debate over whether perception precedes production in an undergoing sound change, it is unclear whether SK and GK high school and university students would exhibit patterns of cue weighting in production that mirrored their perception. It was proposed that GK university students were more advanced than GK high school

students in the perception of the sound change, despite their older chronological age, due to their place in a speech community (i.e. the university) that includes more speakers who are more advanced in the change. The current study therefore investigates how the production of the Korean stop contrast varies in GK speakers across these same age groups (elementary school, high school, university), comparing their productions with both SK speakers, and another group of GK speakers who had moved to Seoul to attend university.

2. METHOD

2.1. Participants

The production data were collected from a total of 197 subjects from three dialect groups: 87 Gyeongsang, 90 Seoul, and 20 transplants who grew up in Gyeongsang but were attending university in Seoul. Gender in each group was split roughly equally between female and male. The number of speakers and mean age of each group is shown in Table 1. The GK and SK speakers were all born and educated in their respective dialect regions, mostly in the city of Changwon (where the South GK dialect is spoken) and the greater Seoul area, respectively. All participants, including the elementary school students, were literate in Korean. Subjects reported no language or hearing problems and were paid for their participation.

	SK	GK	Transplant
Elementary	27 (9.0)	23 (9.0)	
High School	25 (16.0)	29 (16.0)	
University	38 (24.2)	35 (23.8)	20 (19.6)

Table 1: No. of speakers (mean age) in each group.

2.2. Stimuli and recording

Each participant produced 27 disyllabic words, which targeted three initial stop phonation types (fortis, lenis, aspirated), three accent types (HL, HH, LH), and three places of articulation (bilabial, alveolar, velar). Most of the high school and university students produced three repetitions of a randomized word list presented in Korean orthography, but some of the high school students and elementary school students only recorded one or two repetitions due to fatigue. A total of 14,523 tokens were collected using a digital recorder and a Shure condenser microphone (SM81-LC) at a sampling rate of 44,100 Hz.

2.3. Acoustic measurements

The two acoustic measurements analyzed here are the VOT of the word-initial stop and the F0 of the following vowel. All measurements were done in Praat using a semi-automated script. The script stopped at each target word and let the user visually identify the stop burst and onset of voicing in the following vowel. The script then calculated the VOT as the difference between these two landmarks, and F0 as the mean over the first 15 ms after the voicing onset. For tokens with no measurable F0, which was mostly due to creakiness, the script advanced forward in 5 ms steps until F0 could be measured. If no F0 could be measured within 20 ms, the token was discarded. Lastly, F0 in Hz was converted to semitones.

2.4. Analysis

We used mixed effects logistic regression models to determine the extent to which VOT and F0 were used by speakers to signal the lenis-aspirated stop contrast. The present analysis focused only on the lenis and aspirated stops based on their observed generational and regional variation related to sound change. We also tested only the high tone-initial words (i.e. HH, HL), excluding low tone-initial words to control for the effect of lexical pitch accent in GK.

We first built separate models for female and male speakers with all groups except for the transplants. These *full models*, predicting aspirated (vs. lenis) laryngeal category, included age (elementary, high school, university) and region (Seoul, Gyeongsang) as fixed effects, with by-subject random intercepts and slopes for both VOT and F0. Age and region were treatment coded, with Seoul and high school serving as reference levels.

To analyze the effect of region more specifically, we built a second set of *university models* using only the university students, including the transplants, with transplants serving as the reference level. This model included only a three-level region factor (Seoul, Gyeongsang, transplant), and no age factor. Statistical models were run using the lme4 package [13] in R [14].

3. RESULTS

Relative use of VOT and F0: Both SK and GK speakers of all age groups produced aspirated stops with higher F0 and longer VOT than lenis stops. In terms of the relative use of these cues, F0 was more important than VOT, as the models yielded much larger coefficients for F0 than VOT. This result differs from perception patterns reported in other studies (e.g. [12]) in that F0 here served as a stronger

cue than VOT even in GK speakers' productions, and in all age groups.

	Female	Male
	Estimate (SE)	Estimate (SE)
Intercept	**-20.64 (1.93)**	**12.57 (2.17)**
VOT	0.82 (0.44)	**1.32 (0.34)**
F0	**25.40 (2.55)**	**12.56 (1.80)**
VOT×Age_{elem}	-0.18 (0.55)	*-2.21 (0.86)*
VOT×Age_{uni}	0.39 (0.61)	1.07 (0.60)
F0×Age_{elem}	-2.78 (2.29)	**19.00 (3.32)**
F0×Age_{uni}	6.27 (2.63)*	0.41 (1.53)
VOT×$Region_{GK}$	*1.80 (0.60)*	**2.41 (0.71)**
F0×$Region_{GK}$	*8.32 (2.71)*	0.13 (1.62)
VOT×Age_{elem}×$Region_{GK}$	-0.52 (0.79)	-1.60 (0.85)
VOT×Age_{uni} ×$Region_{GK}$	-0.74 (0.86)	-1.34 (1.03)
F0×Age_{elem}×$Region_{GK}$	*-10.59 (3.43)*	-3.67 (4.36)
F0×Age_{uni}×$Region_{GK}$	-9.28 (3.71)*	0.92 (2.26)

Table 2: Regression output from the *full models*. Subscripts indicate factor levels of elementary school (elem), university (uni), and Gyeongsang (GK). (**bold** $p < .001$, *italic* $p < .01$, '*' $p < .05$,)

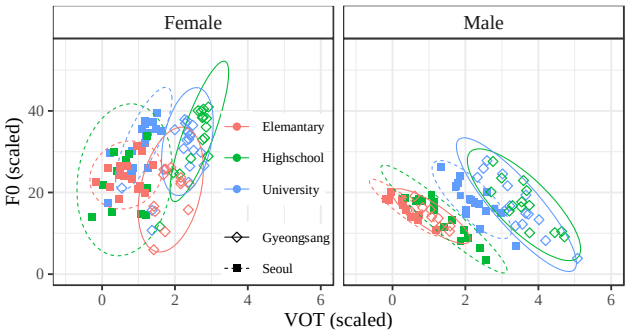

Figure 1: Female (left) and male (right) speakers' coefficient distributions for F0 and VOT from *full models*.

Figure 1 shows the distributions of individual speakers' F0 and VOT coefficients grouped by age and region, in each gender. In terms of gender, male speakers' F0 and VOT coefficients were negatively correlated (right panel; $r(92) = -.21$, $p = 0.059$), indicating a trading relationship between the two cues. In contrast, female speakers' F0 and VOT coefficients were positively correlated (left panel; $r(81) = -.46$, $p < 0.001$), suggesting that female speakers who used VOT more than others also used F0 more.

Differences between regions: As shown in Figure 1, regional differences were observed most clearly in high school speakers relative to elementary school or university speakers (see the solid and dashed green ellipses), in both genders. The regression output confirmed this trend: for males, GK high school students' VOT coefficient was significantly greater than that of their SK counterparts. For female speakers, it was not only VOT but also the F0 coefficients of GK high school students that were significantly greater than those of SK high school students. For the other age groups, regional differences were consistently found in female, but not male, speakers. GK female speakers' VOT coefficients were greater than those of SK female elementary school and university student students (solid vs. dashed blue and red ellipses). By contrast, male elementary school and university speakers did not exhibit regional differences in their VOT and F0 coefficients.

Differences among age groups: The age groups of SK and GK speakers in each gender were not consistently differentiated in terms of their VOT and F0 coefficients. While the university students were consistently different from the elementary school students (blue vs. red ellipses in Figure 1), they were not necessarily different from the high school student groups in using VOT or F0 (blue vs. green ellipses). In an analogous model run with university as the reference level age group, female university students in both regions had significantly greater F0 coefficients than the elementary student groups (SK: $\beta_{F0\times Elem} = -9.67$, SE = 2.63, $p < .001$; GK: $\beta_{F0\times Elem} = -11.01$, SE = 2.56, $p < .001$).

Similarly, male university students in both regions were differentiated from elementary school students in both acoustic dimensions, with significantly greater VOT and smaller F0 coefficients (SK: $\beta_{VOT\times Elem} = -3.47$, SE = 1.11, $p < .005$; $\beta_{F0\times Elem} = 18.96$, SE = 3.54, $p < .001$; GK: $\beta_{VOT\times Elem} = -3.70$, SE = 1.07, $p < .001$; $\beta_{F0\times Elem} = 14.5$, SE = 3.73, $p < .001$).

Between the GK university and high school students in both genders, however, neither VOT nor F0 were significantly different. In SK, however, female university students' F0 coefficient was significantly greater than those of high school students, while SK male university students' VOT coefficient was marginally greater than those of the high school students.

Differences among regions in university students: The *university model* output is given in Table 3, with the individual coefficients plotted in Figure 2. Female transplant speakers were not different from SK or GK speakers in VOT and F0. Male transplant speakers were distinguished from SK (but not GK) speakers, however, by having a significantly greater VOT coefficient.

	Female	Male
	Estimate (SE)	Estimate (SE)
Intercept	**-19.04 (2.35)**	**18.47 (1.73)**
VOT	1.78 (0.77)*	**4.40 (0.73)**
F0	**34.68 (4.57)**	**18.18 (1.93)**
VOT×$Region_{Seoul}$	-0.64 (0.92)	-2.14 (0.85)*

VOT×Region$_{GK}$	0.92 (0.89)	-0.92 (0.88)
F0×Region$_{Seoul}$	-2.74 (3.70)	-1.66 (1.57)
F0×Region$_{GK}$	-4.54 (3.60)	-2.26 (1.58)

Table 3: Regression output from *university models*. Subscripts indicate Seoul and Gyeongsang (GK) factor levels. (**bold** $p < .001$, *italic* $p < .01$, '*' $p < .05$)

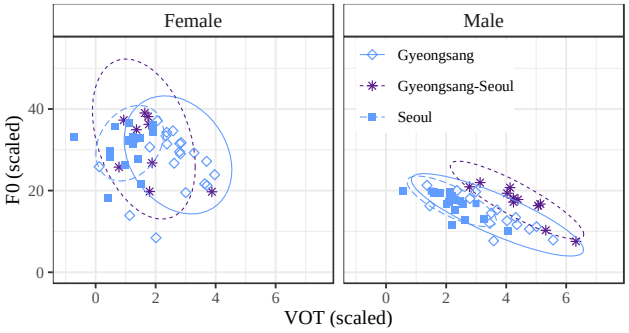

Figure 2: Female (left) and male (right) speakers' coefficient distributions of F0 and VOT from *university models*.

4. DISCUSSION

The current study examined the acoustic patterns of VOT and F0 in the Korean stops produced by Seoul and Gyeongsang speakers of three age groups: children, adolescents and young adults. We aimed to investigate whether the sub-phonemic phonetic change in the lenis-aspirated contrast would differ by age and region, and whether these results were similar to the perception patterns previously reported by [12]. We found that both the GK and SK speakers produced the newer stop variants by enhancing F0. Among the age groups, GK high school students were the most regionally distinctive by their use of VOT, but they were not the most progressive group. Notably, while the GK university students were no different from the high school students, the acoustic cues used by the transplants shifted toward a Seoul pattern.

We may draw two broader conclusions from these findings. First, the fact that the acoustic cues used in production do not necessarily mirror the acoustic cue weighting used in perception supports the idea that perception may be more flexible. [12] showed that the perception of university students seemed to reflect exposure to a wider speech community than the local variety, whether that be the standard variety SK, or a non-standard variety GK. The fact that GK university students' production did not reflect any change due to such exposure could be related to their not actually having left the GK region. This possibility is also suggested by the finding that GK transplant speakers *did* increase their use of F0 compared to local GK university students.

Second, these results lead us to consider adolescent peaks [16, 17] in sound change from a non-standard regional speaker's perspective. In the case of standard Seoul Korean, the sound change seems to be basically complete, so we may not expect an "adolescent peak" in which high school students produce the most innovative forms among age groups. But why don't we see an adolescent peak in GK, where the sound change is still ongoing? In the case of non-standard GK, high school students were not differentiated from university or elementary school students in terms of more use of F0 and/or less use of VOT. It could have to do with the fact that the sound change in question is one from another region, and the current speakers have not physically left the local area. This may be compared with the male GK transplants, who did rely on F0 more than the local GK university students. In the case of SK, the high school students did rely on VOT less than the university students, which could be interpreted as a sort of adolescent peak, although they were not differentiated from elementary school students. Thus, we find some limited evidence for adolescent peaks in the current data from multiple regions.

So, what does this say about the precedence of perception or production in an ongoing sound change? Focusing on GK, in which the sound change is still ongoing, it appears that speakers from all age groups use a more conservative acoustic cue weighting in their perception than their production. The scenario suggested by [6] (and supported by [2]) is that changes in perception will lead changes in production at the beginning of a sound change, but then as it proceeds and nears completion production will become to lead. This can be understood as speakers having advanced through the sound change but are maintaining flexible perceptual cue weightings to accommodate speakers who are more conservative. This appears to be the pattern found in the current study.

To conclude, we first observed an effect of region in the progression of the sound change in Korean stops: the sound change is mostly complete in SK, and still ongoing in GK. In terms of age, the effects are most clear when the current results are compared with findings from much older GK speakers, as in [11]. The age groups in the current study are close enough that the differences are more reflective of their immediate social setting (e.g. high school vs. university setting) than chronological age. We also saw how age and place interact in our comparison between local male GK speakers and male GK transplants living in Seoul, which supports the findings of [18]. These results suggest that although production may be less flexible than perception, physical migration may modulate this effect.

5. REFERENCES

[1] Harrington, J., Kleber, F., Reubold, U. 2008. Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study. *J. Acoust. Soc. Am.* 123(5), 2825-2835.

[2] Kuang, J., Cui, A. 2018. Relative cue weighting in production and perception of an ongoing sound change in Southern Yi. *Journal of Phonetics* 71, 194-214.

[3] Coetzee, A. W., Beddor, P. S., Shedden, K., Styler, W., Wissing, D. 2018. Plosive voicing in Afrikaans: Differential cue weighting and tonogenesis. *Journal of Phonetics* 66, 185-216.

[4] Strickler, A. 2019. Within-speaker perception and production of dialectal /aɪ/-raising. In S. Calhoun, P. Escudero, M. Tabain & P. Warren (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019* (pp. 3205-3209).

[5] Howe, P. 2017. *Tonogenesis in central dialects of Malagasy: Acoustic and perceptual evidence with implications for synchronic mechanisms of sound change* [Doctoral dissertation, Rice University].

[6] Pinget, A.-F., Kager, R., Van de Velde, H. 2020. Linking variation in perception and production in sound change: Evidence from Dutch obstruent devoicing. *Language and Speech* 63(3), 660-685.

[7] Lee, H., Holliday, J. J., Kong, E. J. 2020. Diachronic change and synchronic variation in the Korean stop laryngeal contrast. *Language and Linguistics Compass* 14(7), e12734.

[8] Kang, Y. 2014. Voice onset time merger and development of tonal contrast in Seoul Korean stops: A corpus study. *Journal of Phonetics* 45, 76-90.

[9] Lee, H., Politzer-Ahles, S., Jongman, A. 2013. Speakers of tonal and non-tonal Korean dialects use different cue weightings in the perception of the three-way laryngeal stop contrast. *Journal of Phonetics* 41, 117-132.

[10] Holliday, J. J., Kong, E. J. 2011. Dialectal variation in the acoustic correlates of Korean stops. In W. S. Lee & E. Zee (eds.), *Proceedings of the 17th International Congress on Phonetic Sciences* (pp. 878–881).

[11] Lee, H., Jongman, A. 2012. Effects of tone on the three-way laryngeal distinction in Korean: An acoustic and aerodynamic comparison of the Seoul and South Kyungsang dialects. *Journal of the International Phonetic Association 42*(2), 145-169.

[12] Kong, E. J., Holliday, J. J., Lee, H. 2022. Post-adolescent changes in the perception of regional sub-phonemic variation. *Journal of Phonetics* 90, 101114.

[13] Bates, D., Mächler, M., Bolker, B., Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software 67*(1), 1–48.

[14] R Core Team. 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

[15] Lee, H. 2020. Cross-generational acoustic comparisons of tonal Kyungsang Korean stops. *J. Acoust. Soc. Am. 148*(2) EL172-EL178.

[16] Tagliamonte, S., D'Arcy, A. 2009. Peaks beyond phonology: Adolescence, incrementation and language change. *Language 85*, 58-108

[17] Holmes-Elliott, S. 2016. Ladies First? Adolescent Peaks in a Male-Led Change. *University of Pennsylvania Working Papers in Linguistics 22*(2) 81-90.

[18] Kim, H., Jongman, A. 2022. The influence of inter-dialect contact on the Korean three-way laryngeal distinction: An acoustic comparison among Seoul Korean speakers and Gyeongsang speakers with limited and extended residence in Seoul. *Language and Speech 65*(3), 531-553.