THE EFFECT OF WORD FREQUENCY ON THE TIMECOURSE OF TONOGENESIS IN SEOUL KOREAN

Hye-Young Bang[a], *Morgan Sonderegger*[a], *Yoonjung Kang*[b], *Meghan Clayards*[a] & *Tae-Jin Yoon*[c]

[a]McGill University, Candada; [b]University of Toronto Scarborough, Canada; [c]Sungshin Women's University, Korea
hye-young.bang@mail.mcgill.ca; morgan.sonderegger@mcgill.ca; yoonjung.kang@utoronto.ca; meghan.clayards@mcgill.ca; tyoon@sungshin.ac.kr

ABSTRACT

Seoul Korean is currently undergoing a tonogenetic sound change wherein the traditional consonantal VOT cue has been replaced by the previously intrinsic f0 of the following vowel. This study makes use of a recently available apparent-time corpus of speech to examine how this change has unfolded across the lexicon. In particular, we examine the effect of word frequency to determine whether the origin of the change can be found in reduction of high frequency (presumably hypo-articulated) words, or enhancement of low frequency (presumably hyper-articulated) words. Our results suggest that tonogenesis in Seoul Korean originated in words with high frequency, indicating a likely reduction-driven sound change. Furthermore, we find a parallel change between VOT and f0 over time, which signals that the loss of the VOT contrast proceeds in tandem with the enhancement of the f0 contrast, largely consistent with theories of adaptive behaviour of sound change.

Keywords: tonogenesis, Korean, sound change, VOT, speech corpora

1. INTRODUCTION

Seoul Korean (SK) has a typologically-unusual system of three voiceless stop series (tense, lax, aspirated), which until recently were contrasted primarily by VOT differences (tense $<$ lax $<$ aspirated), and secondarily by f_0 differences (lax $<$ tense, aspirated) [6]. Recent studies converge to show that SK is undergoing a tonogenetic sound change, whereby the tone (i.e. f_0) of the following vowel replaces VOT of the stop as a primary cue for the aspirated-lax stop contrast [14, 24]. While older speakers distinguish aspirated and lax stops using VOT, younger speakers produce aspirated and lax stops with comparable VOTs, and instead use f_0 of the following vowel (high tone following aspirated stops, low tone following lax stops) to signal the contrast. That is, the VOT difference has been lost, and the f_0 difference has increased. While tonogenesis is a common type of sound change of considerable theoretical interest [17], relatively little is known about how and why the trade-off between the consonantal and vocalic cues unfolds over time, spreading through the language and the speech community. In particular, theories differ on whether exaggeration of the vocalic cue leads to collapse of the consonantal cue, or reduction of the consonantal cue leads to exaggeration of the vocalic cue [9, 18, 24]. The fact that this case of tonogenesis is currently in progress, and the availability of a large apparent time corpus in which it can be observed, make Seoul Korean an ideal case study to address such questions.

Specifically, we examine whether the timecourse of this sound change (and tonogenesis more generally) can be understood by reference to (a) the continuum of hypo/hyper-articulated speech, in which contrasts between sounds are reduced/exaggerated, and (b) adaptive behavior by speaker/listeners [18, 19, 20]; three possible scenarios are given in Table1. We test whether this sound change is driven by the *hypoarticulation of the VOT distinction* or by the *hyperarticulation of the* f_0 *distinction*, by examining if and how the changes in VOT and f_0 in aspirated and lax stops affect high-frequency words and low-frequency words differently (c.f. [21]). High frequency words are prone to reductive hypoarticulation and contrast neutralization compared to low frequency words [1, 12]. If the sound change is hypoarticulation driven (Hypothesis 1 or 3), we expect the sound change to be more advanced in high frequency words than low frequency words [5], at least in the VOT dimension. By contrast, if the sound change is driven by hyperarticulation of f_0 (Hypothesis 2 or 3), we expect the sound change to be more advanced in low frequency words than in high frequency words [28], at least in the f_0 dimension.

Furthermore, if loss of VOT distinction and enhancement of f_0 distinction are linked by adaptive compensation of each other [18, 20], we expect

Table 1: Possible relationships between word frequency and VOT reduction/f_0 enhancement. "High $>$ low" means "change in lax/aspirated contrast more advanced in high-frequency words compared to low-frequency words".

Hypothesis	VOT reduction	f_0 enhancement
1. hypoartic.-driven, adaptive	high > low	high > low
2. hyperartic.-driven, adaptive	low > high	low > high
3. hypo/hyperartic.-driven, non-adaptive	high > low	low > high

them to proceed in tandem, both affecting the same high frequency words first (Hypothesis 1) or both the same low frequency words first (Hypothesis 2), depending on whether hyper- or hypoarticulation drives the change. Such parallel frequency effects on VOT reduction and f_0 enhancement are unexpected if VOT and f_0 are independently affected by synchronic pressure for hypo- or hyper-articulation in high and low frequency words (Hypothesis 3).

In this paper, we evaluate Hypotheses 1–3 by addressing two empirical questions in a large apparent-time corpus of Seoul Korean: how have VOT and f_0 of aspirated and lax stops changed over time in SK, and how is this change modulated by word frequency for each cue?

2. DATA & METHODS

The data come from The Speech Corpus of Reading-Style Standard Korean [25], the same corpus used by [14]. The corpus contains a total of 930 sentences read by 120 Seoul dialect speakers (ages 19–71). The corpus had been force-aligned [26, 27] to facilitate data collection. We extracted tokens from 118 speakers and 82 *items*, defined as a particular occurrence of a word in a sentence, chosen to achieve roughly equal distribution among different values of word-level variables (laryngeal category, place of articulation, etc.; see below).[1] As this sound change is only occurring at the Accentual Phrase and higher prosodic domains [10], sentence-medial tokens were only included if followed by a force-aligned pause for a given speaker following the previous finding that a pause very often signals an Intonational Phrase boundary [11].

The resulting dataset consists of 5888 stops, for which VOT and f_0 were measured. VOT was measured by manually correcting automatic measurements made by AutoVOT [15]. f_0 values were measured at vowel midpoints using a Praat script [3], followed by manual correction of outliers. In analyses below, VOT is log-transformed (because its distribution is right-skewed), and f_0 values (in semitones) for each speaker are normalized by subtracting their mean value. Word frequencies were obtained from [13], and log-transformed.

VOT and f_0 are both modeled as a function of a number of variables which are properties of speakers, items, and utterances. Speaker YEAR OF BIRTH (YOB) is included, to account for change over time, as is GENDER, as previous work suggests female speakers lead this change [14]. Of main interest are speaker YOB (to account for change over time), FREQUENCY of the word corresponding to each item, and the word's LARYNGEAL category (3 levels: tense, lax, aspirated). We also include various properties of items and utterances (following VOWEL HEIGHT, PLACE OF ARTICULATION, SENTENCE POSITION, SPEAKING RATE), which are all expected to affect VOT and f_0 [7, 8, 16]. Below, we discuss only the results of this model involving FREQUENCY, LARYNGEAL, and YOB; note that the inclusion of other variables guards againt the possibility that our results are in fact due to some factor which correlates with one of these three variables. (For example, older speakers tend to have lower SPEAKING RATE.)

We modeled VOT and f_0 as a function of all the variables above using two linear mixed-effects models, fitted in R using the `lme4` package [2, 23]. In these models, YOB was coded as a nonlinear spline, with degrees of freedom chosen based on exploratory plots. Helmert coding was used for all factors and all continuous variables were centered, to minimize collinearity. In particular, for LARYNGEAL, the two contrasts correspond to the differences between tense and non–tense (LARYNGEAL_1), and between aspirated and lax (LARYNGEAL_2). The models included fixed effects for: main effects of all variables above; interactions between LARYNGEAL and YOB (to describe how the cues to each stop category are changing over time), and between LARYNGEAL and all other variables (to capture what factors promote VOT reduction or f_0 enhancement), as well as further interactions not discussed here for space.[2] The crucial term for testing Hypotheses 1–3 is the LARYNGEAL_2:FREQUENCY interaction, which describes whether and how the VOT and f_0 difference between aspirated and lax stops depends on word frequency. In other words, this term tests if VOT contrast reduction and f0 contrast enhancement are more advanced in high-frequency words or in low-frequency words.

The model included by-item and by-speaker random intercepts, as well as all possible random slopes corresponding to fixed-effect terms discussed below, to allow for variability among speakers and items. Correlations between random effects were omitted as models with correlations included did not converge.

3. RESULTS

The fixed-effect terms of the model relevant for understanding the results involving word frequency, the aspirated/lax difference (LARYNGEAL$_2$) and year of birth (YOB$_1$, YOB$_2$: components of the spline) are given in Table 2^3.

3.1. Overall timecourse of change

Before turning to the effect of frequency, we first establish how the change is proceeding across the speech community in this corpus—averaging across all words. Both the VOT and f_0 models show strong and significant main effects of LARYNGEAL$_2$ and interactions with YOB (LARYNGEAL$_2$, YOB$_1$:LARYNGEAL$_2$, YOB$_1$:LARYNGEAL$_2$ rows in Table 2) As illustrated in Fig. 1, speakers use VOT progressively less and f_0 progressively more to differentiate aspirated and lax stops, while older speakers rely more on VOT and less on f_0. The timecourse of change differs greatly by speaker gender (significant GENDER and GENDER:LARYNGEAL$_2$ rows in Table 2): male speakers rely on VOT significantly more than female speakers, and female speakers rely significantly more on f_0, as suggested by Fig. 1.

Overall, VOT loss and f_0 enhancement proceed in parallel over time, with female speakers leading the sound change. These findings replicate previous work on this sound change in Seoul Korean, in particular [14], which used the same corpus analyzed in the current study (but with a significantly smaller dataset: 1250 tokens & 12 items).

3.2. Frequency

We now turn to the question of whether and how word frequency modulates the sound change described in the previous section. In both models, the difference between aspirated and lax stops is modulated by word frequency (Table 2, FREQUENCY:LARYNGEAL$_2$ rows). For VOT, the difference between aspirated and lax stops significantly decreases as frequency increases ($\hat{\beta} < 0$, $p = 0.025$), while for f_0, this difference significantly increases as frequency increases ($\hat{\beta} > 0$, $p = 0.027$). Taken together with the overall reduction of VOT and expansion of f_0 over time noted above, these results mean the model predicts that *both VOT reduction and f_0 enhancement take place earlier for high-frequency words than for low-frequency words*. This prediction is illustrated in Fig. 2 (panels 2 and 4), where the lines corresponding to aspirated and lax stops merge earlier (for VOT) or separate earlier (for f_0) for words with progressively higher frequency.

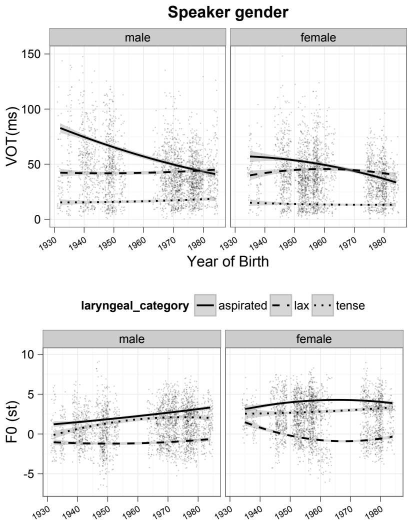


Figure 1: Empirical plots of VOT (top) and f_0 (bottom) versus speaker gender & laryngeal category. The lines and shadings show a quadratic fit and 95% confidence intervals (CIs)

The fact that VOT contrast is smaller but f_0 contrast is greater for high-frequency words suggests that tonogenesis in SK is led by words with high frequency, while shifting from VOT to f_0 is gradually diffusing across the lexicon. Similarly to the effects of age and gender, the parallel change between VOT contrast and f_0 contrast shown in Fig. 2 suggests that word frequency effects impact both cues at the same time. This pattern is thus consistent with Hypothesis 1 in Table 1.

4. DISCUSSION

This paper has examined how and why tonogenesis spreads through the lexicon by probing the effect of word frequency in Seoul Korean. The apparent-time evidence from the NIKL corpus suggests that VOT reduction and f_0 enhancement both began earlier for high frequency words than for low frequency words, and are proceeding in parallel.

The VOT finding supports the hypothesis that this sound change is driven by hypoarticulation of the VOT contrast (Hypotheses 1 and 3), and is consistent with a range of theoretical accounts that provide a link between word frequency and the actuation and spread of sound change, via the effect of

Table 2: Fixed-effect coefficient estimates, standard errors, t-values, and significances, for selected terms from the models of VOT and f_0. (See text.)

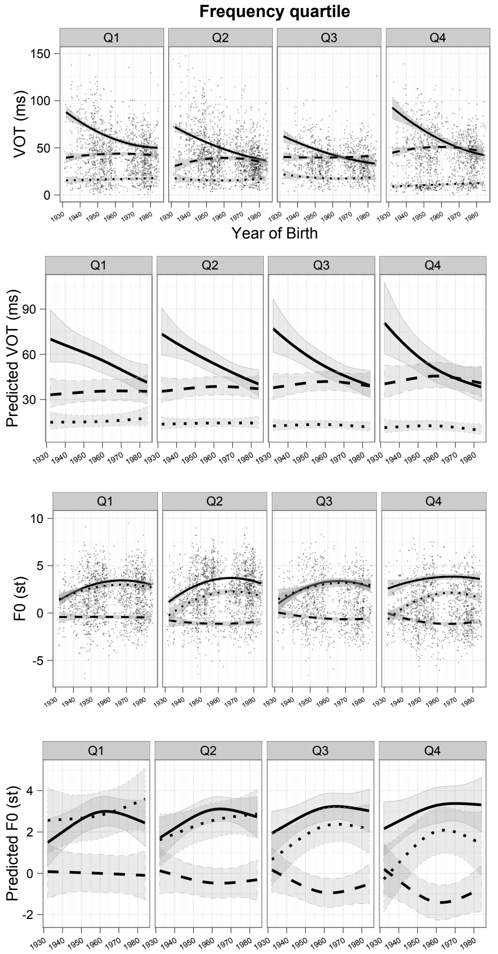
VOT:

Fixed Effects	$\hat{\beta}$	se($\hat{\beta}$)	t	p
(Intercept)	3.392	0.046	73.37	< 0.0001
YOB$_1$	−0.047	0.08	−0.58	0.566
YOB$_2$	−0.089	0.11	−0.79	0.432
GENDER (male)	0.135	0.03	5.2	< 0.0001
LARYNGEAL$_2$ (asp. - lax)	0.226	0.08	2.99	0.003
LARYNGEAL$_2$: GENDER	0.173	0.03	5.37	< 0.0001
YOB$_1$: LARYNGEAL$_2$	−0.484	0.11	−4.33	< 0.0001
YOB$_2$: LARYNGEAL$_2$	0.164	0.15	1.08	0.282
FREQUENCY : LARYNGEAL$_2$	−0.243	0.11	−2.27	0.025

f_0:

Fixed Effects	$\hat{\beta}$	se($\hat{\beta}$)	t	p
(Intercept)	1.781	0.24	7.35	< 0.0001
YOB$_1$	0.683	0.45	1.53	0.129
YOB$_2$	−0.5	0.61	−0.82	0.417
GENDER (male)	−1.15	0.14	−7.94	< 0.0001
LARYNGEAL$_2$ (asp. - lax)	4.335	0.39	10.98	< 0.0001
LARYNGEAL$_2$: GENDER	−1.02	0.2	−5.03	< 0.0001
YOB$_1$: LARYNGEAL$_2$	2.611	0.65	4.03	< 0.0001
YOB$_2$: LARYNGEAL$_2$	−2.581	0.89	−2.9	0.004
FREQUENCY : LARYNGEAL$_2$	1.036	0.46	2.26	0.027

Figure 2: Empirical plots of VOT change (panels 1–2) and f_0 change (panels 3–4) as a function of word frequency & laryngeal category. Lines show a quadratic smooth to empirical data (Panels 1, 3) or the model-predicted effect of YEAR OF BIRTH (Panels 2, 4); shadings are 95% CIs as in Fig 1. *Q1–Q4* refer to word frequency quartiles from lowest (Q1) to highest (Q4).



repeated usage, gestural reduction, or overlapping gestures [4, 5, 22]; all would be expected to cause more VOT reduction in higher-frequency words.

However, hypoarticulation cannot be the only factor at play, because this would predict a frequency effect on f_0 (Hypothesis 3) in the opposite direction of our findings (it predicts a smaller f_0 difference for high-frequency words). The fact that f_0 enhancement and VOT reduction proceed in parallel, even within subsets of the lexicon (defined by frequency), suggests that f_0 is being enhanced as an adaptive response to VOT reduction, to preserve the number of laryngeal contrasts in the language. This finding is consistent with theories of sound change such as that of Lindblom et al. [19, 20], which argues that languages develop in an economical way through continuous negotiation between articulatory ease and perceptual contrast.

Thus, the VOT and f_0 findings together support Hypothesis 1. This account converges with a recent computational study by Kirby [18] which addresses the question of why sound change in SK stops is unfolding in this particular way. Based on simulations of a community of SK speaker/hearer agents under different assumptions, Kirby argues that only assuming that agents have *both* a "bias" in production (e.g., VOT reduction) *and* an adaptive response of "enhancement" (e.g., f_0 enhancement) results in the change observed in SK (rather than e.g. merger). Thus, both our findings and previous work offer converging evidence for Hypothesis 1: tonogenesis in Seoul Korean is driven by a combination of hypoarticulation-driven reduction in one dimension (VOT), and adaptive expansion in another dimension (f_0) to maintain the contrast.

5. REFERENCES

[1] Aylett, M., Turk, A. 2004. The Smooth Signal Redundancy Hypothesis: A Functional Explanation for Relationships between Redundancy, Prosodic Prominence, and Duration in Spontaneous Speech. *Lang. Speech* 47(1), 31–56.

[2] Bates, D., Maechler, M., Bolker, B., Walker, S. 2014. *lme4: Linear mixed-effects models using Eigen and S4*. R package version 3.1.2.

[3] Boersma, P., Weenink, D. 2013. Praat: doing phonetics by computer (Version 5.3. 57). University of Amsterdam. http://www.praat.org/.

[4] Browman, C. P., Goldstein, L. 1991. Gestural Structures: Distinctiveness, phonological processes and historical change. In: Mattingly, I. G., Studdert-Kennedy, M., (eds), *Modularity and the motor theory of speech perception*. Hillsdale, N.J.: Erlbaum 313–338.

[5] Bybee, J. 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Lang. Var. Change* 14, 261–290.

[6] Cho, T., Jun, S.-A., Ladefoged, P. 2002. Acoustic and aerodynamic correlates of Korean stops and fricatives. *J. Phon.* 30(2), 193–228.

[7] Cho, T., Ladefoged, P. Jan. 1999. Variation and universals in VOT: evidence from 18 languages. *J. Phon* 27(2), 207–229.

[8] Higgins, M. B., Netsell, R., Schulte, L. 1998. Vowel-related differences in laryngeal articulatory and phonatory function. *J. Speech Lang. Hear. Res.* 41(4), 712–724.

[9] Hombert, J.-M., Ohala, J., Ewan, W. 1979. Phonetic explanations for the development of tones. *Language* 55, 37–58.

[10] Jun, S.-A. 1993. *The Phonetics and Phonology of Korean prosody*. PhD thesis The Ohio State University.

[11] Jun, S.-A. 1998. The Accentual phrase in the Korean prosodic hierarchy. *Phonology* 15, 189–226.

[12] Jurafsky, D., Bell, A., Gregory, M., Raymond, W. D. 2001. Probabilistic Relations between Words: Evidence from Reduction in Lexical Production. In: Bybee, J., Hopper, P., (eds), *Frequency and the emergence of linguistic structure*. Amsterdam: Benjamins 229–254.

[13] KAIST, 1999. KAIST Concordance Program. http://semanticweb.kaist.ac.kr/research/kcp/.

[14] Kang, Y. 2014. Voice Onset Time merger and development of tonal contrast in Seoul Korean stops: A corpus study. *J. Phon.* 45, 76–90.

[15] Keshet, J., Sonderegger, M., Knowles, T. 2014. AutoVOT: A tool for automatic measurement of voice onset time using discriminative structured prediction [Computer program]. Version 0.91. https://github.com/mlml/autovot/.

[16] Kessinger, R. H., Blumstein, S. E. 1997. Effects of speaking rate on voice-onset time in Thai, French, and English. *J. Phon* 25(2), 143–168.

[17] Kingston, J. 2011. Tonogenesis. In: van Oostendorp, M., Ewen, C. J., Hume, E., Rice, K., (eds), *The Blackwell Companion to Phonology*. Oxford: Blackwell Publishing 2304–2333.

[18] Kirby, J. 2013. The role of probabilistic enhancement in phonologization. In: Yu, A., (ed), *Origins of sound patterns: approaches to phonologization*. Oxford: Oxford University Press 228–246.

[19] Lindblom, B. 1990. Explaining variation: A sketch of the H and H theory. In: Hardcastle, W., Marchal, A., (eds), *Speech Production and Speech Modelling*. Dordrecht: Kluwer 403–439.

[20] Lindblom, B., Guion, S., Hura, S., Moon, S.-J., Willerman, R. 1995. Is sound change adaptive? *Rivista di linguistica* 7, 5–36.

[21] Phillips, B. S. 2006. *Word frequency and lexical diffusion*. New York: Palgrave Macmillan.

[22] Pierrehumbert, J. B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In: Bybee, J., Hopper, P., (eds), *Frequency and the emergence of linguistic structure*. Amsterdam: Benjamins 137–158.

[23] R Core Team, 2014. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria.

[24] Silva, D. J. 2006. Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology* 23, 287–308.

[25] The National Institute of the Korean Language, 2005. *A speech corpus of reading-style standard Korean [DVDs]*. Seoul: NIKL.

[26] Yoon, T.-J. to appear. A corpus-based study on the layered duration in Standard Korean. In: Kenstowicz, M., Levina, T., Masuda, R., (eds), *Japanese/Korean Linguistics 23*. Chicago: University of Chicago Press 139–145.

[27] Yoon, T.-J., Kang, Y. 2014. Monophthong Analysis on a Large-scale Speech Corpus of Read-Style Korean. *Speech Sciences* 6.

[28] Zhao, Y., Jurafsky, D. 2009. The effect of lexical frequency and Lombard reflex on tone hyperarticulation. *J. Phon.* 37(2), 231–247.

---

[1] We refer to “items” instead of “words” because items differ by SENTENCE POSITION, and some words are read repeatedly in different sentences.

[2] The formula for the fixed and random effects of each model, in `lme4` syntax, is: `(RATE + GENDER) * LARYNGEAL + RCS(YOB, 3) * LARYNGEAL * (PLACE + FREQUENCY + VOWEL HEIGHT + SENTENCE POSITION) + (1 + (LARYNGEAL`$_1$` + LARYNGEAL`$_2$`) * (FREQUENCY + PLACE`$_1$` + PLACE`$_2$` + VOWEL HEIGHT + SENTENCE POSITION)||SPEAKER) + (1 + YOB`$_1$` + YOB`$_2$`||ITEM))`

[3] Other terms are omitted for lack of space. Note that the coding scheme used for factors and the fact that continuous variables have been centered means that the rows of Table 2 can be interpreted separately from other terms in the model (which are not shown), as the effects when all other variables are averaged over.