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

Hye-Young Banga, Morgan Sondereggera, Yoonjung Kangb, Meghan Clayards & Tae-Jin Yoonb

aMcGill University, Canada; bUniversity of Toronto Scarborough, Canada; cSungshin 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Seoul Korean is currently undergoing a tonogenetic sound change wherein the traditional consonantal VOT cue has been replaced by the previously intrinsic f0 of the following vowel.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 f0 differences (lax < tense, aspirated) [6]. Recent studies converge to show that SK is undergoing a tonogenetic sound change, whereby the tone (i.e. f0) 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 f0 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 f0 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 Table 1. We test whether this sound change is driven by the hypoarticulation of the VOT distinction or by the hyperarticulation of the f0 distinction, by examining if and how the changes in VOT and f0 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 f0 (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 f0 dimension.

Furthermore, if loss of VOT distinction and enhancement of f0 distinction are linked by adaptive compensation of each other [18, 20], we expect
Table 1: Possible relationships between word frequency and VOT reduction/f0 enhancement. “High > low” means “change in lax/aspirated contrast more advanced in high-frequency words compared to low-frequency words.”

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>VOT reduction</th>
<th>f0  enhancement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. hypoartic.-driven, adaptive</td>
<td>high &gt; low</td>
<td>high &gt; high</td>
</tr>
<tr>
<td>2. hyperartic.-driven, adaptive</td>
<td>low &gt; high</td>
<td>low &gt; low</td>
</tr>
<tr>
<td>3. hypo/hyperartic.-driven, non-adaptive</td>
<td>high &gt; low</td>
<td>low &gt; high</td>
</tr>
</tbody>
</table>

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 f0 enhancement are unexpected if VOT and f0 are independently affected by synchronic pressure for hypo- or hyper-articulation in high and 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 f0 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 Accidental 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 f0 were measured. VOT was measured by manually correcting automatic measurements made by AutoVOT [15]. f0 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 f0 values (in semitones) for each speaker are normalized by subtracting their mean value. Word frequencies were obtained from [13], and log-transformed.

VOT and f0 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 f0 [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 against 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 f0 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. Helmhert 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 (LARYNGEAL1), and between aspirated and lax (LARYNGEAL2). 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 f0 enhancement), as well as further interactions not discussed here for space.2 The crucial term for testing Hypotheses 1–3 is the LARYNGEAL2:FREQUENCY interaction, which describes whether and how the VOT and f0 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 (\texttt{LARYNGEAL2}) and year of birth (\texttt{YOB1}, \texttt{YOB2}; components of the spline) are given in Table 2.

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 \texttt{LARYNGEAL2} and interactions with \texttt{YOB} (\texttt{LARYNGEAL2}, \texttt{YOB1: LARYNGEAL2}, \texttt{YOB1: LARYNGEAL2} 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 \texttt{GENDER} and \texttt{GENDER: LARYNGEAL2} 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, \texttt{FREQUENCY: LARYNGEAL2} 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.

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 actualization 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 f₀. (See text.)

### VOT:

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>( \hat{\beta} )</th>
<th>se(( \hat{\beta} ))</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>3.392</td>
<td>0.046</td>
<td>73.37</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>YOB₁</td>
<td>-0.047</td>
<td>0.08</td>
<td>-5.58</td>
<td>0.566</td>
</tr>
<tr>
<td>YOB₂</td>
<td>-0.135</td>
<td>0.08</td>
<td>106</td>
<td>0.807</td>
</tr>
<tr>
<td>LARYNGEAL₂ (asp. - lax)</td>
<td>0.226</td>
<td>0.08</td>
<td>29.99</td>
<td>0.0001</td>
</tr>
<tr>
<td>LARYNGEAL₂ : GENDER</td>
<td>0.173</td>
<td>0.03</td>
<td>5.37</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>YOB₁ : LARYNGEAL₂</td>
<td>-0.484</td>
<td>0.11</td>
<td>-4.33</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>YOB₂ : LARYNGEAL₂</td>
<td>0.164</td>
<td>0.15</td>
<td>1.08</td>
<td>0.282</td>
</tr>
<tr>
<td>FREQUENCY : LARYNGEAL₂</td>
<td>-0.243</td>
<td>0.11</td>
<td>-2.27</td>
<td>0.025</td>
</tr>
</tbody>
</table>

### f₀:

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>( \hat{\beta} )</th>
<th>se(( \hat{\beta} ))</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.781</td>
<td>0.24</td>
<td>7.35</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>YOB₁</td>
<td>0.683</td>
<td>0.45</td>
<td>1.53</td>
<td>0.129</td>
</tr>
<tr>
<td>YOB₂</td>
<td>-0.5</td>
<td>0.11</td>
<td>-2.2</td>
<td>0.029</td>
</tr>
<tr>
<td>LARYNGEAL₂ (asp. - lax)</td>
<td>4.335</td>
<td>0.19</td>
<td>10.98</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>LARYNGEAL₂ : GENDER</td>
<td>-1.02</td>
<td>0.2</td>
<td>-5.03</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>YOB₁ : LARYNGEAL₂</td>
<td>2.611</td>
<td>0.65</td>
<td>4.03</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>YOB₂ : LARYNGEAL₂</td>
<td>-2.581</td>
<td>0.89</td>
<td>-2.9</td>
<td>0.004</td>
</tr>
<tr>
<td>FREQUENCY : LARYNGEAL₂</td>
<td>1.036</td>
<td>0.46</td>
<td>2.26</td>
<td>0.027</td>
</tr>
</tbody>
</table>

Figure 2: Empirical plots of VOT change (panels 1–2) and f₀ 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 f₀ (Hypothesis 3) in the opposite direction of our findings (it predicts a smaller f₀ difference for high-frequency words). The fact that f₀ enhancement and VOT reduction proceed in parallel, even within subsets of the lexicon (defined by frequency), suggests that f₀ 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 f₀ 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., f₀ 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 (f₀) to maintain the contrast.
5. REFERENCES


---

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: \((\text{RATE} + \text{GENDER}) \ast \text{LARYNGEAL} + \text{FREQUENCY} + \text{PHONEME} + \text{SENTENCE POSITION}) + (1 + (\text{LARYNGEAL} + \text{LARYNGEAL}) \ast (\text{FREQUENCY} + \text{PLACE}) + \text{PLACE} + \text{PHONEME} + \text{SENTENCE POSITION}) \ast \text{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.

---