

# The emergence, progress, and impact of sound change in progress in Seoul Korean: implications for mechanisms of tonogenesis

Hye-Young Bang<sup>\*1</sup>, Morgan Sonderegger<sup>1,2</sup>, Yoonjung Kang<sup>3,4</sup>, Meghan Clayards<sup>1,2</sup>, and Tae-Jin Yoon<sup>5</sup>

<sup>1</sup>*Department of Linguistics, McGill University, Canada*

<sup>2</sup>*Centre for Research on Brain, Language, and Music, McGill University, Canada*

<sup>3</sup>*Centre for French and Linguistics, University of Toronto Scarborough, Canada*

<sup>4</sup>*Department of Linguistics, University of Toronto, Canada*

<sup>5</sup>*Department of English Language and Literature, Sungshin Women's University, Korea*

## Abstract

This study examines the origin, progression, and impact of a sound change in Seoul Korean where the primary cue to a stop contrast in phrase-initial position is shifting from VOT to f0. Because it shares similarities with the initial phase of tonogenesis, investigating this “quasi-tonogenetic” sound change provides insight into the nature of the emergence of contrastive f0 in “tonogenetic” sound changes more generally. Using a dataset from a large apparent-time corpus of Seoul Korean, we built mixed-effects regression models of VOT and f0 to examine the time-course of change, focusing on word frequency and vowel height effects. We found that both VOT contrast reduction and f0 contrast enhancement are more advanced in high-frequency words and in stops before non-high vowels, indicating that the change is spreading across words and phonetic contexts in parallel. Furthermore, speakers suppress non-contrastive variation in f0 as f0 emerges as a primary cue. Our findings suggest that one impetus for tonogenetic change is production bias coupled with an adaptive link between the cues. We further discuss the role of Korean intonational phonology on f0 which may help explain why the phonetic precondition leads to change in Seoul Korean but not in other languages.

**Key Words**— Tonogenesis, Transphonologization, Vowel intrinsic f0, Word frequency, Cue tradeoff, Sound change, Korean

## 1 Introduction

Tonogenesis (Matisoff, 1973) is a linguistic process whereby redundant pitch patterns become phonologized and contrastive over time. It is a common type of sound change, and has occurred across many genetically unrelated languages (Hombert, Ohala & Ewan, 1979; Kingston, 2011). Tonogenesis has its origins in various phonetic sources (Kingston, 2011)

---

\*Correspondence address: 1085 Dr. Penfield, Room 111, Montreal, Quebec, H3A1A7; Tel.: 1 514-398-4222; Fax: 1 514-398-7088; E-mail address: hye-young.bang@mail.mcgill.ca

but the most common and well-documented source of tonogenesis is the  $f_0$  differences in vowels adjacent to consonants with different laryngeal settings developing into contrastive tone (Hombert, 1977; Hombert et al., 1979; Löfqvist, Baer, McGarr & Story, 1989). When the traditional consonantal cue is lost as the tonal contrast emerges, *transphonologization* is said to have taken place (e.g. Hagège & Haudricourt, 1978; Hombert et al., 1979; Hyman, 1976; Kingston, 2011; Maran, 1973). Transphonologization is often assumed to have a functional motivation of contrast maintenance (Hyman, 2008, p. 387).

Many phonetic studies on tonogenetic sound change examine languages in a transitional state from consonantal to tonal contrast (e.g. Chen, 2011; DiCanio, 2012; Mazaudon & Michaud, 2009; Misnadin, Kirby & Remijsen, 2015). Most studies documenting the diachronic trajectory of tonogenesis do so indirectly by comparing different endpoints of sound change in related languages or dialects (Kingston, 2005; Purcell, Villegas & Young, 1978; Svantesson & House, 2006). In addition, a growing body of instrumental studies examine variation within a single speech community to track sound change in progress (Coetzee, Beddor & Wissing, 2014 (as cited in Beddor, 2015), Abramson, L-Thongkum & Nye, 2005; Hyslop, 2009; Kirby, 2014).

For example, Kirby (2014) examines production and perception for an ongoing sound change in Phnom Penh Khmer, where /r/ in consonant clusters in onset position is being replaced by other acoustic cues associated with the following vowel (e.g. breathiness,  $f_0$  contour). The origin of the sound change is argued to lie in perceptual reanalysis of colloquial speech variants. Coetzee et al. (2014) examine an emergent tonogenetic sound change in Afrikaans, which traditionally contrasted prevoiced and voiceless unaspirated stop series in word initial position. However, in present-day Afrikaans VOT is similar for the two stop series, which now differ primarily in  $f_0$ . The focus of this body of work is, however, limited to either the precondition or origin of change at the language level or its spread at the community level.

Building on this existing literature, the current study focuses on Seoul Korean as a case study for understanding the broader pathway of a sound change which bears similarities to cases of tonogenesis, using a large corpus dataset. We address how this sound change originates, progresses, and impacts other aspects of the linguistic system. Seoul Korean provides a rich empirical foundation for understanding tonogenetic sound changes, for several reasons. First, a sound change is currently in progress whereby the primary cue to the aspirated/lax stop distinction in phrase-initial position is shifting from VOT to  $f_0$  over time. (Korean has a three-way aspirated/lax/tense stop contrast, discussed below.) We call this ongoing change *quasi-tonogenesis* because the change does not to date exhibit all features of tonogenesis, where lexical tonal contrast develops from consonant-induced  $f_0$  distinction. The change affects only sounds at the left edge of the accentual phrase (AP) and higher prosodic domains, conditioned by Korean intonational phonology (Jun, 1996, 1998, 2005) (see Section 5.4). Hence, in present day Seoul Korean, for speakers where this change has occurred, high/low tone differentiates the meaning of relevant lexical items only in phrase-initial position. For example, the minimal pair [p<sup>h</sup>al] ‘arm’ vs. [pal] ‘foot’ (where [p] is used for a lax stop) is realized approximately as [pál] vs. [pàl] phrase-initially, while the same words are distinguished by the traditional consonantal cues in phrase-medial position.<sup>1</sup> Despite the fact that  $f_0$  cannot be used to mark arbitrary syllables as H/L in Seoul Korean, we make reference to the tonogenesis literature because we believe our results have implications for a better understanding of tonogenesis. The change in Seoul Korean

---

<sup>1</sup>IPA symbols indicate approximate phonetic realizations, based on previous literature on this sound change discussed below. The use of [pál] in particular should not be taken to indicate total absence of aspiration.

essentially exhibits the same type of transphonologization we find in cases of ‘tonogenesis’ reported in the literature (e.g. Khmer, Afrikaans), where  $f_0$  shifts from a redundant phonetic property of a laryngeal contrast to a primary cue. Furthermore, there is a large phonetic literature on laryngeal contrasts in Seoul Korean and a large apparent-time corpus (The National Institute of the Korean Language, 2005) spanning much of the time period over which the change has occurred. For all these reasons, Seoul Korean is an ideal case study for better understanding the pathway and mechanism of tonogenetic sound change.

Seoul Korean has a three-way laryngeal contrast of *tense* (or *fortis*), *lax* (or *lenis*), and *aspirated*. When described across all speakers of different ages, the three categories are contrasted by a combination of acoustic cues: primarily VOT and  $f_0$  on the following vowel, and also closure duration,  $F_1$  trajectory, and breathiness (Cho, Jun & Ladefoged, 2002; Cho & Keating, 2001; Hardcastle, 1973; Kang & Guion, 2008; Lee & Jongman, 2012; Lisker & Abramson, 1964; Park, 2002). In traditional descriptions, in phrase-initial position, aspirated, lax, and tense stops have progressively shorter VOT, and  $f_0$  on the following vowel is higher for aspirated and tense stops than for lax stops. The contrast between lax and aspirated stops—which is of main interest here—was traditionally realized primarily using VOT with  $f_0$  playing a secondary role (Han & Weitzman, 1967, 1965; Hardcastle, 1973; Kang & Han, 2013; Kim, 1965). For example, Han & Weitzman (1967) found that  $f_0$  values for all three categories overlapped significantly, and Kang & Han (2013) found that a 41-year-old speaker recorded in the 1930s realized the aspirated/lax distinction exclusively using VOT. However, the VOT difference between lax and aspirated stops reported in more recent studies is much smaller compared to those reported for the 1930s–1960s, while the  $f_0$  difference has increased (Beckman, Li, Kong & Edwards, 2014; Silva, 2002).  $f_0$  is also the primary perceptual cue to the lax/aspirated stop contrast in present-day Seoul Korean (Kim, Beddor & Horrocks, 2002; Kong, Beckman & Edwards, 2011; Lee, Politzer-Ahles & Jongman, 2013).

More direct evidence for this sound change has come from apparent-time studies (Bailey, Wikle & Tillery, 1993; Weinreich, Labov & Herzog, 1968) that map out the diachronic change by comparing the realization of aspirated and lax stops among different age groups of Seoul Korean speakers (Kang & Guion, 2008; Kang, 2014; Kang & Nagy, 2016; Silva, 2006; Wright, 2007), or from meta-analysis of studies spanning 60 years (Beckman et al., 2014). These studies have all found that the VOT contrast between aspirated and lax stops is reducing in younger speakers’ speech, while some (but not all) also found that the  $f_0$  contrast is similarly increasing. Kang & Han (2013) examined the lifespan change of a single male speaker of Seoul Korean by comparing his stop productions recorded in 1935 and 2005 (ages 11 & 81), and found change in the direction of the community: the speaker used a greater aspirated/lax stop  $f_0$  contrast in 2005. While based on a single speaker, this finding suggests that age-dependent variation in contemporary Seoul Korean cannot be an artifact of *age-grading* (Wagner, 2012), where speakers adopt age-appropriate speech patterns as they age. Given the attested lifespan change, the apparent time data if anything underestimate the rate of ongoing change in Seoul Korean.

Women have been found to be more advanced in the sound change, for VOT alone (Oh, 2011), or for both VOT and  $f_0$  (Kang, 2014). This gender effect is mirrored in perception, with listeners relying on  $f_0$  more (and VOT less) when responding to female speech (Kong et al., 2011). In sum, previous work suggests a quasi-tonogenetic sound change in Seoul Korean involving VOT contrast reduction and  $f_0$  contrast enhancement gradually spreading across speakers (over time), and that this change is more advanced in female speakers.

While much is known about how the change is spreading across speakers of the language,

little is known about how the change is propagating through different phonological and lexical conditions. These aspects of the change are crucial for understanding its mechanism, as elaborated below (Sec. 2.2, 2.3). The current study uses the same corpus examined in Kang (2014), but a much larger subset of data is studied to explore the questions of how the change is initiated, how it propagates through the language (as well as the speech community), and how it impacts other aspects of the linguistic system. Specifically, we investigate how word frequency and vowel height condition this quasi-tonogenetic change in progress. The next section lays out our research questions in detail and proposes specific hypotheses and predictions.

## 2 Background

### 2.1 Gradual Sound Change

Sociolinguistic studies have documented that variation and change are associated with social factors (Labov, 1990, 2001) Language-internal change or *change from below* (Labov, 1966) is consistently characterized by two factors: younger speakers are more advanced than older speakers, and female speakers typically lead change (Labov, 1990, 1994, 2001, but see Eckert, 1989). Based on the assumption that pronunciation is more or less stable in adulthood (Sankoff, 2004), many ‘apparent-time’ studies have mapped out sound changes in progress by comparing the speech of speakers of different ages in a synchronic sample.

In contrast to the general consensus on the role of social factors in sound change, the role of properties of words has been more controversial. Since the Neogrammarians (late 19th century), phonetically conditioned sound changes have been taken to be phonetically gradual in terms of how a sound’s pronunciation changes over time, but lexically abrupt in that change affects all the relevant words simultaneously where the conditioning environment is met (e.g. Hockett, 1958). Under this view (the *Neogrammarian hypothesis*), exceptional lexical items only occur when analogy or dialect borrowing interferes with the change. The Neogrammarian hypothesis is broadly accepted to hold at the endpoints of change, but it is unclear to what extent it holds—or is expected to—in the intermediate stages of a change. The default assumption would be that there should be little variation in ‘how far along’ different words are which are undergoing a sound change.

In contrast, theorists of *lexical diffusion* (Chen, 1972; Wang, 1969) argue that different groups of words can be affected at different rates until the change gradually spreads to all the lexical items in the conditioning environment of the change. Thus, at a given time while a sound change is taking place, pronunciation variation should exist among words undergoing the change. This viewpoint is supported by studies showing differences among words in ongoing sound changes which cannot be linked to phonetic context or structural factors (which uncontroversially condition regular sound change). Most such studies adopt a usage-based viewpoint, and focus on effects of word frequency—whether words with higher frequency lead or lag in a change, compared to low-frequency words (e.g. Bybee, 2000, 2012; Bybee & Hopper, 2001; Hooper, 1976; Ogura, 2012; Phillips, 1984)—to which we now turn.

### 2.2 Origin of transphonologization: word frequency

The correlation between the direction of frequency effects and the *type* of sound change has proved robust enough that frequency effects have been argued to be diagnostic of the

mechanism of a given sound change. Low-frequency words are thought to lead in analogical changes (e.g. Bybee, 1985; Lieberman, Michel, Jackson, Tang & Nowak, 2007; Phillips, 1984); changes that involve structural generalizations in the phonology of certain word types in the lexicon (e.g. Phillips, 2006); or ambiguity or misperception-driven changes (e.g. Bybee, 2002, 2012; Hay, Pierrehumbert, Walker & LaShell, 2015; Ogura, 2012); due to their weaker availability in memory (Bybee, 2002). In contrast, high-frequency words are thought to lead sound changes driven by a leniting bias or a reduced contrast (e.g. Bybee, 2002; Bybee & Hopper, 2001; Phillips, 2006), because they have a higher probability of occurrence and higher predictability than infrequently used words (Lindblom, Guion, Hura, Moon & Willerman, 1995; Pierrehumbert, 2001), and high predictability is in turn associated with reduction (Aylett & Turk, 2004; Baker & Bradlow, 2009; Bell, Jurafsky, Fosler-Lussier, Girand, Gregory & Gildea, 2003; Bybee, 2000).

These two frequency effects are in line with two known mechanisms by which phonetically motivated sound changes, such as tonogenesis, can be triggered. First, the change can originate in misparsing of the speech signal (Ogura, 2012; Ohala, 1993a; Ohala & Sprouse, 2003), which should impact low-frequency words first, because language users have relatively less experience with these words, which will add more ambiguity in perceptual parsing than for high frequency words (Bybee, 2012; Hay et al., 2015). Ohala (1981) suggests that misperceptions occur, although rarely, when listeners fail to compensate for coarticulatory effects on segments. For the case of tonogenesis, if listeners sufficiently often misattribute the  $f_0$  difference to the vowel itself rather than to the preceding consonant (what speakers intended) (Beddor, 2009; Beddor, McGowan, Boland, Coetzee & Brasher, 2013), the speaker’s production target could shift (a ‘mini sound change’: Ohala, 1993b), which could then spread to other individuals with whom they interact via imitation (Baker, Archangeli & Mielke, 2011; Harrington, 2012; Stevens & Harrington, 2014), eventually leading to the emergence of a tonal system in the language. This is consistent with the view in classic papers on tonogenesis (Hombert, 1974; Hombert et al., 1979; Hyman, 1976; Ohala, 1978) that “phonological change is perception-oriented” (Hyman, 1976, p. 40), and listeners’ eventual selection of novel variants is not necessarily linked to the magnitude of coarticulation. We use the term *misparsing* to refer to the driving factors behind this type of change.

Second, change may originate from production variation, specifically a lenition bias targeting high-frequency words. The general lenition bias in high-frequency words will cause overall shortening of VOT in stops, and is expected to affect long-lag aspirated stops disproportionately more than other stops, based on cross-linguistic work on how VOT is affected in hypospeech (Kessinger & Blumstein, 1997; Miller, Green & Reeves, 1986; Pind, 1995, for English, Icelandic, Thai). In the Korean case, this would lead to reduction of the VOT contrast between lax and aspirated stops. Subsequently, a perceptual reinterpretation of the speech signal by the listener may follow (Beddor, 2009; Bybee, 2012; Harrington, Kleber, Reubold & Siddins, 2015; Lindblom et al., 1995). This account is consistent with the view that “significant change in the phonetic pattern” (Lindblom et al., 1995, p. 16) must be present to trigger reanalysis by listeners. We use the term *production bias* to refer to the driving forces (gestural undershoot, reduction) behind this type of change.

There has been little investigation of the role of word frequency in tonogenetic sound changes. We are aware of one experimental study which examines the degree of coda reduction in laryngealization in Vietnamese as a function of frequency and speech style (Stebbins, 2010), and argues for a relationship with an ongoing sound change. However, because different speaker ages or recording years are not considered, the findings cannot be unambiguously linked to the change.

Our first research question addresses how the change is spreading across words: *are there word frequency effects in how the quasi-tonogenetic sound change in Seoul Korean spreads through the lexicon, and if so, do high or low-frequency words lead the change?* Any word frequency effects found in our apparent-time data would give evidence for the origin of this change in production bias or misparsing. The patterns expected under the production bias and misparsing scenarios are schematized in panels A and D of Figure 1.

## 2.3 Spread of transphonologization: words and vowel contexts

Once transphonologization is triggered, how does the change spread from word to word and from context to context? During intermediate periods of a tonogenetic change, it is unlikely that speakers will use either the ‘traditional’ (maximal VOT contrast) or ‘innovative’ (maximal f0 contrast) system in production. Rather, as sound change is generally phonetically gradual, it is likely that speakers use a mixture of intermediate values of the two cues, and that the consonantal cue is used progressively less and the vocalic cue progressively more over time. Indeed, for Seoul Korean, Kang (2014) found continuous and parallel change in VOT contrast loss and f0 contrast enhancement across speakers of different ages and genders. These findings suggest there is a close, inverse relationship between the role of VOT and f0 in signaling the contrast, and that this relationship shifts over time such that f0 becomes the dominant cue. A similar relationship between two cues was observed in pre-nasal vowels in English by Beddor (2009), who found an inverse relationship between nasality in the vowel and duration of the nasal consonant across contexts. Although this data is from speakers of similar ages, the observed relationship is argued to be the precursor to the diachronic development of nasal vowels.

There are several possible mechanisms for such inverse relationships between cues. Listeners may adjust the roles of different cues to balance the total signaling requirements of the contrast (‘cue enhancement’: Kirby, 2013), or because they perceive both cues as arising from a single articulatory source (Beddor, 2009). Whatever the mechanism, in this study we use the term *adaptivity* to refer to continuous and inverse shift in the role of VOT and f0.

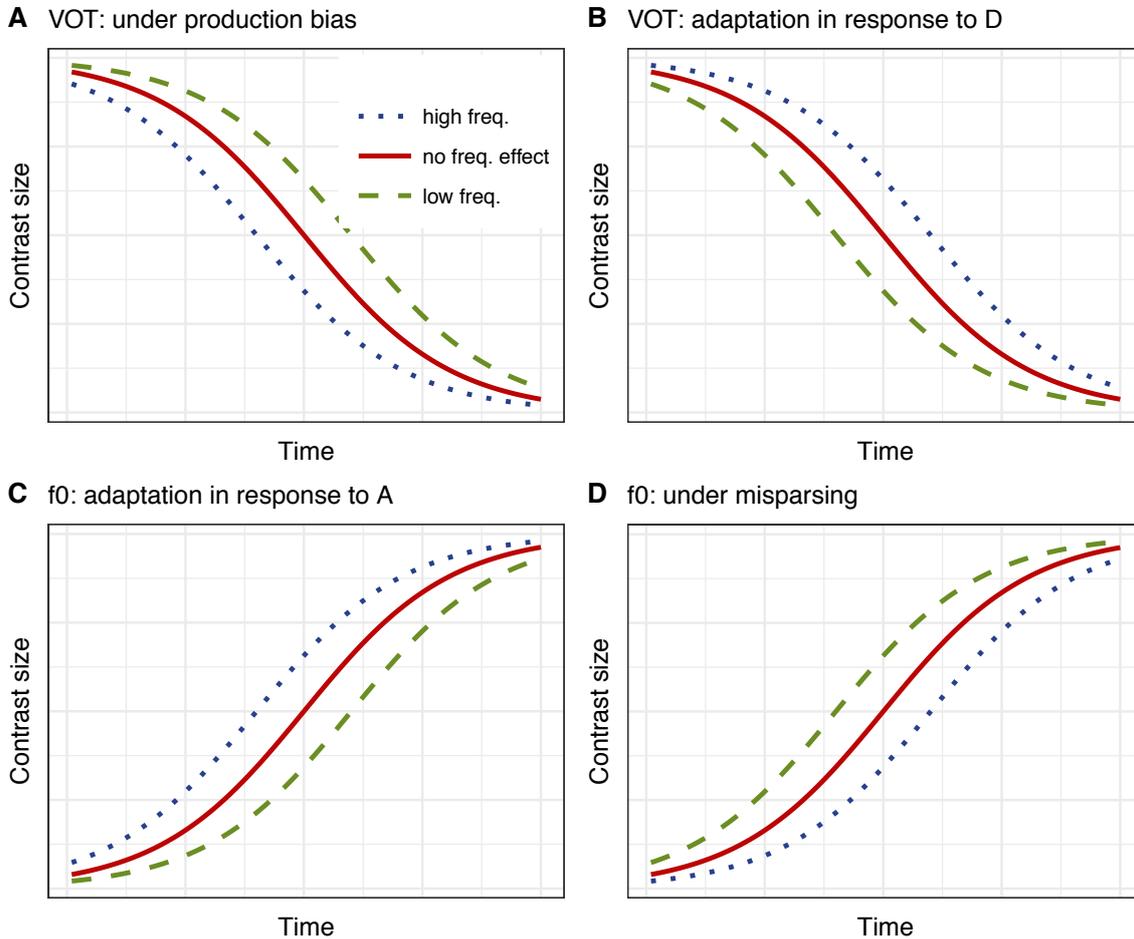
What is not known is at what level the adaptivity operates—whether adaptivity would manifest not just across speakers, but *across different linguistic contexts* as well. Put otherwise, in words and phonetic contexts where the VOT cue is used less, is the f0 cue used more? (One could imagine, alternatively, that the sound change is adaptive for any given speaker, but f0 contrast enhancement is ahead in some words and VOT contrast loss is ahead in others). We predict that if VOT contrast loss and f0 contrast enhancement are linked by adaptivity, they should proceed in tandem, both affecting the same words and phonetic contexts.

Here we discuss possible patterns that could occur during the change and how each pattern is diagnostic of a different underlying mechanism.

### 2.3.1 Predictions: Word frequency

If the sound change originates in production bias we expect to observe the pattern in Figure 1 A, where VOT contrast reduction is more advanced in high-frequency words, while if the sound change originates in misparsing, we expect to observe the pattern in Figure 1 D, where f0 contrast enhancement is more advanced in low-frequency words. Either pattern would be expected if the observed differences in the timecourse of change for words with different frequencies are due to synchronically-motivated word frequency effects: there would be more

reduction in the size of VOT contrast (caused by production bias) for higher-frequency words, and more expansion in the size of f0 contrast (caused by misparsing) for lower-frequency words. Either pattern (A) or (D) occurring independently or both occurring together would be consistent with there being an adaptive link between VOT and f0 across speakers, but *not* across words. This is the first of three possible scenarios:



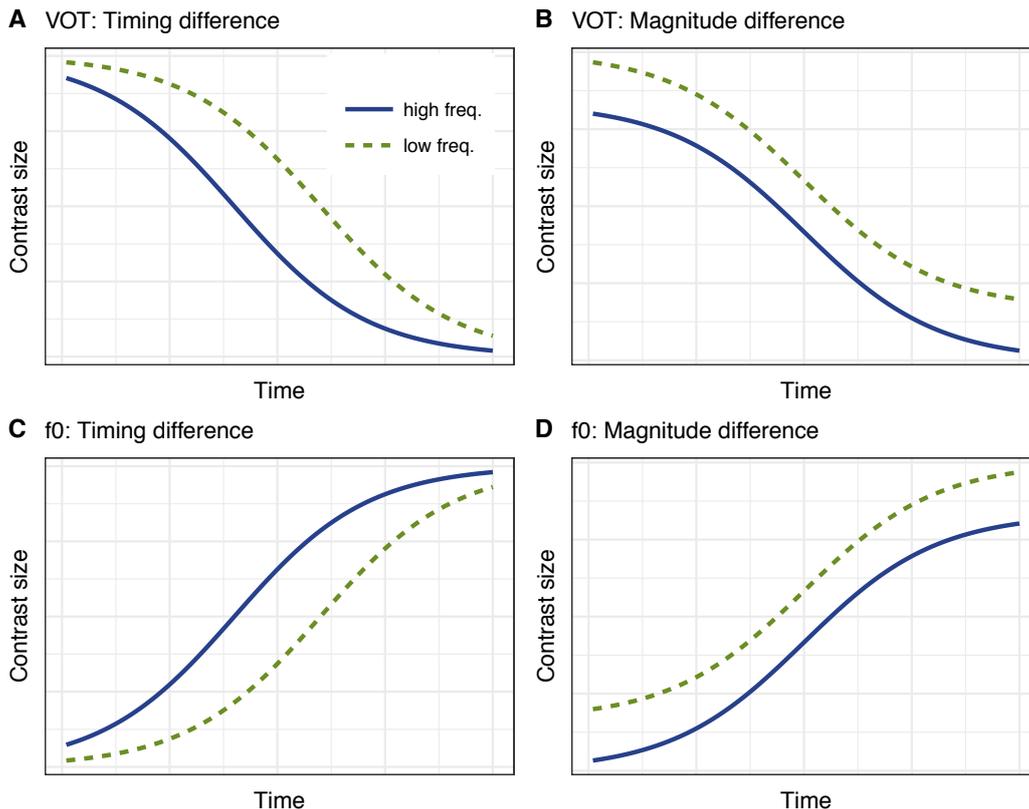
**Figure 1:** Hypothesized effects of word frequency on sound change in Seoul Korean: The S-curves illustrate change over time in the importance of VOT (A, B) and f0 (C, D) in contrasting aspirated and lax stop series. The solid lines represent the expected pattern if there were no frequency effect. The dotted and dashed lines represent the expected trajectories for words with high and low frequency respectively, under different assumptions about the source of the change: production bias (A, C) or misparsing (B, D).

1. (A), (D) or (A) + (D): production bias and/or misparsing & *no* adaptivity
2. (A) + (C): production bias & adaptivity
3. (B) + (D): misparsing & adaptivity

In scenario 2, VOT contrast reduction in high-frequency words is a trigger of f0 contrast enhancement. This pattern would be driven by production bias affecting the VOT contrast, as in (A), and adaptivity compensating for decreased VOT informativity by the f0 contrast being enhanced, as in (C). In Scenario 3, it is the low-frequency words that lead both changes (B + D), as would be expected if the change is driven by misparsing and adaptivity.

The three scenarios just outlined describe *diachronic* change. That is, they assume that any observed difference in the size of the VOT or f0 contrast between high- and low-frequency words at any time point is due to one set of words being ahead of the other.

However, for any given time point, a *synchronic* source is possible. For example, decreased VOT contrast size between high-frequency words relative to low-frequency words could be due to known reduction effects, operating on high-frequency words in a similar way across time points. We call these two possibilities *time-of-inception* (i.e. diachronic) and *magnitude* (i.e. synchronic) effects. Across the full time-course of sound change these two possibilities should have different trajectories, schematized in Figure 2. Panels A and C illustrate a time-of-inception effect where one of the curves is shifted forward in time, while Panels B and D illustrate a magnitude effect where one of the curves is shifted up across time points. Crucially, for a time-of-inception effect, the difference in contrast size across words would change over time.



**Figure 2:** Schematic of effects of word frequency on sound change that would result from timing effects (A, C) versus magnitude effects (B, D). The solid (high frequency) and dotted (low frequency) lines represent the expected trajectories for words with high and low frequency. (A) and (C) are expected if the change is caused by production bias in VOT and an adaptive link to  $f_0$ , as predicted in Scenario 2 (see text).

The patterns in (A) and (C) of Figure 2, where high-frequency words change sooner, could be also explained by a ‘rate effect’ predicted by usage-based accounts of sound change (Bell, Brenier, Gregory, Girand & Jurafsky, 2009; Bybee, 2000; Hay et al., 2015; Pierrehumbert, 2002, 2001): high-frequency words would change at a faster rate than low-frequency words in reduction-driven changes, and vice versa for ambiguity or analogy-driven changes (see Sec. 2.2) In the presence of a rate effect, the difference in contrast size across words again would change over time.

Either a time-of-inception effect or a rate effect would lead to some words being ahead of others in the middle of the sound change. Distinguishing between these two types of effects requires a broader time range than is available in our data, containing a stable time period before the change begins. We use the term *timing effects* to encompass time-of-inception

and rate effects, because what is crucial for our research questions is not to differentiate between these two types of effects, but to distinguish them from (synchronic) magnitude effects. Either type of timing effect would indicate different progression of the change across words, while a magnitude effect would not. Any timing effect is most likely to be detected during a portion of the S-shaped curve of the change when there is large variation across words.

### 2.3.2 Predictions: Vowel height

Word frequency is one way to examine propagation of a tonogenetic change through a language’s lexicon. Another way is to examine change across phonetic contexts. We focus on vowel height in particular because it affects both VOT of the preceding stop and f0 of the vowel (both are increased in high vowel contexts compared to non-high contexts, cross-linguistically: Esposito, 2002; Higgins, Netsell & Schulte, 1998; Honda, 1983; Hoole & Honda, 2011; Klatt, 1975; Whalen & Levitt, 1995; Whalen, Levitt, Hsiao & Smorodinsky, 1995), and because of our interest in intrinsic f0 effects (see Sec. 2.4). Otherwise, the choice of vowel height as a phonetic context (as opposed to e.g. stop place of articulation) is somewhat arbitrary—unlike word frequency, which previous work suggests could play a role in triggering the sound change.

Unlike for word frequency, we do not have a clear prediction for how vowel height affects contrast size: whether high- or non-high-vowel context enhances the f0 distinction or reduces the VOT distinction between stop categories. Therefore, any observed vowel height effect cannot distinguish between production bias and misparsing as the origin of the sound change (as discussed in Sec. 5.2.) Instead, we can only assess the presence or absence of adaptivity across vowel contexts.

There are two possible scenarios with respect to vowel height (where A–D refer to Figure 1, replacing dotted/dashed lines with non-high/high vowels). If an adaptive mechanism does not function across contexts, we would observe that the change in VOT is led by one context (e.g. high vowels) while the change in f0 is led by a different context (e.g. non-high vowels), or that the change in one cue is not modulated by vowel height at all. Alternatively, if the change spreads across contexts in an adaptive way, we would observe a continuous and gradual shift from the VOT dominant pattern to the f0 dominant pattern. In this case, both VOT contrast reduction and f0 contrast enhancement would be more advanced in the same vowel context midway through the change.

The discussion above leads to our second research question, which addresses the goal of a better understanding of the intermediate stages of tonogenetic sound change: *how is the emergence of contrastive f0 in Korean propagating across words with different frequencies and vowel contexts, and does it do so in an adaptive way (Lindblom et al., 1995)?*

## 2.4 Impact of transphonologization: vowel intrinsic f0

As f0 gradually becomes the primary cue, another relevant question is whether and how the innovative f0 contrast affects other aspects of the linguistic system. Languages which use f0 contrastively (for tonal or pitch-accent systems) may be constrained in the functional use of f0 (i.e. intonation, Yip, 2002; Beckman & Pierrehumbert, 1986; c.f. Torreira, Bögels & Levinson, 2015) or phonetic effects on f0 (Connell, 2002), compared to other languages. In the current study, we ask whether the increasing importance of f0 in the stop contrast affects the relationship between f0 and vowel height. To understand this, we must consider the mechanisms underlying f0 realization.

First,  $f_0$  can be deliberately controlled by muscular maneuvers—generally using the cricothyroid (CT) muscle (Atkinson, 1972; Hirose & Gay, 1972; Honda, Hirai & Dang, 1994; Roubeau, Chevrie-Muller & Saint Guily, 1972). Second,  $f_0$  perturbations associated with consonantal laryngeal class (e.g. voiced/voiceless) are generally thought to be due to physiological and/or aerodynamic constraints inherent to consonant voicing production (Bell-Berti, 1975; Hyman, 1976; Löfqvist et al., 1989; Ohala, 1993b, 2000). Third, anatomical links between the tongue and the larynx can affect  $f_0$  (Honda, 1983), which is thought to be responsible for the cross-linguistic tendency of high vowels to have higher  $f_0$  than non-high vowels (i.e. *intrinsic  $f_0$*  effects: IF0 effects; Lehiste, 1976; Whalen & Levitt, 1995; Whalen et al., 1995). Thus, variation in  $f_0$  can be due to physiological factors as well as muscular control and these components can in principle work together to enhance vowel height contrasts or consonant voicing (Hoole, Honda, Murano, Fuchs & Pape, 2006; Kingston, 1992), or against each other to preserve or enhance tonal contrasts (Connell, 2002).

IF0 effects appear to be near-universal: Whalen & Levitt (1995) found an IF0 effect in all 31 languages studied in a meta-analysis, and argue that IF0 is an automatic physiological process. However, the size of IF0 effects differ substantially across speakers and languages (e.g. Van Hoof & Verhoeven, 2011). In particular, based on data from four African tone languages and Whalen & Levitt (1995)’s survey, Connell (2002) argues that IF0 effects in tonal languages are generally smaller than in intonational languages, and concludes that IF0 effects may be smaller in a language where they would obscure tonal contrasts.

These studies lead to the question of whether the emergence of contrastive  $f_0$  in tonogenetic sound change could affect non-contrastive variation in  $f_0$ . While previous work has compared across different languages, the change in progress in Korean affords an interesting opportunity to observe the relationship between the size of the IF0 effect and the role of  $f_0$  within a single language, where other variables are held constant. Because  $f_0$  variation arises from both mechanical factors and active control (Solé, 2007), one possibility is that speakers actively attenuate the mechanical factors in order to enhance the contrastive use of  $f_0$  as transphonologization occurs. In this case, *the size of the IF0 effect would differ before and after the tonogenetic sound change*. IF0 effects could be also affected by the fact that the direction and magnitude of the  $f_0$  change differs by stop in Seoul Korean— $f_0$  decreases for lax stops and increases for both aspirated and tense stops, but less so for tense stops (Kang, 2014). It has been argued that IF0 attenuation is primarily constrained by the mechanical status of the larynx in low tone production (Ladd & Silverman, 1984; Whalen & Levitt, 1995). If this is correct, IF0 effects may be attenuated to a greater degree for lax stops, which have the lowest  $f_0$ , than other categories. Alternatively, if IF0 effects are largely constrained by pressure to maintain tonal contrast (Hoole et al., 2006), the degree of change in the IF0 effect over time may depend on the degree of the importance of  $f_0$  for signaling phonological contrasts of a particular stop category.

Our third research question is: *does the IF0 effect in Seoul Korean change as contrastive  $f_0$  emerges, and does the magnitude of change in the IF0 effect differ by stop?*

### 3 Data and Methods

We address our research questions on the origin, progression, and impact of tonogenetic sound change, using apparent-time corpus data from Seoul Korean.

### 3.1 Corpus data

The data come from The Speech Corpus of Reading-Style Standard Korean (The National Institute of the Korean Language, 2005), henceforth the *NIKL Corpus*. The corpus consists of recordings of 120 Seoul dialect speakers, aged 19 to 71 years old, reading essays and children’s stories. The recordings were made in sound attenuated booths in the Seoul metropolitan area in 2003, and each sentence was stored as an individual audio file. We used a version of the corpus which is force-aligned at the word and segment level using the Korean Phonetic Aligner (Yoon, 2014; Yoon & Kang, 2014). This corpus was also used by Kang (2014), who examined a subset of 1250 tokens from 11 words, across 118 speakers, in utterance-initial position. (Following Kang (2014), we excluded two speakers for whom all sound files contained recording errors.) Given our focus on the spread of the sound change across words and lexical contexts, we expanded the dataset as much as possible to include many more words. We also considered positions besides utterance-initial, in order to increase the amount of data per speaker and word, to maximize our statistical power for detecting word-level effects. To examine the pronunciation of different words over time, it was important to use words pronounced by speakers from all age groups. We therefore limited ourselves to the 11 stories (out of 19) read by speakers from all age groups.

Using the data from these 11 stories for the 118 speakers, we first extracted all words beginning with any of the nine stops ( $\{\text{alveolar, bilabial, velar}\} \times \{\text{tense, lax, aspirated}\}$ ). The dataset was then constructed by restricting it by prosodic context and other factors, as follows.

The nature of the sound change affecting lax and aspirated stops crucially depends on prosodic structure. Korean is often analyzed as having three prosodic units larger than a Prosodic Word (PW): the Accentual Phrase (AP), Intermediate Phrase (ip), and Intonation Phrase (IP) (Jun, 2005). Each higher-order prosodic unit consists of one or more lower units. For example, an AP consists of one or more PW’s. The sound change in progress in Seoul Korean is thought to affect only sounds at the left edge of the AP (and thus higher prosodic domains). Because of the difficulty of annotating AP boundaries, we limited our investigation to IP-initial stops (Jun, 1993, 1996): all tokens in sentence-initial position, as well as a subset of tokens in sentence-medial position, were selected as follows:

- Only stops preceded by a force-aligned pause longer than 30 ms (to lessen the possibility of including stop closures mislabeled as pauses)—since IP’s are almost always preceded by some pause.
- Among these stops, tokens were selected if there was a syntactic clause boundary (e.g. after a conjunctive morpheme or a topic marker).
- In other cases where there was a force-aligned pause, the first author manually identified IP boundaries which were cued by pitch resetting (a secondary cue for IP’s).

This subset of the data, consisting only of IP-initial stops, was then further restricted to a subset of *items*, defined as a particular occurrence of a word in a sentence. Each item was present for a different number of speakers (since speakers differ in whether utterance-medial items were produced with a preceding pause). In order to address our research questions about how the change is impacted by properties of words and phonetic contexts (i.e., items), we selected items to give a roughly equal distribution among different values of item-level variables (laryngeal category, place of articulation, and vowel height), and we prioritized items which occurred for a larger number of speakers. The final dataset consisted of 6916 tokens from 81 items.

## 3.2 Dataset construction

For each token in this dataset, we measured VOT, f0, and other variables. We measured VOT using a semi-automatic method (similar to Stuart-Smith, Sonderegger & Rathcke, 2015): automatic measurement, followed by manual correction. Automatic measurements were obtained using the software package ‘AutoVOT’ (Keshet, Sonderegger & Knowles, 2014), which uses an algorithm trained on a small set of hand-annotated tokens to measure VOT. For the training dataset, VOT onset was determined at the time of the burst and VOT offset at the time of the first visible indication of voicing, based on the initiation of periodicity in the waveform. The algorithm was separately trained for each of the three laryngeal categories based on 100 manually-coded VOTs, then used to assign automatic measurements to each stop in the full dataset. All automatic measurements were manually checked (by the first author), and hand-adjusted if necessary based on the same criteria applied to the training dataset.

For each token, f0 was extracted at the vowel midpoint using a Praat script (25 ms analysis window; f0 range of 80–450 Hz; time step = 5 ms). To detect pitch tracking errors, we examined histograms of the resulting f0 values by gender, decade of birth, and stop category (lax, aspirated, tense); values at histogram edges were manually checked and remeasured if necessary. Errors due to devoiced high vowels were removed ( $n = 67$ ), due to undefined f0, leaving a total of 6849 tokens in the final dataset. Summary statistics for f0 and VOT by stop category and speaker decade of birth are shown in Table 1.

The measurement of f0 varies across speakers as a function of age and gender (Titze, 1989; Torre & Barlow, 2009): in addition to higher overall f0 for female speakers, there is a general lowering of f0 for women and raising of f0 for men in older age (Soltani, Ashayeri, Modarresi, Salavati & Ghomashchi, 2014; Torre & Barlow, 2009); pitch range varies as well as a function of age and gender, as a higher mean f0 is associated with a larger pitch range. Such age and gender-related variation must be controlled for when examining a diachronic change in an f0 contrast (Reubold & Harrington, 2015). We do so by converting f0 to semitones, which represent equal perceptual intervals relative to each speaker’s mean f0 (Nolan, 2003). Each speaker’s mean f0 was estimated by averaging f0 over all vowels ( $n=504$ ) in one story (*Sungnyungyi Jihye*), and used to convert raw f0 values into semitones. On this logarithmic scale, positive and negative values indicate f0 values higher and lower than a speaker’s mean.

We also used two measures of speech rate. Raw speech rate was defined as syllables per second in a sentence. We then calculated each speaker’s *mean speech rate* (mean of raw speech rate across all sentences), and the difference between each token’s raw speech rate and the speaker’s mean rate (*speech rate deviation*). These two measures account for two ways speech rate might affect VOT (following Stuart-Smith et al., 2015): within speakers, VOT may be shorter for faster speech; across speakers, VOT may be shorter for faster speakers.

Finally, wordform frequency information was taken from the KAIST Concordance program (KAIST, 1999) based on the 70 million-word KAIST Corpus (Yoon & Choi, 1999) and log-transformed.

## 3.3 Statistical models

### 3.3.1 Variables

We model VOT and f0 as a function of a number of variables that are properties of speakers, items, and utterances (termed *speaker-level* variables, etc.), indicated in SMALL CAPS.

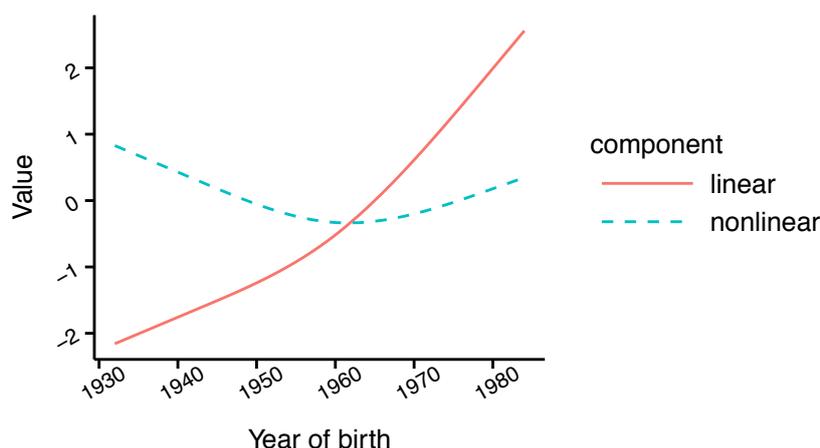
The speaker-level variables year of birth (YOB) and GENDER are included in the models

**Table 1:** Summary statistics for VOT (ms) and f0 (Hz, before normalization) by stop category and speaker decade of birth: mean, standard deviation, and number of tokens ( $n$ ). Number of speakers per decade is shown in parentheses.

Decade of birth	Laryngeal class	Stop	VOT (msec)		f0 (Hz)		$n$
			mean	SD	mean	SD	
1930s (6)	tense	p*	8.78	5.17	146.53	38.16	20
		t*	9.72	5.28	138.59	27.74	27
		k*	23.28	8.73	150.23	32.87	25
	lax	p	39.66	15.78	135.44	29.51	58
		t	37.26	17.28	131.00	27.54	61
		k	51.69	15.53	139.73	30.38	66
	aspirated	p <sup>h</sup>	84.31	18.87	161.78	34.05	37
		t <sup>h</sup>	53.06	20.26	148.28	37.46	71
		k <sup>h</sup>	106.77	24.48	164.41	35.13	22
1940s (21)	tense	p*	10.19	4.94	190.81	56.31	78
		t*	11.11	5.08	187.90	50.26	82
		k*	24.73	10.42	192.07	61.09	91
	lax	p	44.34	16.11	165.28	44.21	197
		t	41.34	16.60	160.24	44.72	204
		k	50.41	17.80	165.33	45.03	228
	aspirated	p <sup>h</sup>	72.77	22.65	201.92	58.49	126
		t <sup>h</sup>	52.60	16.46	199.12	62.25	251
		k <sup>h</sup>	89.26	16.74	200.65	63.00	65
1950s (29)	tense	p*	10.40	5.46	247.49	51.34	97
		t*	10.47	5.22	241.33	49.27	107
		k*	18.51	8.38	252.68	52.32	124
	lax	p	42.10	14.74	203.52	39.98	252
		t	42.51	17.93	201.07	39.85	250
		k	49.94	16.47	209.25	38.63	307
	aspirated	p <sup>h</sup>	57.05	17.30	260.19	54.70	160
		t <sup>h</sup>	40.81	13.67	265.01	56.43	335
		k <sup>h</sup>	79.25	17.43	254.86	56.40	78
1960s (11)	tense	p*	14.28	10.56	192.08	82.00	38
		t*	11.26	5.07	193.57	80.78	45
		k*	23.54	8.41	196.12	81.92	47
	lax	p	45.12	16.39	153.36	54.08	102
		t	45.23	15.20	156.38	56.97	90
		k	54.08	18.96	158.84	57.90	127
	aspirated	p <sup>h</sup>	61.94	18.82	213.94	97.70	62
		t <sup>h</sup>	43.66	11.68	214.14	104.03	126
		k <sup>h</sup>	79.20	12.72	206.78	96.91	29
1970s (37)	tense	p*	12.54	6.69	180.71	58.28	123
		t*	12.76	6.50	180.89	56.27	120
		k*	22.32	8.16	186.57	53.43	137
	lax	p	41.38	14.09	151.56	44.29	306
		t	42.77	15.33	153.10	44.96	307
		k	49.27	15.53	154.15	44.80	400
	aspirated	p <sup>h</sup>	51.17	15.69	200.26	59.14	176
		t <sup>h</sup>	34.45	11.76	197.83	63.73	415
		k <sup>h</sup>	69.39	17.34	188.54	56.15	77
1980s (14)	tense	p*	12.96	7.20	214.88	76.40	34
		t*	11.55	5.70	217.10	72.72	39
		k*	18.79	9.82	225.46	69.63	52
	lax	p	39.54	16.54	180.63	52.18	101
		t	43.23	17.14	179.42	51.29	106
		k	47.39	16.33	183.35	54.17	136
	aspirated	p <sup>h</sup>	48.72	20.98	231.48	72.27	66
		t <sup>h</sup>	32.20	10.14	239.02	74.41	143
		k <sup>h</sup>	62.95	14.76	217.37	74.06	26

to account for the diachronic change and the expectation that it is led by female speakers (Kang, 2014; Kong et al., 2011; Oh, 2011). Based on exploratory plots, as well as the nonlinear relationship between year of birth and VOT/f0 evident in previous work (Kang, 2014), YOB was coded as linear and nonlinear effects. Specifically, we coded YOB using a restricted cubic spline with three knots, using `RCS()` in the `rms` package (Harrell & Frank, 2015) in R, with degrees of freedom chosen based on exploratory plots. This corresponds to two variables for YOB, called *components*, which are shown in Figure 3 to aid in interpreting model results involving YOB. The first component, which looks roughly like a line, we call the ‘linear’ component. The second component, which looks roughly like a quadratic function, we call the ‘nonlinear’ component. Thus, the two components can be interpreted roughly as the linear and quadratic terms of a polynomial, which are a common way to model nonlinear effects that “look quadratic” (e.g. as used in Zellou & Tamminga, 2014), but with the crucial property that they grow linearly rather than quadratically at the minimum and maximum of the range of YOB, which is preferable for accurately predicting near these endpoints (see Baayen, 2008; Harrell, 2001). Both components are included in each model below, to jointly represent the effect of YOB.<sup>2</sup>

YOB was first centered and divided by two standard deviations (*standardized*; see Gelman & Hill, 2007), and GENDER was coded using sum contrasts (female < male).



**Figure 3:** Values of the first (‘linear’) and second (‘nonlinear’) components of the restricted cubic spline coding of YOB, for the range of years of birth represented in the dataset.

Four item-level variables were included in the model. Of primary interest is how the contrast between lax and aspirated stops changes over time and depends on other variables; thus, laryngeal class (LARYNGEAL) was coded using Helmert contrasts, corresponding to tense vs. non-tense stops (LARYNGEAL1) and lax vs. aspirated stops (LARYNGEAL2). Each item’s word FREQUENCY and vowel HEIGHT (of the vowel following the stop) are included; the effects of these variables are critical for our research questions. Log-transformed frequency was standardized, and HEIGHT was coded using sum contrasts (non-high < high). PLACE of

<sup>2</sup>A reviewer suggests instead using a logistic function of time, reflecting the ‘S-shaped curve’ characteristic of linguistic change. We experimented with doing so, but found that it was not possible to fit logistic functions because the data is not from a large enough time range to infer the full S-shape, and is thus ambiguous between different possible diachronic trajectories (e.g. magnitude versus timing effects). We believe this situation in fact obtains for most cases of phonetic change in progress, and we follow other recent work on such cases by coding time using a linear or non-logistic nonlinear function (Fruehwald, 2016; Hay & Foulkes, 2016; Hay et al., 2015; Kang, 2014; Zellou & Tamminga, 2014). The broader issue of what can be inferred about the overall trajectory of change from data from only part of the change is an interesting one for future work.

articulation of the stop was included as a control variable (coded using Helmert contrasts: labial vs. nonlabial; alveolar vs. velar), due to its strong effect on VOT cross-linguistically and in Seoul Korean (expected: labial < alveolar < velar; Cho et al., 2002; Cho & Ladefoged, 1999; Lisker & Abramson, 1964).

Several utterance-level variables are also included in the model. Recall that the data comes from IP-initial words, which may be sentence-initial or follow a pause. Both utterance position and the quantitative strength of a prosodic boundary (using the proxy of pause duration) are expected to affect both VOT and f0 in Seoul Korean (Cho & Keating, 2001; Jun, 1996, 1998; Kang & Guion, 2008; Keating, Cho, Fougeron & Hsu, 2003). We coded both sentence position and pause duration as a single POSITION factor with four levels, with pause duration cutoffs chosen using *cut2* in the `Hmisc` package (Harrell, 2015) in R: (1) utterance-initial stops; utterance-medial stops preceded by (2) a short pause (< 280 ms); (3) a medium pause (280–430 ms); (4) a long pause ( $\geq$  430 ms). POSITION was coded using Helmert contrasts: utterance-initial stops vs. utterance-medial stops (POSITION1); stops after a short pause vs. after medium–long pauses (POSITION2); stops after a medium pause vs. after a long pause (POSITION3). Thus, POSITION1 encodes utterance position, while POSITION2 and POSITION3 encode pause length for sentence-medial stops.

Each speaker’s mean speech rate (SPEAKER MEAN RATE; a speaker-level variable) and deviation from the mean for each token (RATE DEVIATION; an utterance-level variable) were included in the models. Cross-linguistically, faster speech is strongly negatively correlated with VOT for stops signaled with long-lag VOT, while short-lag categories show small or null effects (Kessinger & Blumstein, 1997; Miller et al., 1986; Pind, 1995). Because all three stop categories are signaled with positive VOT in Seoul Korean, we expect that speech rate will negatively affect VOT, but possibly only for long-lag stops (i.e., especially for aspirated stops in the case of Korean). In particular, we expect these effects for RATE DEVIATION, which corresponds to slower/faster speech by a given speaker relative to his/her mean speaking rate. In addition to a speech rate effect on VOT, both speech rate measures may index the degree of hyperarticulation, which may play a role in this sound change (see above Sec. 2.2), thus influencing both VOT and f0. Including SPEAKER MEAN RATE also controls for an important confound for any effect of speaker age (which is of primary interest, for inferring change over time): older speakers may speak slower than younger speakers (e.g. Jacewicz, Fox, O’Neill & Salmons, 2009), which could in turn affect VOT and f0 for the reasons just mentioned, potentially interfering with inferences about *change* in VOT and f0. Both speech rate measures were standardized.

The dependent variables VOT and f0 were transformed before inclusion in the models. The distribution of VOT, which can only be positive (for Korean stops), is heavily right-skewed; VOT was thus log-transformed, to bring its distribution closer to normality. f0 was normalized by converting to semitones, as discussed above.

### 3.3.2 Model structure

VOT and f0 were modeled as a function of the nine independent variables introduced above, using linear mixed-effects models, fitted using the *lmer* function from the `lme4` package (Douglas, Martin, Ben & Steve, 2015) in R. The models for VOT and f0 had identical structure (fixed and random effects), which allows us to assess to what extent VOT and f0 are changing in parallel across speakers, words, and phonetic contexts.

**Fixed effects:** Main effect terms were included for the nine independent variables. Interaction terms were chosen to address our research questions and control for known factors affecting VOT and f0. Two-way interactions between laryngeal category and speaker-

level variables (LARYNGEAL:YOB, LARYNGEAL:GENDER) were included to capture how both cues to the stop contrast are changing over time, across speakers. Interactions between laryngeal category and (1) frequency and (2) vowel height (LARYNGEAL:FREQUENCY, LARYNGEAL:HEIGHT) were included to examine how the change is spreading across words of different frequencies and across vowel contexts (Questions 1–2). The interaction between HEIGHT and YOB was included to examine whether and how the IF0 effect is modulated by the sound change (Question 3). The interaction between LARYNGEAL and RATE DEVIATION was included to account for expected speech rate effects on VOT, which should differ between stop categories, as well as any hyperarticulation effects on VOT and f0. The interaction between LARYNGEAL and POSITION was included to control for expected prosodic effects on both cues.

We included two types of three-way interactions to address dynamic aspects of the sound change (related to Questions 2–3). The YOB:LARYNGEAL:FREQUENCY and YOB:LARYNGEAL:HEIGHT interactions assess whether word frequency and vowel height tease apart synchronic magnitude effects and diachronic timing effects. The YOB:LARYNGEAL:HEIGHT interaction further addresses whether there is a difference in the magnitude of the IF0 change over time between laryngeal categories. Note that we do not include a YOB:LARYNGEAL:GENDER interaction—this effect has already been discussed by Kang (2014) for this dataset, and is not related to our research questions.

**Random effects:** The models included by-item and by-speaker random intercepts, to account for variability in VOT and f0 of speakers and items beyond the effects of variables included in the models. The models also included all possible by-item and by-speaker random slopes, to account for variability among speakers and items in the effects of variables on VOT and f0 (Barr, Levy, Scheepers & Tily, 2013). Correlations between random-effect terms were omitted to facilitate model convergence.

We note that our statistical methodology is highly conservative: we do not omit non-significant fixed-effect terms from models—all of which are either related to our research questions or motivated based on prior work—and include all possible random slopes. By doing so, we prioritize *accurate coefficient estimates* and minimize spurious effects (Type I errors), at the risk of lower statistical power (i.e., overly conservative significances). (For discussion of these issues, see e.g. Barr et al., 2013; Bates, Kliegl, Vasishth & Baayen, 2015; Gelman & Hill, 2007; Matuschek, Kliegl, Vasishth, Baayen & Bates, 2015.) As a result, it is crucial when discussing our results to discuss the direction and values of coefficient estimates corresponding to our research questions, regardless of whether they reach a conventional significance threshold (e.g.  $p < 0.05$ ).

## 4 Results

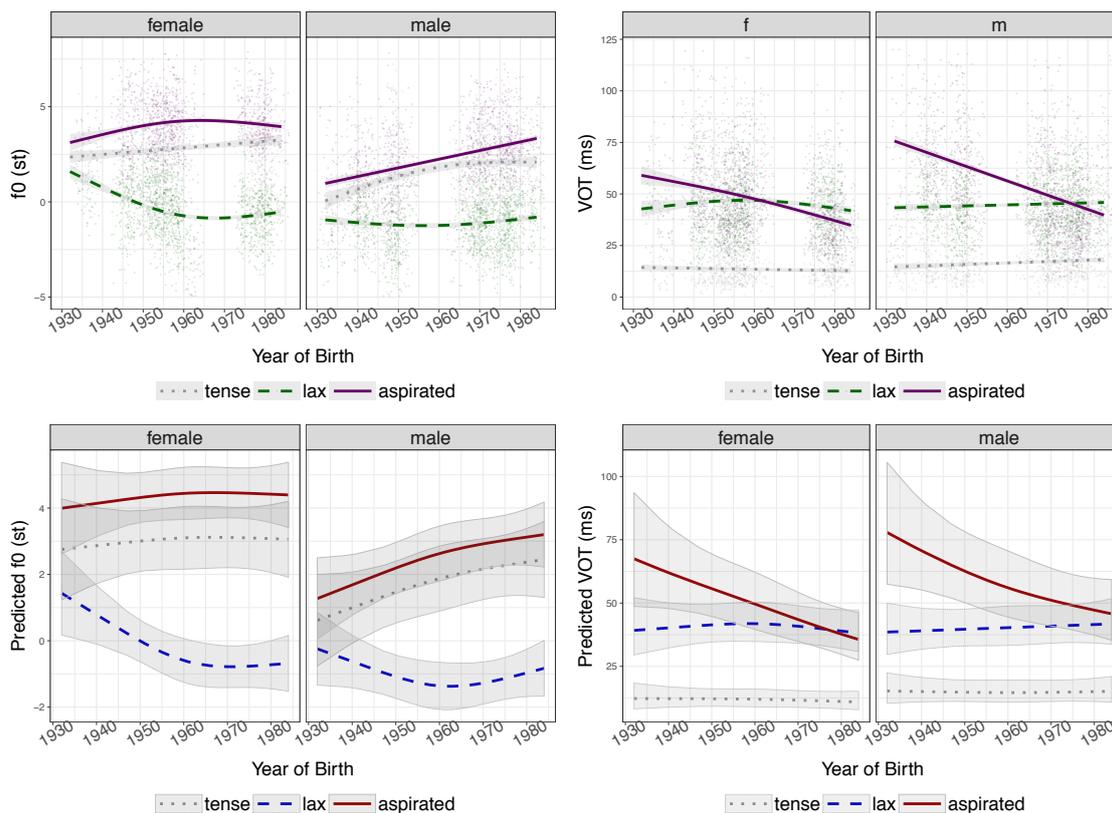
The fixed effects for the statistical models of VOT and f0 are summarized in Table 2: each fixed-effect coefficient is shown with its associated standard error, degrees of freedom, test statistic, and significance, calculated using the Satterthwaite approximation as implemented in the `lmerTest` package (Kuznetsova, Brockhoff & Christensen, 2015). We present these results in stages, showing different aspects of how the sound change progresses. (Random effects are not shown.) We first discuss how VOT and f0 for aspirated and lax stops are affected by the speaker-level variables (year of birth, gender; Sec. 4.1) addressed in previous work; we then turn to the effects of word frequency and following vowel height (word-level variables: Sec. 4.2), which are the foci of our research questions; and briefly discuss the effects of other variables included as controls (Sec. 4.3). For each subset of fixed-effect

terms, we summarize the model results quantitatively (using the regression table results) and graphically, by showing model predictions corresponding to these terms (how they are predicted to affect VOT and f0, holding other variables constant),<sup>3</sup> as well as the empirical trends corresponding to these predictions (where other variables are *not* held constant).

Our primary interest is to assess the change in the way lax and aspirated stops are contrasted (LARYNGEAL2) over time and how other variables modulate the change. Therefore, most of the main effects are discussed in terms of their interaction with LARYNGEAL2. In both models, all the categorical predictors were coded using Helmert or sum contrasts and all continuous predictors were centered. Therefore, the coefficient for a main effect term of a variable  $X$  can be interpreted as its “average” effect, marginalizing over any other variables which are part of interactions involving  $X$  (holding continuous variables at average values; averaging over categorical variables).

## 4.1 Change across speakers

We first present the model results with respect to the speaker-level variables age (YOB: linear and nonlinear components) and GENDER, and their interactions with the aspirated/lax contrast (LARYNGEAL2), which establishes the basic pattern of sound change in the aspirated/lax contrast for VOT and f0. Figure 4 shows the empirical distributions and the model predictions of f0 and VOT by stop category, speaker year of birth, and gender.



**Figure 4:** Empirical plots (top) and model prediction plots (bottom) for f0 (left) and for VOT (right) of three laryngeal categories for female and male speakers as a function of speaker year of birth: Lines show a quadratic smooth to empirical data or the model-predicted effect; shadings are 95% confidence intervals (CIs).

<sup>3</sup>95% confidence intervals for model predictions in Figures 4–8 were calculated using the variance-covariance matrix of the fixed-effect terms.

**Table 2:** Summary of all fixed-effect coefficients for the models of f0 (left) and log(VOT) (right): coefficient estimates, standard errors, degrees of freedom (df),  $t$ -values, and significances. YOB' and YOB'' refer to the linear and nonlinear components of the YEAR OF BIRTH variable. Note that LARYNGEAL2 compares lax and aspirated stops.

FULL MODELS	f0					VOT				
	Estimate	SE	df	$t$	$P(> t)$	Estimate	SE	df	$t$	$P(> t)$
Intercept	1.573	0.104	154.926	15.135	< 0.001	3.409	0.028	97.544	122.569	< 0.001
YOB'	0.095	0.065	120.354	1.451	0.149	-0.034	0.012	122.56	-2.798	0.006
YOB''	-0.089	0.31	115.042	-0.285	0.776	0.003	0.058	115.529	0.052	0.959
LARYNGEAL1(tense vs. nontense)	-1.038	0.193	76.845	-5.366	< 0.001	1.242	0.063	78.149	19.705	< 0.001
LARYNGEAL2(lax vs. aspirated)	4.149	0.196	113.849	21.118	< 0.001	0.221	0.056	75.249	3.909	< 0.001
HEIGHT(h)	0.96	0.196	72.842	4.89	< 0.001	0.134	0.062	67.434	2.139	0.036
FREQUENCY	-0.304	0.155	68.342	-1.958	0.054	-0.11	0.05	65.925	-2.2	0.031
POSITION1(initial vs. medial)	-0.215	0.115	118.191	-1.862	0.065	0.1	0.035	97.702	2.848	0.005
POSITION2(short vs. longer pause)	0.243	0.061	147.696	3.969	< 0.001	0.001	0.018	160.767	0.066	0.948
POSITION3(medial vs. long pause)	0.115	0.07	162.122	1.63	0.105	-0.025	0.02	137.155	-1.253	0.212
RATE DEVIATION	0.157	0.061	457.146	2.567	0.011	-0.007	0.015	5591.925	-0.439	0.661
GENDER(m)	-1.248	0.145	121.793	-8.63	< 0.001	0.127	0.027	124.629	4.706	< 0.001
PLACE1(labial vs. non-labial)	-0.022	0.148	69.631	-0.15	0.881	0.123	0.048	68.565	2.575	0.012
PLACE2(alveolar vs. velar)	0.225	0.192	70.632	1.175	0.244	0.314	0.062	68.326	5.1	< 0.001
SPEAKER MEAN RATE	0.218	0.185	115.914	1.178	0.241	-0.021	0.035	120.983	-0.605	0.546
YOB':LARYNGEAL1	-0.145	0.052	85.275	-2.802	0.006	-0.061	0.017	124.312	-3.561	0.001
YOB':LARYNGEAL2	0.362	0.077	138.585	4.705	< 0.001	-0.118	0.013	114.175	-8.826	< 0.001
YOB'':LARYNGEAL1	0.614	0.25	86.267	2.452	0.016	0.037	0.09	109.225	0.408	0.684
YOB'':LARYNGEAL2	-1.692	0.427	120.036	-3.96	< 0.001	0.173	0.069	106.083	2.505	0.014
YOB':HEIGHT	-0.142	0.051	81.597	-2.779	0.007	-0.011	0.013	84.856	-0.858	0.393
YOB':FREQ.	0.044	0.038	57.598	1.154	0.253	-0.015	0.01	71.298	-1.492	0.14
YOB'':HEIGHT	0.583	0.242	88.341	2.405	0.018	0.154	0.063	85.177	2.449	0.016
YOB'':FREQ.	0.024	0.17	55.637	0.144	0.886	-0.041	0.048	67.411	-0.854	0.396
LARYNGEAL1:HEIGHT	-0.539	0.474	67.977	-1.138	0.259	-0.233	0.152	65.32	-1.527	0.132
LARYNGEAL2:HEIGHT	-0.692	0.331	76.972	-2.094	0.04	0.326	0.103	66.717	3.157	0.002
LARYNGEAL1:FREQ.	0.417	0.354	67.225	1.179	0.243	0.193	0.114	66.345	1.686	0.096
LARYNGEAL2:FREQ.	0.625	0.342	68.8	1.827	0.072	-0.185	0.109	64.986	-1.695	0.095
LARYNGEAL1:POSITION1	0.283	0.219	213.937	1.294	0.197	-0.037	0.067	204.377	-0.553	0.581
LARYNGEAL2:POSITION1	0.527	0.292	72.891	1.801	0.076	0.158	0.092	65.725	1.709	0.092
LARYNGEAL1:POSITION2	0.182	0.132	156.386	1.38	0.17	-0.031	0.047	139.183	-0.657	0.512
LARYNGEAL2:POSITION2	-0.047	0.126	175.065	-0.37	0.712	0.009	0.032	5565.071	0.296	0.768
LARYNGEAL1:POSITION3	0.165	0.171	142.551	0.968	0.335	0.048	0.048	127.345	1.015	0.312
LARYNGEAL2:POSITION3	0.183	0.143	5205.008	1.28	0.201	0.075	0.038	5617.112	1.989	0.047
LARYNGEAL1:RATE DEV.	0.013	0.142	525.564	0.091	0.928	-0.079	0.041	633.019	-1.922	0.055
LARYNGEAL2:RATE DEV.	-0.139	0.101	322.279	-1.374	0.17	-0.005	0.026	4381.899	-0.177	0.86
LARYNGEAL1:GENDER	-0.062	0.136	76.176	-0.454	0.651	-0.155	0.046	123.46	-3.374	0.001
LARYNGEAL2:GENDER	-1.048	0.213	127.636	-4.914	< 0.001	0.16	0.036	118.285	4.486	< 0.001
YOB':GENDER	0.285	0.103	115.084	2.757	0.007	0.029	0.019	112.736	1.521	0.131
YOB'':GENDER	-0.382	0.615	110.94	-0.621	0.536	0.06	0.114	110.257	0.524	0.601
YOB':LARYNGEAL1:HEIGHT	-0.018	0.118	60.096	-0.154	0.878	0.068	0.031	71.561	2.151	0.035
YOB':LARYNGEAL2:HEIGHT	-0.04	0.098	84.148	-0.405	0.687	0.024	0.022	63.715	1.105	0.273
YOB':LARYNGEAL1:FREQ.	-0.083	0.088	56.599	-0.939	0.352	0.006	0.025	78.795	0.233	0.816
YOB':LARYNGEAL2:FREQ.	0.061	0.083	53.802	0.735	0.466	-0.005	0.021	55.19	-0.241	0.81
YOB'':LARYNGEAL1:HEIGHT	-0.532	0.528	58.321	-1.008	0.318	-0.169	0.147	69.039	-1.151	0.254
YOB'':LARYNGEAL2:HEIGHT	0.815	0.474	80.076	1.72	0.089	-0.073	0.101	60.468	-0.723	0.472
YOB'':LARYNGEAL1:FREQ.	0.953	0.383	51.374	2.489	0.016	0.103	0.118	72.869	0.866	0.389
YOB'':LARYNGEAL2:FREQ.	-0.424	0.366	58.429	-1.159	0.251	0.023	0.094	57.789	0.249	0.804

### 4.1.1 f0

The significant main effects of LARYNGEAL2 ( $\hat{\beta} = 4.149, p < 0.001$ ) and LARYNGEAL1 ( $\hat{\beta} = -1.038, p < 0.001$ ) show that lax stops have lower f0 than aspirated stops and non-tense stops have lower f0 than tense stops, averaging over other variables. No main effects of YOB reach significance. There is a significant interaction between LARYNGEAL2 and YOB (linear:  $\hat{\beta} = -0.362, p < 0.001$ ; nonlinear:  $\hat{\beta} = 1.692, p < 0.001$ ), which can be interpreted using Figure 4 (lower-left): the difference in f0 between lax and aspirated stops increases over time, confirming that Seoul Korean is undergoing a sound change. In addition, this change slows down among speakers born after 1960. There is also a significant interaction between LARYNGEAL1 and YOB (linear:  $\hat{\beta} = 0.145, p = 0.006$ ; nonlinear:  $\hat{\beta} = -0.614, p = 0.016$ ), whose interpretation (Figure 4, lower-left) is that the difference in f0 between tense and nontense stops is increasing over time, and that the change in tense stops slows down, keeping pace with aspirated stops.

Turning to gender effects: male speakers use a smaller f0 difference in contrasting aspirated and lax stops than female speakers (LARYNGEAL2:GENDER:  $\hat{\beta} = -1.048, p < 0.001$ ), which can be interpreted as the sound change (f0 contrast enhancement) being more advanced for female speakers. The f0 difference between tense and non-tense stops does not significantly differ by gender (LARYNGEAL1:GENDER:  $p = 0.651$ ). The significant main effect of gender (GENDER:  $\hat{\beta} = -1.248, p < 0.001$ ) and interaction with time (YOB':GENDER:  $\hat{\beta} = -0.285, p = 0.007$ ; YOB'':GENDER:  $p = 0.536$ ) also plausibly reflect the sound change: speakers for whom the sound change is more advanced (female speakers, younger speakers) have higher 'average f0' across the three laryngeal classes (Figure 4, lower-left).

### 4.1.2 VOT

There is a significant main effect of YOB (linear:  $\hat{\beta} = -0.034, p = 0.006$ ; nonlinear:  $p = 0.959$ ), with VOT, averaged across laryngeal categories, becoming shorter over time. Aspirated stops have significantly longer VOT than lax stops, averaged across other variables (LARYNGEAL2:  $\hat{\beta} = 0.221, p < 0.001$ ). VOT is also greater for non-tense stops than for tense stops (LARYNGEAL1:  $\hat{\beta} = 1.242, p < 0.001$ ), which is consistent with VOT continuing to serve as the primary cue differentiating tense from lax/aspirated stops. The significant interaction between LARYNGEAL2 and YOB (linear:  $\hat{\beta} = -0.118, p < 0.001$ ; nonlinear:  $\hat{\beta} = 0.173, p = 0.014$ ) can be interpreted using Figure 4 (lower-right): the difference in VOT between lax and aspirated stops is decreasing over time, confirming that part of the ongoing sound change is the loss of the aspirated/lax VOT contrast. In addition, the change slows down over time (nonlinear term), though not as dramatically as was the case for f0. Finally, the VOT difference between tense and nontense stops also decreases over time (YOB':LARYNGEAL1:  $\hat{\beta} = -0.061, p = 0.001$ ; YOB'':LARYNGEAL1:  $p = 0.684$ ), primarily due to change in aspirated stop VOT (Figure 4, lower-right).

Male speakers have significantly longer VOT than female speakers, across laryngeal categories (GENDER:  $\hat{\beta} = 0.127, p < 0.001$ ), and the VOT differences between aspirated and lax stops and between tense and non-tense stops are larger for male speakers (LARYNGEAL2:GENDER:  $\hat{\beta} = 0.16, p < 0.001$ ; LARYNGEAL1:GENDER:  $\hat{\beta} = -0.155, p = 0.001$ ). All these effects can be interpreted using Figure 4 (right panels), as the sound change being more advanced for female speakers.<sup>4</sup> Interestingly, the VOT values for aspirated stops and lax stops are reversed for the youngest speakers. This is consistent with Silva (2006), who found a negative

---

<sup>4</sup>Note that the overall gender difference in VOT is unlikely to be due to physiological differences, which would if anything suggest women should have *higher* VOT than men (Morris, McCrea & Herring, 2008).

aspirated/lax VOT difference for a handful of young speakers.

### 4.1.3 Summary

We found that the aspirated/lax distinction in Seoul Korean has shifted over time from primarily VOT-based to primarily f0-based, this change is more advanced for female speakers, VOT contrast reduction and f0 contrast enhancement are proceeding in parallel, and tense stops pattern together with aspirated stops in f0 change (but to a lesser extent). These findings all replicate Kang (2014) on a significantly larger dataset.

## 4.2 Change across words

### 4.2.1 Word Frequency

We now discuss the effects of word frequency on VOT and f0 predicted by the models, which addresses our first two research questions: is there a word frequency effect in this sound change, and how is this sound change spreading across the lexicon of Seoul Korean? We examine the directionality of any word frequency effect, whether this directionality is the same for VOT and f0, and whether the role of frequency changes over time, all of which offer evidence for the mechanism behind this sound change. The relationship of word frequency with VOT and f0 are captured in the models (Table 2) by terms for the main effect of FREQUENCY and its interactions with LARYNGEAL2 and YOB. Three-way interactions will be discussed in Sec. 4.2.3. Figures 5–6 show the empirical and model-predicted effects of word frequency on VOT and f0.

#### 4.2.1.1 f0

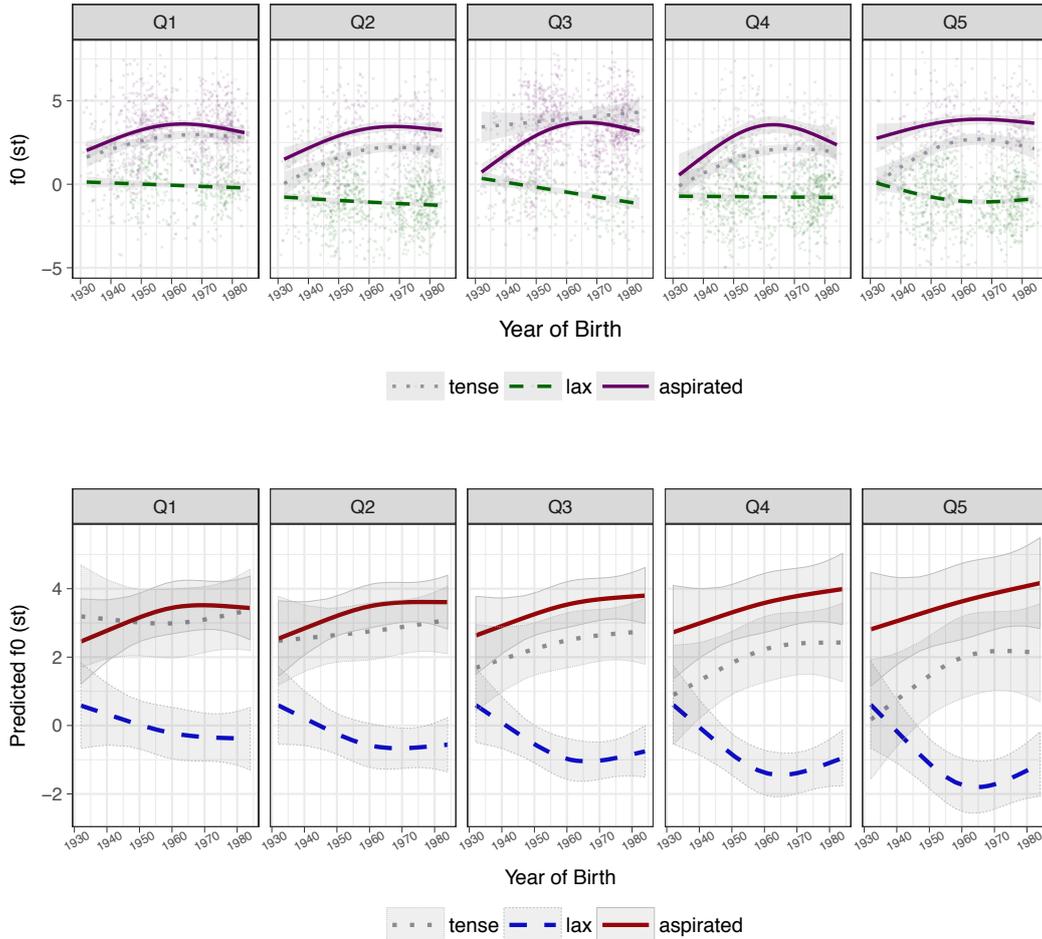
There is a marginal negative effect of word frequency on f0 (FREQUENCY:  $\hat{\beta} = -0.304$ ,  $p = 0.054$ ), suggesting that frequently used words have lower f0 than infrequently used words. This may be due to factors observed cross-linguistically: high-frequency words tend to be produced with lower pitch (Cantonese: Zhao & Jurafsky, 2007, 2009) and phrasal prominence is reduced with higher predictability (English: Pan & Hirschberg, 2000).

We find a marginal interaction between laryngeal class and frequency (LARYNGEAL2:FREQUENCY  $\hat{\beta} = 0.625$ ,  $p = 0.072$ ), such that the difference in f0 between aspirated and lax stops is greater for high-frequency words (averaging across speakers of different ages). This effect is visible in Figure 5 as an increasing distance between the lines corresponding to aspirated and lax stops, as frequency increases.<sup>5</sup> When this frequency effect is interpreted with the significant YOB:LARYNGEAL2 interaction seen above, the diachronic divergence in f0 between laryngeal classes is more advanced for high-frequency words. Note that this diachronic pattern is unlikely to result from a synchronic magnitude effect, which would if anything predict *smaller* f0 differences between laryngeal classes for higher frequency words (since they would be more predictable, and hence less informative; e.g. Aylett & Turk, 2006), the opposite of the pattern observed here.

An additional observation can be made from Figure 5 for tense stops, for which f0 appears to be increasing over time along with aspirated stops, as a member of the same natural class

---

<sup>5</sup>A reviewer notes discrepancies between the empirical trends and model fits in word frequency effects on both VOT and f0 (in Figure 5 and Figure 6). These discrepancies are largely due to unbalanced data in terms of frequency and vowel height. Low-frequency words are skewed towards nonhigh vowel contexts and high-frequency words are skewed towards high vowel contexts. When the same plots are made for just tokens with a fixed vowel height, the empirical plots look much closer to the model prediction plots.



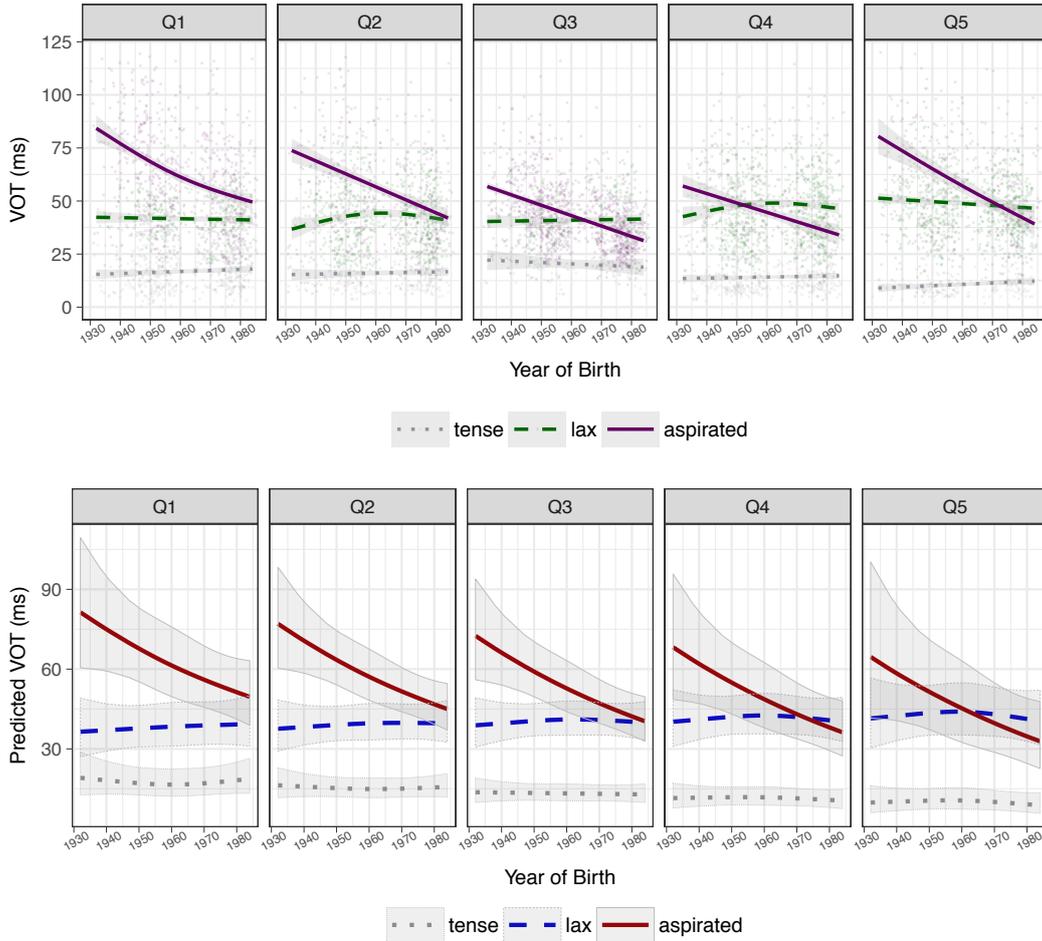
**Figure 5:** Empirical plots (top) and model prediction plots (bottom) of  $f_0$  as a function of word frequency & laryngeal category. Lines and shadings as in Figure 4. Q1–Q5 refer to word frequency quantiles from lowest (Q1) to highest (Q5).

(as proposed by Kang, 2014). However, for tense stops, the change in  $f_0$  is more advanced before *lower* frequency words. This pattern makes sense if  $f_0$  in tense stops is changing by analogy with aspirated stops—since low-frequency words are expected to lead analogical sound changes.<sup>6</sup>

#### 4.2.1.2 VOT

High-frequency words have significantly shorter VOT than low-frequency words (FREQUENCY:  $\hat{\beta} = -0.11$ ,  $p = 0.031$ ), averaged across speakers and stop categories. This directionality is expected, as a synchronic effect, independent of sound change in progress: higher-frequency words show shorter segmental durations due to hypoarticulation (e.g. Aylett & Turk, 2004; Baker & Bradlow, 2009; Bell et al., 2003). There is also a marginal interaction of frequency with laryngeal class (LARYNGEAL2:FREQUENCY:  $\hat{\beta} = -0.185$ ,  $p = 0.095$ ), such that the VOT difference between lax and aspirated stops is smaller for high-frequency words (averaging across speakers of different ages). This effect is visible in Figure 6 as a decreasing distance between the lines corresponding to aspirated and lax stops, as frequency increases, due primarily to VOT for aspirated stops decreasing. When this frequency effect is interpreted in view of the diachronic change (LARYNGEAL2:YOB), it suggests that the diachronic merger of

<sup>6</sup>We thank an anonymous reviewer for this suggestion.



**Figure 6:** Empirical plots (top) and model prediction plots (bottom) of VOT as a function of word frequency & laryngeal category. Lines and shadings as in Figure 4. Q1–Q5 refer to word frequency quantiles from lowest (Q1) to highest (Q5).

VOT happens earlier for high-frequency words.

We also note the marginal interaction of LARYNGEAL1 with word frequency (LARYNGEAL1:FREQUENCY:  $\hat{\beta} = 0.193$ ,  $p = 0.096$ ): the difference in VOT between tense and nontense stops is larger for words with higher frequency; this is due to a negative relationship between word frequency and VOT for tense stops and a positive relationship for lax stops (Figure 6 bottom). We do not have an explanation for this pattern, and leave the more general question of the role of tense stops in this sound change to future work.

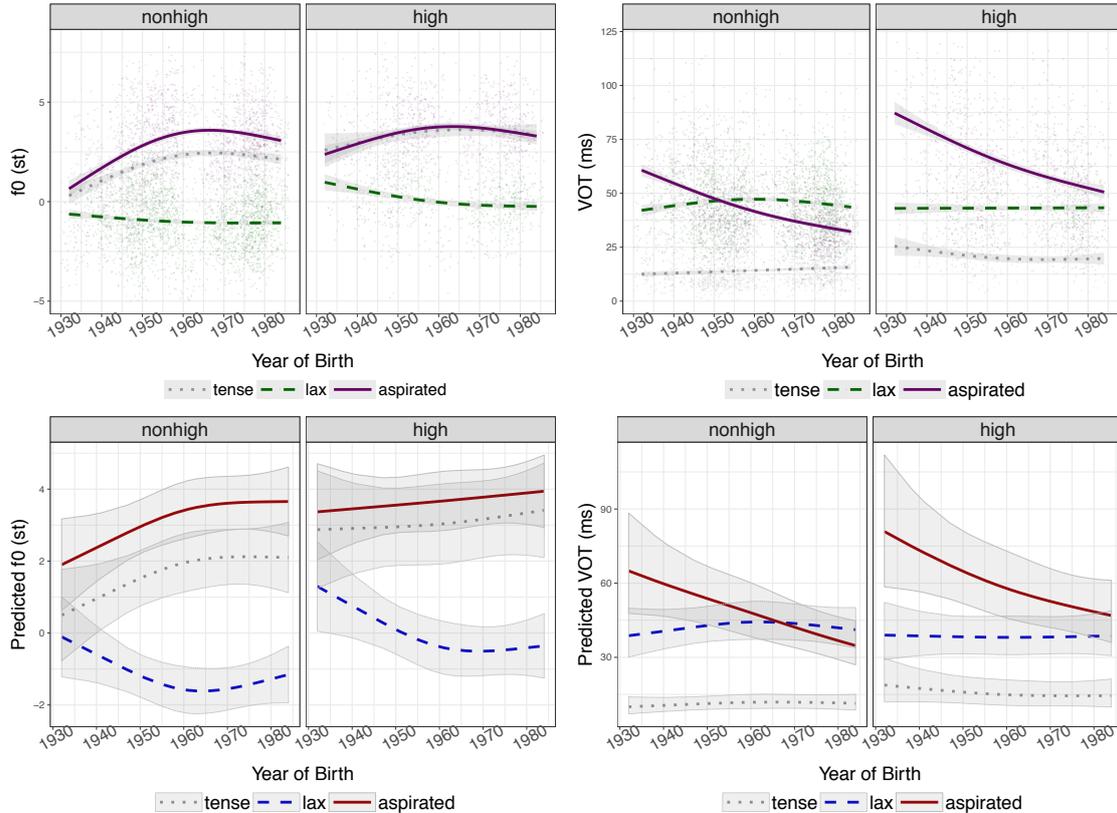
#### 4.2.2 Vowel Height

We turn to the effect of vowel height on VOT and  $f_0$ , which addresses our second and third questions: how is the change propagating across vowel contexts, and how is the magnitude of vowel-height dependent IF0 effects influenced by the emergence of contrastive  $f_0$ ? We examine the directionality of any vowel height effect, whether this directionality is the same or different for VOT and  $f_0$ , and whether the IF0 effect varies over time and across stop categories.

The relationship between vowel height and each cue (VOT,  $f_0$ ), and how it changes over time, are captured in the models (Table 2) by terms for the main effect of HEIGHT and its interaction with YOB. Differences in IF0 effects and how the IF0 effect changes over time for each laryngeal class are captured by LARYNGEAL:HEIGHT and YOB:LARYNGEAL:HEIGHT

interaction terms.

Figure 7 shows the empirical and model-predicted effects of vowel height on VOT and  $f_0$ , and Figure 8 shows the diachronic development of this effect for each stop category.



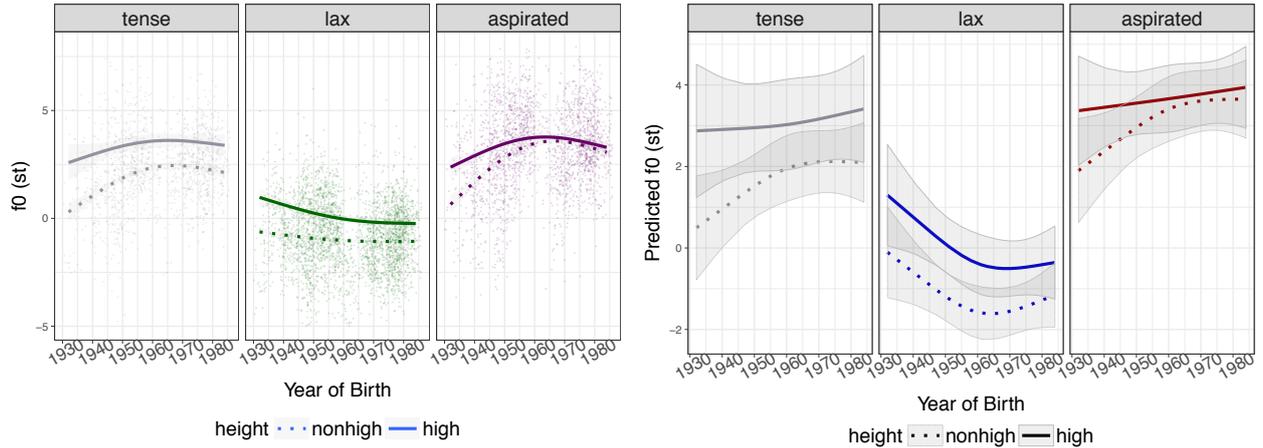
**Figure 7:** Empirical plots (top) and model prediction plots (bottom) of  $f_0$  (left) and VOT (right), as a function of vowel height and laryngeal category. Lines and shadings as in Figure 4.

#### 4.2.2.1 $f_0$ : across vowel context

Concerning our second research question, the  $f_0$  difference between aspirated and lax stops is modulated by vowel height. As can be seen in Figure 7, the difference in  $f_0$  between aspirated and lax stops is greater for stops in nonhigh vowel context than for those in high vowel context (averaging across speakers of different ages) (LARYNGEAL2:HEIGHT:  $\hat{\beta} = -0.692$ ,  $p = 0.04$ ). When this height effect is interpreted in reference to the ongoing sound change across speakers, it indicates that the divergence of  $f_0$  over time is more advanced in nonhigh vowel context than in high vowel context.

#### 4.2.2.2 IF0 effects

There is a significant main effect of HEIGHT: as expected (Sec. 2.4), high vowels have intrinsically higher  $f_0$  than low vowels ( $\hat{\beta} = 0.96$ ,  $p < 0.001$ ). More importantly, as illustrated in Figure 8, we find a significant interaction between YOB and HEIGHT: the linear term suggests that the intrinsic difference in  $f_0$  between high and nonhigh vowels is attenuated over time as contrastive  $f_0$  emerges in the language ( $\hat{\beta} = 0.142$ ,  $p = 0.007$ ), while the nonlinear term suggests that this attenuation in IF0 effects is slowing down ( $\hat{\beta} = -0.583$ ,  $p = 0.018$ ). The pattern of slowing down fits with the significant interaction between YOB (nonlinear) and LARYNGEAL2 observed for change across speakers. Together, the YOB'' :HEIGHT and



**Figure 8:** Empirical plots (left) and model prediction plots (right) showing a change in the size of IF0 effects over time by each laryngeal category. Lines and shadings as in Figure 4.

YOB'':LARYNGEAL2 effects suggest that IF0 attenuation is decelerating as the sound change is nearing completion in phrase-initial position.

#### 4.2.2.3 VOT

There is a significant effect of HEIGHT in the cross-linguistically expected direction (Higgins et al., 1998; Stevens, 1998): VOT is longer for stops before a high vowel than before a non-high vowel ( $\hat{\beta} = 0.134$ ,  $p = 0.036$ ). This difference is attenuated over time (YOB':HEIGHT:  $p = 0.393$ ; YOB'':HEIGHT:  $\hat{\beta} = 0.154$ ,  $p = 0.016$ ).

Crucially, the VOT difference between lax and aspirated stops is significantly smaller in non-high vowel context than in high vowel context (LARYNGEAL2:HEIGHT:  $\hat{\beta} = 0.326$ ,  $p = 0.002$ ). Similarly to the results for f0, this vowel height effect has a clear interpretation in terms of sound change when interpreted together with the community-level change results: the diachronic merger of VOT for the aspirated/lax stop contrast observed across the speech community occurs earlier in nonhigh vowel context.

#### 4.2.3 Magnitude versus timing effects

In Sec. 2.3.1 above we considered the issue of whether the effects we have observed can be interpreted as effects of magnitude (i.e. pre-existing synchronic differences between classes of words that are maintained during diachronic change) or timing (i.e. diachronic change proceeding faster or earlier in some environments). So far we have interpreted our results to mean that non-high vowels and high-frequency words are leading the change in VOT and f0 contrasts—that is, we have interpreted them as timing effects. We now consider to what extent we have evidence for this claim.

As explained in Sec. 2.3.1, timing differences should manifest themselves across the full time range of the sound change as differences in the rate of change over time—corresponding to three-way interaction terms in the statistical models between year of birth, laryngeal class and either frequency or vowel height. We consider only terms involving LARYNGEAL2 (aspirated/lax contrast), which are of interest for the sound change, and do not discuss terms involving LARYNGEAL1 (tense/non-tense contrast). In unpacking these terms, we will use the plots in Figure 9, which show the model-predicted difference in VOT and f0 between aspirated and lax stops over time, for words with different frequencies and with different

vowel heights (with other variables held constant, as above).<sup>7</sup>

We first consider three-way interactions with frequency, the evidence for which was mixed. For  $f_0$ , the direction of the interaction between LARYNGEAL2, year of birth and frequency is consistent with a timing effect, where the sound change has progressed more over time for high-frequency words, as can be seen in Figure 9(c). However, this interaction does not reach significance (YOB':LARYNGEAL2:FREQUENCY:  $p = 0.466$ ; YOB'':LARYNGEAL2:FREQUENCY:  $p = 0.251$ ). For VOT, the interaction between laryngeal class, and frequency has both very small effect size and does not reach significance ( $p = 0.389$ ), as is clear in Figure 9(d).

Turning to the three-way interaction terms with vowel height: for  $f_0$ , there is a marginal interaction of LARYNGEAL2 with vowel height and year of birth (YOB'':LARYNGEAL2:HEIGHT:  $p = 0.089$ ). This trend indicates that the magnitude of the nonlinear change in LARYNGEAL2 over time differs by vowel context, as shown in Figure 9(a): the enhancement of the  $f_0$  contrast is more advanced in nonhigh vowel contexts than high vowel contexts, as expected for a timing effect where stops in nonhigh vowel contexts lead the change. (Alternatively, this trend may be interpreted as a difference in the magnitude of IF0 attenuation over time between stop categories.) For VOT, the interaction of LARYNGEAL2 with vowel height and year of birth has small effect size and does not reach significance (YOB':LARYNGEAL2:HEIGHT:  $p = 0.273$ ; YOB'':LARYNGEAL2:HEIGHT:  $p = 0.472$ ) (Figure 9(b)).

In sum, the three-way interactions (between LARYNGEAL2, year of birth, and frequency/vowel height) for  $f_0$  were generally in the direction predicted under a timing effect interpretation, but the weak significances of these terms mean that they do not offer strong evidence for this interpretation over a magnitude interpretation. Any three-way interactions for VOT were negligible. Like all null results, the  $f_0$  and VOT three-way interaction results are not meaningful a priori since there are many reasons a “real” effect may have not been detected if it existed. One such reason is suggested by the model-predicted VOT and  $f_0$  contrasts over time, for different classes of words in Figure 9, which can be compared directly to the trajectories that were predicted under magnitude versus timing effects in Figure 2.

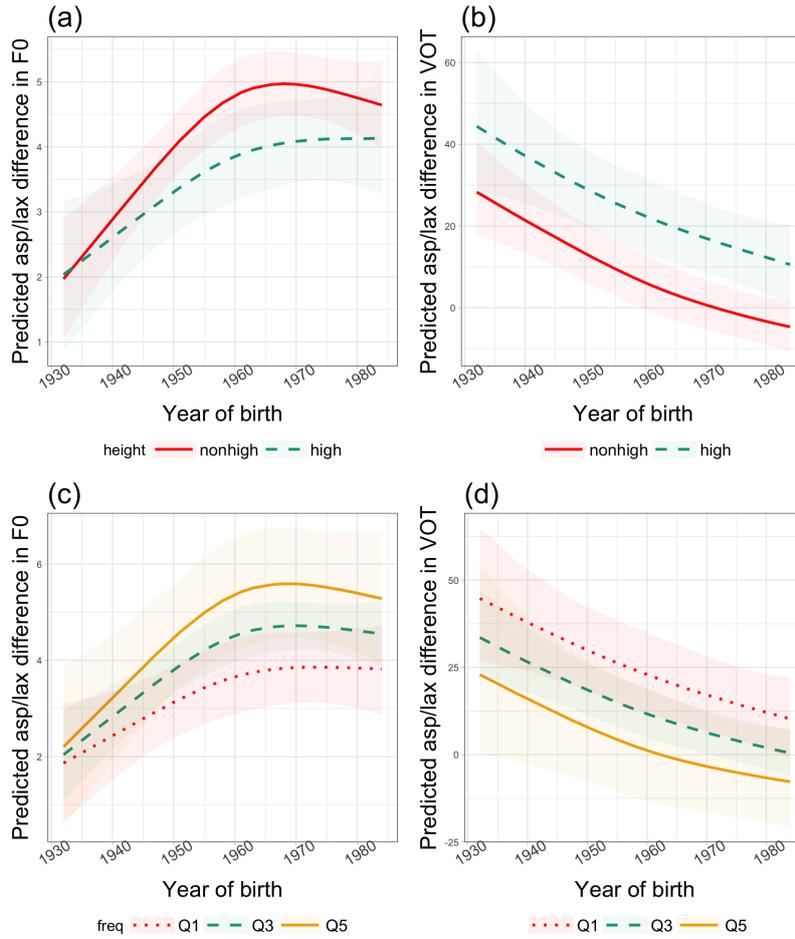
Crucially, the slopes for  $f_0$  difference seem to vary across words with different frequency (Figure 9(c)) and stops in different vowel contexts (Figure 9(a)) until the change becomes stabilized (compare to Figure 2(c)). In contrast, the slopes for VOT differences do not exhibit noticeable differences across words (Figure 9(d)) and vowel contexts (Figure 9(b)). Thus, the null effects for  $f_0$  in the three-way interactions involving YOB—particularly for the linear term (YOB':LARYNGEAL2:FREQUENCY and YOB':LARYNGEAL2:HEIGHT)—may be due in part to reduced variation in the stable portion at the endpoint of the S-curve.

To test this idea, we carried out a post-hoc analysis by building a new  $f_0$  model on just data from speakers born before 1965. The time band was chosen because the empirical and model prediction plots show the beginning of stabilization in the 1960s, consistent with previous work (Silva, 2006) that found a critical divide around 1965 between ‘traditionalists’ (VOT users) and ‘innovators’ ( $f_0$  users).

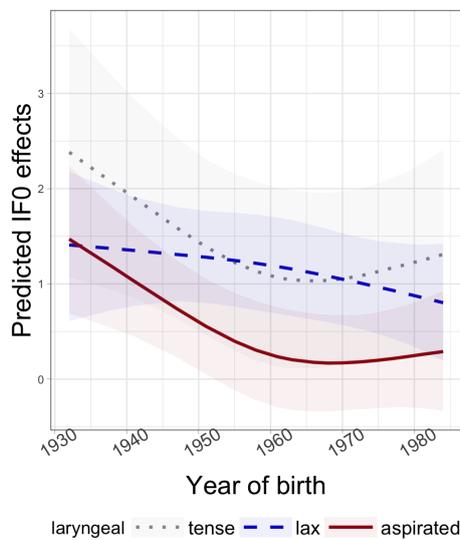
On the subsetted data, the new model was constructed in the same way as the previous model, keeping most terms the same. Because we intended this model to only include the linear trend for year of birth, the nonlinear term YOB'' and all interaction terms involving YOB'' were excluded. The fixed effects for the new  $f_0$  model are summarized in Table A1 in the appendix. We omit discussion of most results of this model, which largely overlap with our previous  $f_0$  model, and report only the two three-way interaction terms of interest:

---

<sup>7</sup>Model predictions and 95% prediction intervals were approximated by simulation. For each model (VOT and  $f_0$ ),  $n = 10000$  draws of the fixed effect coefficients ( $\vec{\beta}$ ) from the model’s posterior distribution were taken using the `sim` function in the `arm` package (Gelman & Su, 2015), then used to calculate a median prediction and 95% prediction intervals, which correspond to the lines and shading in in Figure 9 and Figure 10.



**Figure 9:** Model-predicted differences between aspirated and lax stop VOT and  $f_0$  over time, for different vowel heights (top row) and word frequencies (bottom row). Lines and ribbons are median model predictions and 95% prediction intervals calculated by simulation from the model posterior. Q1–Q5 refer to word frequency quantiles from lowest (Q1) to highest (Q5).



**Figure 10:** Model-predicted IF0 effect ( $f_0$  difference between high and non-high vowels) over time, for each class of stops. Lines and ribbons are as in Figure 9.

YOB:LARYNGEAL2:FREQUENCY and YOB:LARYNGEAL2:HEIGHT.

Crucially, both terms are statistically significant with notably increased effect sizes relative to the earlier f0 model (YOB:LARYNGEAL2:FREQUENCY:  $\hat{\beta} = 1.186$ ,  $p = 0.019$ ; YOB:LARYNGEAL2:HEIGHT:  $\hat{\beta} = -1.561$ ,  $p = 0.01$ ). This indicates that for speakers born up to 1965, high-frequency words and stops preceding a nonhigh vowel are *ahead* in the change in f0 contrast enhancement, and these effects are beyond synchronic magnitude effects. Thus, this model provides our best evidence of observing timing rather than magnitude effects.

The three-way interaction between year of birth, laryngeal class, and vowel height (YOB:LARYNGEAL2:HEIGHT) also adds to our interpretation of how the IF0 effect changes over time. The IF0 attenuation is significantly greater for aspirated stops than for lax stops, as can be seen in Figure 10, which shows the model-predicted IF0 effect (f0 for high minus non-high vowels) of stops over time (with other variables held constant).<sup>8</sup> By 1965, the predicted IF0 difference approaches zero for aspirated stops, but is still positive for lax stops. Another interesting pattern is that the IF0 effect for tense stops always remains larger than for aspirated stops. The different development of IF0 effects over time for different stop classes is discussed further below (Sec. 5.3).

#### 4.2.4 Frequency versus vowel height effects

Before proceeding, we note that the frequency effects observed in our data are weaker than the corresponding vowel height effects—especially for VOT—with the frequency effects of interest often having higher p-values and smaller effect sizes than the analogous vowel height effects of interest. One possible explanation for the asymmetry between frequency and vowel height effects is that more meaningful frequency effects exist, but are masked due to the distribution of the data and our statistical methodology. In this dataset, we found high multicollinearity between frequency, place of articulation, and vowel height, leading to unstable models when terms for all of their interactions with laryngeal category were included. Because word frequency and vowel height are central to our research questions, we had to exclude the interaction of place of articulation and laryngeal category. However, because place of articulation is a priori expected to affect VOT, we retained the main effect of PLACE. We also included all possible random slopes. Both aspects of our modeling methodology may lead to conservative p-values, while prioritizing accurate coefficient estimates (see Sec. 3.3.2).<sup>9</sup> Thus, in the remainder of this paper, we acknowledge the weakness of some frequency effects ( $p < 0.1$ ) in our results by labeling them as ‘tentative’, but discuss the *direction* of these effects nonetheless.

Another possibility is that the true frequency effects in this dataset are weaker than the height effects—as reflected by the model results. Yet another possibility is that a frequency effect on VOT exists as a synchronic effect, but its role is limited to triggering the change. We return to these possibilities below (Sec. 5.2), in the context of what each one would mean for our research questions.

#### 4.2.5 Summary

We found that VOT contrast reduction and f0 contrast enhancement are greater in stops preceding a nonhigh vowel, and tentatively greater in words with high frequency. In a further analysis exploring a period of time (year of birth < 1965) where there is more variation in

---

<sup>8</sup>These model predictions and prediction intervals are calculated using the same simulation-based method as for the aspirated/lax differences.

<sup>9</sup>Indeed, removing all terms for PLACE lowered the p-values of all frequency effects.

f0, we found evidence that this f0 pattern can be interpreted as a timing effect: f0 contrast enhancement is spreading across words of different frequencies and vowels of different heights in a non-uniform way. The parallel frequency and vowel height effects on VOT merger and f0 contrast enhancement offer important evidence for our proposal, discussed below, that this sound change results from a combination of contrast reduction in one dimension (VOT) and adaptive behavior in another (f0) to preserve the contrast. We also found evidence that the universal trend of IF0 difference between high and nonhigh vowels is attenuated over time as contrastive f0 emerges in the language, and that the effect differs by stop category.

### 4.3 Other Factors

We briefly discuss the f0 and VOT results for variables included in our model as controls (POSITION, SPEAKER MEAN RATE, RATE DEVIATION, PLACE), restricting ourselves to terms which are significant ( $p < 0.05$ ) or are relevant for our research questions.

f0 was higher for utterance-medial stops before a longer pause than before a shorter pause (POSITION2:  $\hat{\beta} = 0.243$ ,  $p < 0.001$ ), perhaps due to larger f0 resets at prosodic boundaries signaled by longer pause durations (Fant, Kruckenberg & Gustafson, 2002). There is a marginal trend for the f0 difference between lax and aspirated stops to be greater for utterance-medial stops than for utterance-initial stops (LARYNGEAL2:POSITION1:  $\hat{\beta} = 0.527$ ,  $p = 0.076$ ).

f0 increases for faster speech within a speaker, averaging across stops (RATE DEVIATION:  $\hat{\beta} = 0.157$ ,  $p = 0.011$ ). There is a trend for faster speech to be associated with reduction in the f0 contrast (LARYNGEAL2:RATE DEVIATION:  $p = 0.17$ ), but this effect does not reach significance.

VOT is higher utterance-medially than utterance-initially (POSITION1:  $\hat{\beta} = 0.1$ ,  $p = 0.005$ ), and there is a trend for the VOT difference between aspirated and lax stops to be larger utterance-medially (LARYNGEAL2:POSITION1:  $\hat{\beta} = 0.158$ ,  $p = 0.092$ ). Among utterance-medial stops, the aspirated/lax contrast in VOT is greater following a long pause than before shorter pauses (LARYNGEAL2:POSITION3:  $\hat{\beta} = 0.075$ ,  $p = 0.047$ ). The lack of a significant SPEAKER MEAN RATE effect on VOT ( $p = 0.546$ ) suggests that the reduced VOT contrast for younger speakers cannot be attributed to age-dependent speech rate variation. Faster speech does not significantly affect the aspirated/lax VOT contrast (LARYNGEAL2:RATE DEVIATION:  $p = 0.86$ ). Finally, VOT is larger for less anterior places of articulation, as expected cross-linguistically (PLACE1 (labial vs. nonlabial):  $\hat{\beta} = 0.123$ ,  $p = 0.012$ ; PLACE2 (alveolar vs. velar):  $\hat{\beta} = 0.314$ ,  $p < 0.001$ ).

## 5 Discussion

In the current study, we first confirmed previous findings (Kang, 2014) on the quasi-tonogenetic sound change underway in Seoul Korean in phrase-initial position: the change is led by female speakers, for both VOT contrast reduction and f0 contrast enhancement; the diffusion of the change through the speech community proceeds by a gradual parallel change of VOT and f0 in the inverse direction over time; and the change is slowing down, suggesting it is nearing completion in the speech community. During this sound change, VOT ‘reduction’ (over time) comes largely from aspirated stops, while lax stops show little change. This asymmetry between aspirated and lax stops parallels how reduction affects VOT of different stop classes cross-linguistically (synchronically): in hypospeech, the contrast between long and short-lag stops is attenuated, mainly due to decrease in the long-lag stop’s VOT, as

observed in languages including English, Icelandic, and Thai (Kessinger & Blumstein, 1997; Miller et al., 1986; Pind, 1995). Our proposal that the sound change is driven in part by production bias in VOT provides a natural explanation for why change in VOT affects only aspirated stops, rather than lax stops or both stop classes.

We then provided three novel findings. First, while not definitive, our results lead us to tentatively conclude that sound change impacted high-frequency words before low-frequency words, suggesting that lenition may be a driving factor of the sound change. Second, the change is spreading through words and vowel contexts as well as speakers in an adaptive manner: both VOT contrast reduction and  $f_0$  contrast enhancement are greater in the same conditions. Third, the vowel intrinsic  $f_0$  difference between high and non-high vowels is attenuated as contrastive  $f_0$  emerges over time. These findings suggest that transphonologization in Seoul Korean is driven by production bias and adaptive reinterpretation of the speech signal.

We now discuss our results on the quasi-tonogenetic change in progress in Seoul Korean, which shed light on the origin, progression, and the impact of tonogenesis.

## 5.1 Origin: Production bias

Our first research question concerned word frequency effects in tonogenesis. There is almost no previous work addressing the role of word frequency in tonogenetic sound change. The results tentatively suggested that high-frequency words are produced with more innovative pronunciation than low-frequency words, by having both a greater  $f_0$  difference and smaller VOT difference between lax and aspirated stops. From this finding we suggested that quasi-tonogenesis in Seoul Korean may be driven by contrast reduction—namely, production bias affecting VOT—and that production bias in VOT may be one source of tonogenetic sound change more generally.

This interpretation of the word frequency effect is based on work showing lenition-driven sound change tends to affect high-frequency words first (see Sec. 2.2). We do not commit to “when” frequency plays a role (acquisition vs. adulthood), or “where” (mental representation), which are the subject of significant debate (Bell et al., 2009; Gahl, 2008; Harrington, Kleber & Stevens, 2016; Kang, Yoon & Han, 2015; Labov, 2007; Pierrehumbert, 2002).

Much work on frequency effects in sound change assumes an exemplar-theoretic model of mental representation. Hay et al. (2015) argue that such a model implies that in ongoing changes, high-frequency words should show different *rates* of change. We included terms in our statistical models to test the possibility that the pattern of  $f_0$  contrast enhancement or VOT contrast reduction over time differs depending on word frequency (a ‘timing effect’).

We did not find clear evidence that it does for VOT. One possible explanation of this null result is that smaller VOT contrasts in high frequency words are limited to the role of a synchronic trigger of diachronic  $f_0$  enhancements. In this case, the effects of word frequency on VOT in our data could be the result of synchronic phonetic bias maintained during the community change (Fruehwald, Gress-Wright & Wallenberg, 2013). Alternatively, like any null result, the lack of timing effects may not be meaningful (e.g. insufficient statistical power). Even if the prediction of different trajectories for high and low-frequency words were correct, their trajectories would not necessarily differ during the late stages of a sound change, when there is relatively little variation between speakers. Indeed, several aspects of our data suggest that this sound change is nearing completion in phrase-initial position, especially for VOT. In addition, in subsequent work on the same dataset (Bang, Sonderegger & Clayards, 2017), we have found that age explains much less by-speaker variability in VOT

contrasts than by-speaker variability in f0 contrasts, supporting the idea that VOT contrast reduction has ended earlier than f0 contrast enhancement.

There was more evidence for timing effects on f0, with a strong trend in the direction predicted by Hay et al. when only the dynamic middle of the change was considered, suggesting that words with higher frequency are further along in the change in f0. Our data cannot distinguish between two possible sources of this effect: a difference in the rate of change and a difference in the time of inception. Both distinguishing these two possibilities and reaching a more definitive conclusion for VOT require data from a broader time range, including the period when the change initiated. We leave this to future studies.

In sum, the word frequency effects in our data, while tentative, are compatible with the idea that transphonologization in Seoul Korean was triggered by production bias, affecting VOT. We now consider how and why this change was propagated across VOT *and* f0.

## 5.2 Progression: Adaptive link

Our second research question was how a new f0 contrast propagates across words and vowel contexts as it spreads across speakers. We found strong evidence for a gradual tradeoff between the two cues across words (i.e. high- vs. low-frequency) and vowel contexts (i.e. high vs. nonhigh vowels), supporting the idea that the change involves adaptivity at the language level as well as the speaker level.

The effect of word frequency on f0 contrast enhancement, though tentative, is important evidence for an adaptive link between VOT and f0, because it cannot be explained as a synchronic effect. Based on previous work, one would normally expect a contrastive difference in f0 to be enhanced in words with *lower* frequency, by the logic that higher-frequency words are generally more hypoarticulated (e.g. Aylett & Turk, 2004, 2006; Baker & Bradlow, 2009; Bell et al., 2003). We are not aware of work testing this prediction for a non-tonal language, but it is borne out for Cantonese (Zhao & Jurafsky, 2007, 2009), where tonal range is enlarged for lower-frequency words. By this logic, in the Korean case, if there were no adaptive link between VOT and f0, we would expect VOT contrast reduction to impact high-frequency words first and f0 contrast enhancement to impact low-frequency words first (A and D in Figure 1) due to independent synchronic pressures operating on each cue, combined with an ongoing sound change progressing at the community level. The pattern we observe instead (A and C in Figure 1), suggests that the sound change is driven by production bias in VOT (A), and the change in f0 (C) is an adaptive response.

Our data also showed a parallel pattern for phonetic contexts. We found more enhanced f0 difference and more reduced VOT difference between lax and aspirated stops in non-high vowel contexts than in high vowel contexts. Crucially, by considering just speakers born before 1965, we found that stops preceding non-high vowels are leading the change in f0. As for the frequency effects, our logic is that this parallelism is expected if the cues are adaptivity-linked, and not expected otherwise. An important caveat is that, unlike for the frequency effects, we do not know whether this parallelism in the vowel height effect would be expected independent of adaptive sound change, as there is (to our knowledge) no work investigating how the size of the contrast across voicing categories depends on following vowel height by vowel context, for VOT or f0, in any language (but see Bang et al., 2017 for preliminary results).

In the absence of such studies, the trajectory of change in VOT over time in Korean, together with established effects of vowel height on VOT cross-linguistically, suggest a possible relationship between the non-high vowel context and VOT merger. In our data, the diachronic reduction in VOT contrast is primarily due to aspirated stops becoming shorter.

Thus, any phonetic factor conditioning lower VOT for aspirated stops could be thought of as a phonetic precursor for this diachronic change. Cross-linguistically, VOT for long-lag stops is shorter before non-high vowels (e.g. Esposito, 2002; Higgins et al., 1998; Klatt, 1975, for Italian, English). By this logic, VOT shortening for aspirated stops before non-high vowels could be another phonetic precursor (along with frequency effects) triggering VOT contrast reduction. This account would further support our view that reduction in the VOT contrast triggered the sound change, and change in  $f_0$  is an adaptive response. While this account would explain the relationship between vowel height and VOT contrast reduction in Korean, the broader question of whether it is before high or non-high vowels that VOT and  $f_0$  contrasts to laryngeal status are reduced requires future cross-linguistic work.

Our findings on parallel change in VOT and  $f_0$  via production bias and adaptive compensation fit well with a computational study by Kirby (2013) which addresses the question of why it is transphonologization from VOT to  $f_0$  that is affecting Seoul Korean stops, rather than another possible change (e.g. merger). Based on simulations of a community of Seoul Korean speaker/hearer agents under different assumptions, Kirby argues that only assuming that agents have both a *bias* in production (e.g. VOT contrast reduction) and an adaptive response of *enhancement* (e.g.  $f_0$  contrast enhancement) results in the observed diachronic change (Hyman, 2008).

These findings are also in line with the inverse relationship between nasal stop duration and (preceding) vowel nasalization duration observed by Beddor (2009) for English. The shift towards longer vowel nasalization and shorter nasal duration was more advanced in some speakers than in others, and more advanced before voiceless stops than before voiced stops, which parallels our result in that two cues are inversely affected in the same environments. Beddor hypothesizes that production bias on the primary cue (shorter nasal duration; here, VOT) is also the trigger of the enhancement on the co-articulatory effect (longer vowel nasalization; here,  $f_0$ ). This type of compensatory link may also underly the inverse relationship between VOT and  $f_0$  we observed in Korean. Determining whether such covariations are precursors to sound change requires further cross-linguistic work building on existing results on individual languages (Bang et al., 2017).

Before proceeding, the weakness of frequency effects in our data (especially for VOT) bears further discussion (Sec. 4.2.4), since frequency effects are implicated in our account of the ‘origin’ and ‘progression’ of this sound change. In addition to the possibility that multicollinearity and statistical methodology are masking the true strength of frequency effects, it is of course possible that the frequency effects in this data *are* weak, even if such issues did not exist, for other reasons. First, ‘word frequency’ may not be well-defined: frequencies are likely not uniform across age groups (Kim, 2016; Walker & Hay, 2011), and it is not obvious what unit to use in ‘word’ frequency calculations in an agglutinative language such as Korean. Second, factors beyond word frequency contribute to probabilistic speech reduction, such as the co-occurrence frequency and conditional probability of sequences of words (e.g. Bush, 1999; Jurafsky, Bell, Gregory & Raymond, 2001), meaning that frequency effects alone may underestimate the role of reduction in a sound change. These considerations may help explain why frequency effects found in this study are weak—and why reported frequency effects in sound change tend to be weak in general (e.g. Hay et al., 2015; Kang et al., 2015). Weak does not, however, mean unimportant (but see Labov, 2010), for purposes of understanding the mechanism of a sound change. What is crucial for our purposes is that the observed frequency effects offer additional evidence for our main claims: this change is spreading through different groups of words in parallel in both VOT and  $f_0$ , due to production bias coupled with an adaptive response.

### 5.3 Impact: Attenuation of IF0

Our third research question concerned whether the typologically common IF0 difference between high and nonhigh vowels would be maintained or attenuated, as f0 assumes a more contrastive role.

We found that the size of the IF0 effect is attenuated over time. After transphonologization, an unrestricted IF0 effect may create a challenge for listeners in attributing f0 variation to its source, which could threaten contrast preservation, since f0 is now the primary cue. In fact, previous studies suggest that IF0 effects can act as a phonetic precursor for tone splits (Kingston, 2011) or tonal merger (Siddins & Harrington, 2015), though such changes are rare. These studies together with our findings suggest that IF0 is controlled in modern-day Seoul Korean to satisfy language-specific perceptual and phonological needs. The Korean case, where IF0 effects in the *same language* change over time, strongly supports the idea that IF0 effects are to some extent ‘controlled’, previously known from cross-linguistic variation in IF0 effect size (e.g. Berry & Moyle, 2011; Connell, 2002; Fischer-Jørgensen, 1990; Hoole & Honda, 2011).

The attenuation of the IF0 effect in our data was larger for aspirated stops than for lax or tense stops, with the IF0 effect for aspirated stops approaching zero over time. This category-dependent variation may be motivated by the language-specific implementation of stops in Korean. Speakers may not attenuate IF0 effects for tense stops as much as for aspirated stops because there is no functional pressure to do so: the primary cue contrasting tense stops from lax/aspirated stops is VOT, before and after the sound change. The diachronic development of f0 for tense stops also supports this view: f0 increases for tense stops, but not as much as for aspirated stops (Figure 4). However, this account does not explain why the IF0 effect is attenuated less over time for lax stops, compared to aspirated stops. This pattern is puzzling because lax stops have lower f0 than aspirated stops, and IF0 effects are smaller in the lower part of a speaker’s pitch range cross-linguistically (Ladd & Silverman, 1984; Whalen & Levitt, 1995). One possibility is that the low f0 target for Korean lax stops is not in the lowest region of a speaker’s f0 range, while the f0 target for aspirated stops is in the higher region of a speaker’s f0 range. Increased activity of the cricothyroid muscle involved in the production of higher f0 (Hoole & Honda, 2011; Löfqvist et al., 1989) may therefore result in greater attenuation of IF0 for aspirated stops. Thus, IF0 attenuation that accompanies the enhancement of f0 differences or the emergence of contrastive f0 may be due to the ‘controlled’ mechanism suppressing the ‘automatic’ mechanism. The controlled mechanism may play a more important role in the stop category that is both prone to merger and occurs in the highest f0 range (i.e. aspirated stops), compared to other categories.

In sum, we find that as f0 becomes contrastive in Seoul Korean, the size of the IF0 effect is attenuated, especially for the stop category most affected by the sound change (aspirated stops), suggesting that speakers suppress non-contrastive variation in f0 (due to vowel height) as a *consequence* of its rise as a primary cue.

### 5.4 Actuation: Korean intonational phonology

We have argued that frequently used words and stops before nonhigh vowels lead the change in both VOT contrast reduction and f0 contrast enhancement, suggesting that transphonologization in Seoul Korean was triggered by lenition affecting the VOT contrast, which led to the phonologization of the f0 contrast. However, this does not explain why transphonologization has happened in Korean but not in other languages, even though reduction of long-lag stops in certain speech conditions presumably exists as a precondition in every

language with long-lag stops. The role of f0 in prosodic marking in Seoul Korean may help resolve this issue, as follows.

Phonological control of f0 in Seoul Korean related to adjacent consonants is in fact not new to the language, but is a long-standing part of the intonational phonology (Jun, 1993, 1996, 1998). Korean is unusual in that the way prosodic domains (APs) are marked is influenced by the identity of segments at domain edges. In Seoul Korean, tense and aspirated stops and affricates (as well as /s/, tense /s\*/, /h/) condition a high (H) boundary tone, while other consonants condition a low (L) boundary tone. This segment-induced f0 distinction is argued to be phonologically-controlled ‘phrase initial strengthening’ which is functionally motivated by perceptual enhancement (Jun, 1993, 1996, 2005). The prosodic tone-bearing segments pattern exactly together with the segments that are undergoing f0 change in Seoul Korean diachronically: the distance in f0 between the H-tone bearing segments and the L-tone bearing segments increases over time in parallel with VOT contrast reduction.

We suggest that the language-specific implementation of f0 for domain strengthening, which makes the f0 difference between categories larger than expected from purely physiological factors, may mean Seoul Korean listeners are more attuned to f0 than would be the case in other languages—especially in contexts where the VOT contrast is weaker (e.g. high-frequency words). In other words, we conjecture that the process of quasi-tonogenetic change we describe—hypoarticulation-driven reduction of VOT contrast and adaptive f0 contrast enhancement—may have begun when the domain-initial f0 distinction was already in place. When the VOT merger began, the prosodic f0 distinction was readily available to listeners, which facilitated adaptive enhancement. This language-specificity may help explain the broader long-standing question of why tonogenetic sound change is not more common cross-linguistically, given that segment-induced f0 perturbations are present in most languages.<sup>10</sup> Our account is also consistent with Kirby (2013), who argues that transphonologization to f0 *as opposed to another outcome* occurred in Seoul Korean because of the high ‘informativity’ of the f0 contrast which existed at the outset of the change.

The current study has a weakness which could be addressed in future cross-linguistic work. We have taken parallelism between VOT contrast reduction and f0 contrast enhancement as evidence for language adaptivity during a diachronic change. However, we are not certain whether this parallelism is expected cross-linguistically, or is unique to languages undergoing (quasi-)tonogenetic sound change. That is, is parallelism between VOT and f0 a *cause* of transphonologization (a “phonetic precursor”), or a *consequence* of the sound change in progress? While the relationship between VOT and f0 in voicing contrasts cross-linguistically is well-studied (e.g. Kirby & Ladd, 2015; McCrea & Morris, 2005), we are aware of only three studies addressing the relationship between VOT and f0 contrast strength (i.e. cue weights) across speakers, all in English, which reach conflicting results: Shultz, Francis & Llanos (2012) and Clayards (2008) find that talkers who contrast voicing categories with a larger VOT cue weight produce the contrast with a smaller f0 weight (that is, the same direction observed in Korean), while Clayards (2018) finds the opposite pattern. We are not aware of any studies addressing VOT and f0 contrast tradeoff across words (e.g. frequency) or contexts (e.g. vowel height). Future work could examine whether the trading pattern seen in Seoul Korean is found in other languages, and shed light on the more general issues of what the phonetic precursors to tonogenesis are, and the relationship of tonogenesis to synchronic variability in how laryngeal contrasts are implemented.

As a final remark, we note that the contrastive use of f0 in Seoul Korean is still constrained

---

<sup>10</sup>This is a special case of the more general ‘actuation problem’ (why any sound change is not more common; Sóskuthy, 2015; Weinreich et al., 1968).

by phrase-level intonational phonology, which makes the Korean sound change a sub-optimal case study to address general issues of tonogenesis. We believe the quasi-tonogenetic sound change in Seoul Korean shares enough similarity with ‘true’ cases of tonogenesis—including the rise of contrastive  $f_0$  by a combination of prosodic and segmental sources (Kingston, 2011)—for our findings to offer insight into tonogenetic sound changes more generally. However, at this point Seoul Korean is clearly not a tonal language. For Seoul Korean to develop into a true tonal language, where its lexical items are specified and distinguished by a paradigmatic set of more than one contrastive pitch, the use of contrastive  $f_0$  would need to descend from phrase-initial position to lower prosodic levels, for example through the process of ‘domain-narrowing’ (Bermúdez-Otero, 2015). Only time will tell whether Seoul Korean will follow this pathway to develop lexical tone.

## 6 Conclusion

We examined the origin, propagation, and impact of a quasi-tonogenetic sound change in Seoul Korean, and related our findings to these aspects of tonogenetic sound changes more broadly. We found that VOT contrast reduction and  $f_0$  contrast enhancement spread across phonetic contexts (vowels of different heights), and possibly words (of different frequencies), in parallel. These findings suggest that the sound change is propagating across speakers and the language in an adaptive manner, driven by a combination of production bias leading to contrast reduction in one dimension (VOT), and adaptive expansion of contrast in another dimension ( $f_0$ ), plausibly to avoid merger. We also found evidence that the vowel intrinsic pitch difference is attenuated as contrastive  $f_0$  emerges, possibly due to the combined effect of controlled and automatic mechanisms. These findings shed light on how the sound system of a language dynamically changes in an incremental and adaptive manner via continuous adjustments in speech production and perception.

## 7 Acknowledgements

A preliminary version of this paper was published as Bang, Sonderegger, Kang, Clayards & Yoon (2015). We thank audiences at the 2014 Montreal-Ottawa-Toronto Phonology workshop, ICPHS 2015, and ICKL & Harvard-ISOKL 2015, as well as James Kirby for useful feedback. We especially thank three anonymous reviewers and the editor (Taehong Cho) for their helpful comments. This work was supported by SSHRC grant 430-2014-00018 and FRQSC grant 183356 to MS, SSHRC grant 435-2014-1504 to MC, and SSHRC Partnership Development Grant 890-2012-25 to YK.

## References

- Abramson, A. S., L-Thongkum, T., & Nye, P. (2005). Voice register in Suai (Kuai): An analysis of perceptual and acoustic data. *Phonetica*, *61*(2-3), 147–171.
- Atkinson, J. E. (1972). Correlation analysis of the physiological features controlling fundamental voice frequency. *Journal of the Acoustical Society of America*, *63*, 211–222.
- Aylett, M. & Turk, A. (2004). The Smooth Signal Redundancy Hypothesis: A Functional Explanation for Relationships between Redundancy, Prosodic Prominence, and Duration in Spontaneous Speech. *Language and Speech*, *47*(1), 31–56.
- Aylett, M. & Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *Journal of the Acoustical Society of America*, *119*(5), 3048–3058.
- Baayen, R. (2008). *Analyzing linguistic data*. Cambridge: Cambridge University Press.
- Bailey, G., Wikle, T., & Tillery, J. (1993). Some patterns of linguistic diffusion. *Language Variation and Change*, *5*(3), 359–390.
- Baker, A., Archangeli, D., & Mielke, J. (2011). Variability in American English s-retraction suggests a solution to the actuation problem. *Language Variation and Change*, *23*(03), 347–374.
- Baker, R. E. & Bradlow, A. R. (2009). Variability in Word Duration as a Function of Probability, Speech Style, and Prosody. *Language and Speech*, *52*(4), 391–413.
- Bang, H.-Y., Sonderegger, M., & Clayards, M. (2017). Speaker Variability in Cue Weighting for Laryngeal Contrasts: the Relationship to Sound Change. Edinburgh, Scotland: University of Edinburgh.
- Bang, H. Y., Sonderegger, M., Kang, Y., Clayards, M., & Yoon, T.-J. (2015). The Effect of Word Frequency on the Timecourse of Tonogenesis in Seoul Korean. In *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, Scotland, UK.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. arXiv preprint arXiv:1506.04967.
- Beckman, M. E., Li, F., Kong, E. J., & Edwards, J. (2014). Aligning the timelines of phonological acquisition and change. *Laboratory Phonology*, *5*(1), 151–194.
- Beckman, M. E. & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology*, *3*(1), 255–309.
- Beddor, P. (2015). The Relation between Language Users' Perception and Production Repertoires. In for ICPHS 2015, T. S. C. (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*, (pp. Paper 1041.1–9)., Glasgow, UK. The University of Glasgow.
- Beddor, P. S. (2009). A Coarticulatory Path to Sound Change. *Language*, *85*(4), 785–821.
- Beddor, P. S., McGowan, K. B., Boland, J. E., Coetzee, A. W., & Brasher, A. (2013). The time course of perception of coarticulation. *Journal of the Acoustical Society of America*, *133*, 2350–2366.
- Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, *60*(1), 92–111.
- Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America*, *113*(2), 1001–1024.
- Bell-Berti, F. (1975). Control of pharyngeal cavity size for English voiced and voiceless stops. *Journal of the Acoustical Society of America*, *57*, 456–461.
- Bermúdez-Otero, R. (2015). Amphichronic explanation and the life cycle of phonological processes. In *The Oxford handbook of historical phonology* (pp. 374–399). Oxford University Press.
- Berry, J. & Moyle, M. (2011). Covariation among vowel height effects on acoustic measures. *Journal of the Acoustical Society of America*, *130*(5), 365–371.

- Bush, N. (1999). The predictive value of transitional probability for word-boundary palatalization in English. Masters' thesis, University of New Mexico.
- Bybee, J. L. (1985). *Morphology: A Study of the Relation Between Meaning and Form*. Philadelphia, PA: John Benjamins Publishing.
- Bybee, J. L. (2000). The phonology of the lexicon: Evidence from lexical diffusion. In M. Barlow & S. Kemmer (Eds.), *Usage-based models of language* (pp. 65–85). Stanford, CA: CSLI.
- Bybee, J. L. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, 14, 261–290.
- Bybee, J. L. (2012). Patterns of lexical diffusion and articulatory motivation for sound change. In M.-J. Solé & D. Recasens (Eds.), *The Initiation of Sound Change, Perception, Production, and Social Factors* (pp. 210–234). Amsterdam, the Netherlands: Benjamins.
- Bybee, J. L. & Hopper, P. (2001). *Frequency and the Emergence of Linguistic Structure*. John Benjamins Publishing.
- Chen, M. (1972). The time dimension: contribution toward a theory of sound change. *Foundations of language*, 8(4), 457–498.
- Chen, Y. (2011). How does phonology guide phonetics in segment–f0 interaction? *Journal of Phonetics*, 39(4), 612–625.
- Cho, T., Jun, S. A., & Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics*, 30(2), 193–228.
- Cho, T. & Keating, P. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics*, 29, 155–190.
- Cho, T. & Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics*, 27(2), 207–229.
- Clayards, M. (2008). *The ideal listener: Making optimal use of acoustic-phonetic cues for word recognition*. PhD thesis, University of Rochester.
- Clayards, M. (2018). Individual talker and token covariation in the production of multiple cues to stop voicing. *Phonetica*, 75(1), 1–23.
- Coetzee, A. W., Bendor, P. S., & Wissing, D. P. (2014). Emergent tonogenesis in Afrikaans. *Journal of the Acoustical Society of America*, 135, 2421.
- Connell, B. (2002). Tone languages and the universality of intrinsic F0: evidence from Africa. *Journal of Phonetics*, 30(1), 101–129.
- DiCiano, C. T. (2012). Coarticulation between tone and glottal consonants in Itunyoso Trique. *Journal of Phonetics*, 40(1), 162–176.
- Douglas, B., Martin, M., Ben, B., & Steve, W. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Eckert, P. (1989). The whole woman: Sex and gender differences in variation. *Language Variation and Change*, 1(3), 245–267.
- Esposito, A. (2002). On vowel height and consonantal voicing effects: Data from Italian. *Phonetica*, 59, 197–231.
- Fant, G., Kruckenberg, A., & Gustafson, K. (2002). A new approach to intonation analysis and synthesis of Swedish. In *Proceedings of Fonetik, TMH-QPSR*, (pp. 161–164)., Aix en Provence.
- Fischer-Jørgensen, E. (1990). Intrinsic F0 in tense and lax vowels with special reference to German. *Phonetica*, 47(3-4), 99–140.
- Fruehwald, J. (2016). The early influence of phonology on a phonetic change. *Language*, 92, 376–410.
- Fruehwald, J., Gress-Wright, J., & Wallenberg, J. C. (2013). Phonological rule change: The constant rate effect. In *Proceedings of the 40th Annual Meeting of the North East Linguistic Society*. Massachusetts: GLSA Publications.
- Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, 84, 474–496.
- Gelman, A. & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

- Gelman, A. & Su, Y.-S. (2015). *arm: Data Analysis Using Regression and Multilevel/Hierarchical Models*. R package version 1.8-6.
- Hagège, C. & Haudricourt, A.-G. (1978). *La phonologie panchronique*. Paris: Presses Universitaires de France.
- Han, M. S. & Weitzman, R. S. . (1967). Studies in the phonology of Asian languages V: Acoustic features in the manner-differentiation of Korean stop consonants. In *Studies in the phonology of Asian languages V*. Los Angeles: Acoustic Phonetics Research Laboratory, University of Southern California.
- Han, M. S. & Weitzman, R. S. (1965). Studies in the phonology of Asian Languages III: Acoustic characteristics of Korean stop consonants. In *Studies in the phonology of Asian languages III*. Los Angeles: Acoustic Phonetics Research Laboratory, University of Southern California.
- Hardcastle, W. J. (1973). Some observations on the tense-lax distinction in initial stops in Korean. *Journal of Phonetics*, 1, 263–272.
- Harrell, F. (2001). *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*. New York: Springer Verlag.
- Harrell, J. (2015). Hmisc: Harrell miscellaneous. R package version 3.17-1.
- Harrell, J. & Frank, E. (2015). rms: Regression modeling strategies. R package version 4.4-1.
- Harrington, J. (2012). The relationship between synchronic variation and diachronic change . In A. C. Cohn, C. Fougeron, & M. K. Huffman (Eds.), *Handbook of Laboratory Phonology* (pp. 321–332). Oxford University Press.
- Harrington, J., Kleber, F., Reubold, U., & Siddins, J. (2015). The relationship between prosodic weakening and sound change: evidence from the German tense/lax contrast. *Laboratory Phonology*, 6(1), 87–117.
- Harrington, J., Kleber, F., & Stevens, M. (2016). The Relationship Between the (Mis)-Parsing of Coarticulation in Perception and Sound Change: Evidence from Dissimilation and Language Acquisition. In A. Esposito, M. Faundez-Zanuy, A. Esposito, G. Cordasco, T. Drugman, J. Sol-Casals, & F. Morabito (Eds.), *In Recent Advances in Nonlinear Speech Processing* (pp. 15–34). Switzerland: Springer International Publishing.
- Hay, J. & Foulkes, P. (2016). The evolution of medial /t/ over real and remembered time. *Language*. in press.
- Hay, J. B., Pierrehumbert, J. B., Walker, A. J., & LaShell, P. (2015). Tracking word frequency effects through 130 years of sound change. *Cognition*, 139, 83–91.
- Higgins, M. B., Netsell, R., & Schulte, L. (1998). Vowel-related differences in laryngeal articulatory and phonatory function. *Journal of Speech and Hearing Research*, 41(4), 712–724.
- Hirose, H. & Gay, T. J. (1972). The activity of the intrinsic laryngeal muscles in voicing control: an electromyographic study. *Phonetica*, 25, 140–164.
- Hockett, C. F. A. (1958). *A course in modern linguistics*. New York: Mcmillan.
- Hombert, J.-M. (1974). Universals of downdrift: their phonetic basis and significance for a theory of tone. *Studies in African Linguistics*, 5, 169–183.
- Hombert, J.-M. (1977). Consonant types, vowel height and tone in Yoruba. *Studies in African Linguistics*, 8(2), 1–18.
- Hombert, J.-M., Ohala, J. J., & Ewan, W. (1979). Phonetic explanations for the development of tones. *Language*, 55(1), 37–58.
- Honda, K. (1983). Relation between pitch contral and vowel articulation. *Haskins Laboratories Status Report on Speech Research*, SR 73, 269–282.
- Honda, K., Hirai, H., & Dang, J. (1994). A physiological model of speech production and the implication of tongue-larynx interaction. In *Proceedings of the 1994 International Conference on Spoken Language Processing (ICSLP 94)*, (pp. 157–178)., Yokohama, Japan.
- Hoole, P. & Honda, K. (2011). Automaticity vs. feature-enhancement in the control of segmental F0. In G. Clements & N. Ridouane (Eds.), *Where Do Phonological Features Come From?: Cognitive, Physical And Developmental Bases Of Distinctive Speech Categories* (pp. 131–171). Amsterdam: John Benjamins Publishing Company.

- Hoole, P., Honda, K., Murano, E., Fuchs, S., & Pape, D. (2006). Go with the flow: Between automaticity and enhancement in control of segmental F0. In *Proceedings of the 7th International Seminar on Speech Production*.
- Hooper, J. B. (1976). Word frequency in lexical diffusion and the source of morphophonological change. In W. Christie (Ed.), *Current progress in historical linguistics* (pp. 96–105). Amsterdam: North-Holland.
- Hyman, L. M. (1976). Phonologization. In A. Juilland (Ed.), *Linguistic Studies Offered to Joseph Greenberg*, volume 2. Saratoga, CA: Anma Libri.
- Hyman, L. M. (2008). Universals in phonology. *The Linguistic Review*, 25, 83–137.
- Hyslop, G. (2009). Kurtöp tone: a tonogenetic case study. *Lingua*, 119(6), 827–845.
- Jacewicz, E., Fox, R. A., O'Neill, C., & Salmons, J. (2009). Articulation rate across dialect, age, and gender. *Language Variation and Change*, 21(2), 233–256.
- Jun, S. A. (1993). *The Phonetics and Phonology of Korean prosody*. PhD thesis, the Ohio State University.
- Jun, S. A. (1996). Intonational Phonology of Seoul Korean revisited. In *the 14th Japanese/Korean Linguistics conference*, (pp. 14–25)., Tucson Arizona. UCLA Working Papers in Phonetics.
- Jun, S. A. (1998). The Accentual phrase in the Korean prosodic hierarchy. *Phonology*, 15, 189–226.
- Jun, S. A. (2005). Intonational phonology of Seoul Korean revisited. *UCLA Working Papers in Phonetics*, 104, 14–25.
- Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee & P. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure* (pp. 229–254). Amsterdam, The Netherlands: John Benjamins Publishing Company.
- KAIST (1999). KAIST Concordance Program. <http://semanticweb.kaist.ac.kr/research/kcp/>.
- Kang, K.-H. & Guion, S. G. (2008). Clear speech production of Korean stops: changing phonetic targets and enhancement strategies. *Journal of the Acoustical Society of America*, 124(6), 3909–3917.
- 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.
- Kang, Y. & Han, S. (2013). Tonogenesis in early Contemporary Seoul Korean: A longitudinal case study. *Lingua*, 134, 62–74.
- Kang, Y. & Nagy, N. (2016). VOT Merger in Heritage Korean in Toronto. *Language Variation and Change*, 28(2), 249–272.
- Kang, Y., Yoon, T.-J., & Han, S. (2015). Frequency effects on the vowel length contrast merger in Seoul Korean. *Laboratory Phonology*, 6(3-4), 469–503.
- Keating, P., Cho, T., Fougeron, C., & Hsu, C. (2003). Domain-initial articulatory strengthening in four languages. In J. Local, R. Ogden, & R. Temple (Eds.), *Papers in Laboratory Phonology VI: Phonetic interpretations* (pp. 143–161). Cambridge, UK: Cambridge University Press.
- 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/>.
- Kessinger, R. H. & Blumstein, S. E. (1997). Effects of speaking rate on voice-onset time in Thai, French, and English. *Journal of Phonetics*, 25(2), 143–168.
- Kim, C.-W. (1965). On the autonomy of the tensivity feature in stop classification. *Word*, 21, 339–359.
- Kim, J. (2016). Perceptual associations between words and speaker age. *Journal of the Association for Laboratory Phonology*, 7(1), 1–22.
- Kim, M.-R., Beddor, P. S., & Horrocks, J. (2002). The contribution of consonantal and vocalic information to the perception of Korean initial stops. *Journal of Phonetics*, 30(1), 77–100.
- Kingston, J. (1992). The phonetics and phonology of perceptually motivated articulatory covariation. *Language and Speech*, 35, 99–113.

- Kingston, J. (2005). The phonetics of Athabaskan tonogenesis. In S. Hargus & K. Rice (Eds.), *Athabaskan Prosody* (pp. 137–184). Amsterdam: Benjamins.
- Kingston, J. (2011). Tonogenesis. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology* (pp. 2304–2333). Oxford: Oxford University Press.
- Kirby, J. & Ladd, D. R. (2015). Stop voicing and f0 perturbations: evidence from French and Italian. In *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow.
- Kirby, J. P. (2013). The role of probabilistic enhancement in phonologization. In A. C. L. Yu (Ed.), *Origin of Sound Change: Approaches to Phonologization*. OUP: Oxford.
- Kirby, J. P. (2014). Incipient tonogenesis in Phnom Penh Khmer: Acoustic and perceptual studies. *Journal of Phonetics*, 43, 69–85.
- Klatt, D. H. (1975). Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research*, 18(4), 686–706.
- Kong, E. J., Beckman, M. E., & Edwards, J. (2011). Why are Korean tense stops acquired so early?: The role of acoustic properties. *Journal of Phonetics*, 39(2), 196–211.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). lmerTest: Tests in linear mixed effects models. R package version 2.0-29.
- Labov, W. (1966). *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.
- Labov, W. (1990). The intersection of sex and social class in the course of linguistic change. *Language Variation and Change*, 2(02), 205–254.
- Labov, W. (1994). *Principles of Linguistic Change, Vol. 1: Internal Factors*. Malden, MA: Wiley-Blackwell.
- Labov, W. (2001). *Principles of Linguistic Change, Social Factors, Vol. 2, Social factors*. Oxford: Wiley-Blackwell.
- Labov, W. (2007). Transmission and diffusion. *Language*, 83, 344–387.
- Labov, W. (2010). *Principles of Linguistic change. Volume III: Cognitive & Cultural Factors*. Oxford: Wiley Blackwell.
- Ladd, D. R. & Silverman, K. E. A. (1984). Vowel intrinsic pitch in connected speech. *Phonetica*, 41, 31–40.
- 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(02), 145–169.
- 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.
- Lehiste, I. (1976). Suprasegmental features of speech. In N. J. Lass (Ed.), *Contemporary Issues in Experimental Phonetics* (pp. 225–239). New York: Academic Press.
- Lieberman, E., Michel, J.-B., Jackson, J., Tang, T., & Nowak, M. A. (2007). Quantifying the evolutionary dynamics of language. *Nature*, 449(7163), 713–716.
- Lindblom, B., Guion, S. G., Hura, S., Moon, S.-J., & Willerman, R. (1995). Is sound change adaptive? *Rivista di Linguistica*, 7, 5–36.
- Lisker, L. & Abramson, A. S. (1964). A Cross-language Study of Voicing in Initial Stops: acoustical measurements. *Word*, 20, 384–422.
- Löfqvist, A., Baer, T., McGarr, N. S., & Story, R. S. (1989). The cricothyroid muscle in voicing control. *Journal of the Acoustical Society of America*, 85(3), 1314–1321.
- Maran, L. R. (1973). On becoming a tone language: a Tibeto-Burman model of tonogenesis. *Consonant types and tone, southern California occasional papers in linguistics*, 1, 97–114.
- Matisoff, J. A. (1973). Tonogenesis in Southeast Asia. In L. M. Hyman (Ed.), *Consonant Types and Tones* (pp. 72–95). Southern California Occasional Papers in Linguistics.
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2015). Balancing type I error and power in linear mixed models. arXiv preprint arXiv:1511.01864.

- Mazaudon, M. & Michaud, A. (2009). Tonal contrasts and initial consonants: a case study of Tamang, a missing link in tonogenesis. *Phonetica*, 65(4), 231–256.
- McCrea, C. R. & Morris, R. J. (2005). The effects of fundamental frequency level on voice onset time in normal adult male speakers. *Journal of Speech and Hearing Research*, 48(5), 1013–1024.
- Miller, J. L., Green, K. P., & Reeves, A. (1986). Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast. *Phonetica*, 43, 106–115.
- Misnadin, M., Kirby, J., & Remijsen, B. (2015). Temporal and spectral properties of Madurese stops. In for ICPHS 2015, T. S. C. (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*, (pp. Paper 789.1–5), Glasgow, UK. The University of Glasgow.
- Morris, R. J., McCrea, C. R., & Herring, K. D. (2008). Voice onset time differences between adult males and females: Isolated syllables. *Journal of Phonetics*, 36(2), 308–317.
- Nolan, F. (2003). Intonational equivalence: an experimental evaluation of pitch scales. In Solé, M.-J., Recasens, D., & Romero, J. (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, Spain.
- Ogura, M. (2012). The Timing of Language Change. In J. M. Hernandez-Campoy & J. C. Conde-Silvestre (Eds.), *The handbook of historical sociolinguistics* (pp. 427–450). Oxford: Wiley-Blackwell.
- Oh, E. (2011). Effects of speaker gender on voice onset time in Korean stops. *Journal of Phonetics*, 39, 59–67.
- Ohala, J. (1981). The Listener as a Source of Sound Change. In C. Masek, R. A. Hendrick, & M. F. Miller (Eds.), *Papers from the Parasession on Language and Behavior Chicago Linguistic Society* (pp. 178–203). Chicago: Chicago Linguistics Society.
- Ohala, J. J. (1978). Production of Tone. In V. A. Fromkin (Ed.), *Tone: A linguistic survey* (pp. 1–18). New York: Academic Press.
- Ohala, J. J. (1993a). Sound change as nature's speech perception experiment. *Speech Communication*, 13(1), 155–161.
- Ohala, J. J. (1993b). The phonetics of sound change. In C. Jones (Ed.), *Historical linguistics: Problems and perspectives* (pp. 237–278). London: Longman.
- Ohala, J. J. (2000). The physiology of tone. In L. Hyman (Ed.), *Consonant types and tone* (pp. 3–14). Los Angeles: University of Southern California.
- Ohala, J. J. & Sprouse, R. (2003). Effects on speech of introducing aerodynamic perturbations. In *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona.
- Pan, S. & Hirschberg, J. (2000). Modeling local context for pitch accent prediction. In *Proceedings of the ACL 2000*, (pp. 233–240), Hong Kong.
- Park, H. (2002). The Time Courses of F1 and F2 as a Descriptor of Phonation Types. *Acta Otolaryngologica*, 33, 87–108.
- Phillips, B. S. (1984). Word frequency and the actuation of sound change. *Language*, 60(2), 320–342.
- Phillips, B. S. (2006). *Word frequency and lexical diffusion*. Basingstoke: Palgrave Macmillan.
- Pierrehumbert, J. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology* (pp. 101–139). Berlin: Mouton de Gruyter.
- Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. L. Bybee & P. Hopper (Eds.), *Frequency effects and the emergence of lexical structure* (pp. 137–157). Amsterdam: John Benjamins.
- Pind, J. (1995). Speaking rate, voice-onset time, and quantity: the search for higher-order invariants for two Icelandic speech cues. *Perception and Psychophysics*, 57(3), 291–304.
- Purcell, E., Villegas, G., & Young, S. (1978). A before and after for tonogenesis. *Phonetica*, 35(5), 284–293.
- Reubold, U. & Harrington, J. (2015). Disassociating the effects of age from phonetic change. In A. Gerstenberg & A. Voeste (Eds.), *Language Development: The life span perspective* (pp. 9–37). Amsterdam/Philadelphia: John Benjamins Publishing Company.

- Roubeau, B., Chevrier-Muller, C., & Saint Guily, J. L. (1972). Electromyographic activity of strap and cricothyroid muscles in pitch change. *Acta Otolaryngologica*, 117, 459–464.
- Sankoff, G. (2004). Adolescents, young adults and the critical period: Two case studies from 'Seven Up'. In C. Fought (Ed.), *Sociolinguistic variation: Critical reflections* (pp. 121–139). Oxford: Oxford University Press.
- Shultz, A. A., Francis, A. L., & Llanos, F. (2012). Differential cue weighting in perception and production of consonant voicing. *Journal of the Acoustical Society of America*, 132(2), EL95–E101.
- Siddins, J. & Harrington, J. (2015). Does vowel intrinsic F0 affect lexical tone? In *Proceedings of the 18th International Congress of Phonetic Sciences*, (pp. 27–43)., Glasgow, Scotland.
- Silva, D. J. (2002). Consonant aspiration in Korean: a retrospective. In S.-O. Lee & G. K. Iverson (Eds.), *Pathways into Korean language and culture: essays in honor of Young-Key Kim-Renaud* (pp. 447–469). Seoul: Pagijong Press.
- Silva, D. J. (2006). Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology*, 23(02), 287–308.
- Solé, M.-J. (2007). Controlled and mechanical properties in speech: A review of literature. In M.-J. Solé, P. Beddor, & M. Ohala (Eds.), *Experimental Approaches to Phonology* (pp. 302–321). Oxford: Oxford University Press.
- Soltani, M., Ashayeri, H., Modarresi, Y., Salavati, M., & Ghomashchi, H. (2014). Fundamental frequency changes of persian speakers across the life span. *Journal of Voice*, 28(3), 274–281.
- Sóskuthy, M. (2015). Understanding change through stability: A computational study of sound change actuation. *Lingua*, 163, 40–60.
- Stebbins, J. (2010). *Usage frequency and articulatory reduction in Vietnamese tonogenesis*. PhD thesis, University of Colorado, Boulder.
- Stevens, K. (1998). *Acoustic Phonetics*. MA: MIT Press.
- Stevens, M. & Harrington, J. (2014). The individual and the actuation of sound change. *Loquens*, 1, e003.
- Stuart-Smith, J., Sonderegger, M., & Rathcke, T. (2015). The private life of stops: VOT in a real-time corpus of spontaneous Glaswegian. *Laboratory Phonology*, 6(3-4), 505–549.
- Svantesson, J.-O. & House, D. (2006). Tone production, tone perception and Kammu tonogenesis. *Phonology*, 23(02), 309–333.
- The National Institute of the Korean Language (2005). *A speech corpus of reading-style standard Korean [DVDs]*. Seoul: NIKL.
- Titze, I. R. (1989). Physiologic and acoustic differences between male and female voices. *Journal of the Acoustical Society of America*, 85, 1699–1707.
- Torre, III, P. & Barlow, J. A. (2009). Age-related changes in acoustic characteristics of adult speech. *Journal of Communication Disorders*, 42(5), 324–333.
- Torreira, F., Bögels, S., & Levinson, S. C. (2015). Intonational phrasing is necessary for turn-taking in spoken interaction. *Journal of Phonetics*, 52, 46–57.
- Van Hoof, S. & Verhoeven, J. (2011). Intrinsic vowel F0, the size of vowel inventories and second language acquisition. *Journal of Phonetics*, 39(2), 168–177.
- Wagner, S. E. (2012). Age grading in sociolinguistic theory. *Language and Linguistics Compass*, 6(6), 371–382.
- Walker, A. & Hay, J. (2011). Congruence between 'word age and 'voice age facilitates lexical access. *Laboratory Phonology*, 2(1), 219–237.
- Wang, W. S. Y. (1969). Competing changes as a cause of residue. *Language*, 45, 9–25.
- Weinreich, U., Labov, W., & Herzog, M. I. (1968). Empirical Foundations for a Theory of Language Change. In W. Lehmann & Y. Malkiel (Eds.), *Directions for Historical Linguistics* (pp. 95–195). Austin: University of Texas Press.
- Whalen, D. H. & Levitt, A. G. (1995). The universality of intrinsic F0 of vowels. *Journal of Phonetics*, 23(3), 349–366.

- Whalen, D. H., Levitt, A. G., Hsiao, P.-L., & Smorodinsky, I. (1995). Intrinsic F0 of vowels in the babbling of 6-, 9-, and 12-month-old French- and English-learning infants. *Journal of the Acoustical Society of America*, 97(4), 2533–2539.
- Wright, J. D. (2007). *Laryngeal contrast in Seoul Korean*. PhD thesis, University of Pennsylvania.
- Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.
- Yoon, J. & Choi, K.-S. (1999). *Study on KAIST corpus*. CS-TR-99-139, KAIST CS.
- Yoon, T.-J. (2014). A corpus-based study on the layered duration in Standard Korean. In M. Kenstowicz, T. Levin, & R. Masuda (Eds.), *Japanese/Korean Linguistics 23*. Chicago: University of Chicago Press.
- Yoon, T.-J. & Kang, Y. (2014). Monophthong Analysis on a Large-scale Speech Corpus of Read-Style Korean. *Phonetics and Speech Sciences*, 6, 139–145.
- Zellou, G. & Tamminga, M. (2014). Nasal coarticulation changes over time in Philadelphia English. *Journal of Phonetics*, 47, 18–35.
- Zhao, Y. & Jurafsky, D. (2007). The effect of lexical frequency on tone production. In *Proceedings of the 16th International Congress of Phonetic Sciences*, (pp. 477–480), Saarbrücken.
- Zhao, Y. & Jurafsky, D. (2009). The effect of lexical frequency and Lombard reflex on tone hyperarticulation. *Journal of Phonetics*, 37(2), 231–247.

# Appendices

**Table A1:** Summary of fixed-effect coefficients in the static model of  $F_0$  on the subsetted data (speaker year of birth < 1965)

F0 MODEL	F <sub>0</sub>				
	Estimate	SE	df	<i>t</i>	<i>P</i> (> <i>t</i> )
Intercept	1.382	0.291	63.866	4.745	< 0.001
YOB'	0.064	0.536	55.972	0.119	0.906
LARYNGEAL1(tense vs. nontense)	-1.047	0.265	111.444	-3.957	< 0.001
LARYNGEAL2(lax vs. aspirated)	5.245	0.371	88.568	14.133	< 0.001
HEIGHT(h)	0.497	0.263	119.659	1.892	0.061
FREQUENCY	-0.212	0.187	95.897	-1.133	0.26
POSITION1(initial vs. medial)	-0.174	0.13	117.535	-1.342	0.182
POSITION2(short vs. longer pause)	0.134	0.084	57.126	1.597	0.116
POSITION3(medial vs. long pause)	0.114	0.095	73.652	1.202	0.233
RATE DEVIATION	0.275	0.106	74.443	2.585	0.012
GENDER(m)	-1.303	0.553	55.96	-2.357	0.022
PLACE1(labial vs. non-labial)	-0.089	0.16	65.993	-0.555	0.581
PLACE2(alveolar vs. velar)	0.348	0.208	67.565	1.672	0.099
SPEAKER MEAN RATE	-0.112	0.302	56.336	-0.369	0.713
YOB':LARYNGEAL1	-0.624	0.371	59.385	-1.684	0.097
YOB':LARYNGEAL2	3.213	0.645	59.476	4.981	< 0.001
YOB':HEIGHT	-1.087	0.369	74.015	-2.948	0.004
YOB':FREQ.	0.181	0.223	59.094	0.813	0.419
LARYNGEAL1:HEIGHT	0.066	0.59	99.328	0.112	0.911
LARYNGEAL2:HEIGHT	-1.517	0.424	111.722	-3.578	0.001
LARYNGEAL1:FREQ.	-0.133	0.42	88.886	-0.317	0.752
LARYNGEAL2:FREQ.	1.271	0.414	110.056	3.07	0.003
LARYNGEAL1:POSITION1	0.525	0.258	196.741	2.038	0.043
LARYNGEAL2:POSITION1	0.694	0.314	69.484	2.209	0.03
LARYNGEAL1:POSITION2	0.021	0.174	2728.918	0.118	0.906
LARYNGEAL2:POSITION2	-0.032	0.177	74.523	-0.179	0.858
LARYNGEAL1:POSITION3	-0.043	0.199	73.416	-0.215	0.83
LARYNGEAL2:POSITION3	0.129	0.182	2260.517	0.708	0.479
LARYNGEAL1:RATE DEV.	-0.34	0.232	66.041	-1.466	0.147
LARYNGEAL2:RATE DEV.	-0.137	0.178	43.178	-0.767	0.447
LARYNGEAL1:GENDER	-0.055	0.226	66.732	-0.241	0.81
LARYNGEAL2:GENDER	-0.768	0.354	67.457	-2.173	0.033
YOB':GENDER	0.927	1.027	54.976	0.903	0.371
YOB':LARYNGEAL1:HEIGHT	0.906	0.746	59.331	1.215	0.229
YOB':LARYNGEAL2:HEIGHT	-1.561	0.583	57.792	-2.679	0.01
YOB':LARYNGEAL1:FREQ.	-1.398	0.493	49.953	-2.838	0.007
YOB':LARYNGEAL2:FREQ.	1.186	0.498	99.121	2.38	0.019