# The structure of multiple cues to stop categorization and its implications for sound change

Hye-Young Bang

Department of Linguistics

McGill University, Canada

October 2017

*A thesis submitted to McGill University in partial fulfillment for the requirements of the degree of Doctor of Philosophy*

# Abstract

The central goal of this dissertation is to understand how multiple acoustic cues that signal phonetic contrasts are structured across linguistic and socio-linguistic factors, and what elements in these structures contribute to a long-term change in pronunciation norms in a speech community. The acoustic cues this dissertation focuses on are VOT and f0 which covary as cues to stop voicing categorization. The covariation of these cues is of particular interest because it is related to a cross-linguistically common type of sound change called 'transphonologization', where the relative importance of the cues that signal a set of phonetic categories changes over time.

There are many questions that remain unresolved regarding acoustic cue covariation in speech production, including: (1) what is the mechanism and path of transphonologization; (2) which aspects of the structure of cue covariation are language-independent, and which uniquely appear during sound change? This dissertation addresses these two questions in two studies, using large datasets from spoken corpora of three languages.

Study 1 examines how transphonologization from VOT to f0 originates, how it propagates, and what consequences the change has for other aspects of the sound system of the language undergoing change. This question was examined using a dataset from a large apparent-time corpus of Seoul Korean (the NIKL corpus). The time-course of change in VOT and f0 was examined focusing on linguistic (word frequency and vowel height) and social (speakers' age and gender) factors. The results showed a consistent trade-off between cues across talkers and contexts, and that the shift in cue primacy is more advanced in younger females' speech and in the linguistic conditions where the VOT contrast is more reduced (high frequency words and before non-high vowels). These findings suggest that VOT reduction and f0 enhancement are spreading in parallel across speakers and the language in an adaptive manner, driven by a combination of production bias in one dimension (VOT) and adaptive expansion in another (f0).

Study 2, building on the results of Study 1, sought to better describe how linguistic and social variables structure cue covariation across languages. It further aimed to identify aspects of cue covariation that are unique to languages undergoing change, by comparing Seoul Korean with two languages that are not undergoing transphonologization, English and German. The analysis examined corpus data from the three languages, focusing on two linguistic factors (word frequency and vowel height) one social factor (gender), and variation across individual speakers, and their effects on the relative use of VOT and f0 in signalling stop voicing categories. Regression models captured the effects of the linguistic and social factors, and cue use (or 'weights') for each individual was quantified using several methods: Cohen's $D$, linear discriminant analyses, and support vector machines. For German and English, in high-frequency words and before non-high vowels, one cue is

consistently used less than the other cue, reducing total cue informativity for stop voicing contrasts. Interestingly, these are the same conditions where transphonologization in Seoul Korean is more advanced. The results for gender and individual speakers, on the other hand, showed consistent cue trade-offs in all languages.

Taken together, the results from these two studies show that the cue trade-offs across linguistic conditions, observed only during the change in Seoul Korean, may have initiated to compensate for reduction in 'cue informativity' to avoid contrast merger, a pattern which is likely unique to transphonologization. On the other hand, the consistent cue trade-offs across speakers and gender in all three languages suggest that speakers have a goal of constant total cue informativity, while maintaining socially conditioned stylistic variation. The resulting structured variability across speakers may further serve as a catalyst to sound change, and change may be propagating by strengthening these existing structures. Broadly speaking, the findings provide evidence that variability in speech is not random but structured in multiple dimensions across contexts and talkers. This structure potentially reduces the variability that listeners must cope with in real-time processing and helps us understand how sound change occurs over generations.

# Résumé

Cette thèse a comme objectif principal d'éclaircir (a) comment sont structurés les multiples indices acoustiques qui signalent les contrastes phonétiques, une structure qui reflète des facteurs linguistiques et sociolinguistiques, et (b) quels éléments de ces structures contribuent aux changements à long terme des normes de prononciation dans une communauté linguistique. Les indices acoustiques sur lesquels se focalise cette thèse sont le délai d'établissement du voisement (DEV) et la fréquence fondamentale (f0), qui sont des indices covariantes permettant de classer les plosives. La covariation de ces indices est d'un intérêt particulier car elle est liée à un type de changement de son commun nommé la 'transphonologisation', où l'importance relative des indices acoustiques change au cours du temps.

Il y a beaucoup de questions non résolues en ce qui concerne la covariation des indices acoustiques dans la production de la parole, y compris: (1) quels sont le mécanisme et le chemin de transphonologisation; et (2) quels aspects structurels de cette covariation des indices ne sont pas uniques à une langue donnée et quels aspects structurels ne sont présents que lors des changements de son? Cette thèse aborde ces deux questions dans le cadre de deux études à grande échelle, pour lesquelles les données ont été extraites de corpus parlés de trois langues.

La première étude examine comment a lieu la transphonologisation du DEV à la f0, comment ce phénomène se propage et quelles sont les conséquences du changement sur d'autres aspects du système phonologique de la langue. Cette question a été abordée grâce à des données du changement en temps apparent provenant d'un corpus de coréen de Séoul (le corpus NIKL). La chronologie du changement de DEV et de f0 a été examinée en se concentrant sur les facteurs linguistiques (la fréquence des mots et la hauteur des voyelles) et sociaux (l'âge et le sexe des locuteurs). Les résultats ont démontré d'abord qu'il existe un compromis entre les indices acoustique selon le locuteur et le contexte et ensuite que le changement de primauté des indices est plus avancé dans le parler des jeunes femmes et dans les conditions linguistiques où le contraste du DEV est réduit (dans les mots à haute fréquence et devant les voyelles non hautes). Ces résultats suggèrent que la réduction du DEV et le renforcement de la f0 se propagent en parallèle de manière adaptative entre les locuteurs et la langue, le tout motivé par un biais de production dans une dimension (le DEV) et l'expansion adaptative dans une autre (la f0).

La deuxième étude, s'appuyant sur les résultats de la première, a cherché à mieux décrire comment les facteurs sociaux et linguistiques sont impliqués dans la structure de la covariation des indices dans les langues. En outre, visant à identifier les aspects de la covariation des indices qui sont propres aux langues en changement, la deuxième étude offre une comparaison entre le coréen de Séoul et deux langues qui ne sont pas

en cours de transphonologisation, soit l'anglais et l'allemand. L'analyse a examiné des données de corpus pour les trois langues, se focalisant principalement sur deux facteurs linguistiques (la fréquence des mots et la hauteur des voyelles), sur un facteur social (le sexe) et sur la variation entre les locuteurs individuels, ainsi que les effets de cette variation individuelle sur l'utilisation relative du DEV et de la f0 dans la signalisation des groupes de plosives. Nos modèles de régression ont estimé l'importance des facteurs linguistiques et sociaux. De plus, l'emploi des indices acoustiques (ce que l'on nomme également le «poids») pour chaque locuteur a été quantifié de plusieurs façons: le $D$ de Cohen, des analyses de discriminantes linéaires et des machines à vecteurs de support. En allemand et en anglais, un indice acoustique est systématiquement moins utilisé dans les mots à haute fréquence et devant les voyelles non hautes, réduisant ainsi la quantité d'information offerte par cet indice pour ce qui est des contrastes entre les plosives. De plus, ce sont les mêmes conditions où la transphonologisation en coréen de Séoul est la plus avancée. Les résultats pour les sexes et pour les individus, par contre, ont démontré qu'il existe des corrélations systématiques dans toutes les langues.

Pris dans son ensemble, les résultats de ces deux études démontrent que le compromis d'importance relative des indices acoustiques associé aux contexte linguistique, observé uniquement en coréen de Séoul Coréen où l'on retrouve la transphonologisation, aurait pu s'instaurer pour contrer la réduction de « l'informativité » des indices pour éviter la perte de contraste, une tendence qui est probablement unique aux cas de transphonologisation. Cependant, les changements de l'importance des indices acoustiques selon le sexe et l'individu sont systématiques dans les trois langues, ce qui suggère que les locuteurs ont un but partagé par rapport à la quantité d'information linguistique qu'ils souhaitent transmettre tout en produisant des variantes stylistiques socialement conditionnées. La variabilité structurée entre les locuteurs peut également servir de catalyseur pour la transphonologisation et ce changement peut se propager en renforçant ces structures existantes. De façon générale, les résultats démontrent que la variabilité linguistique n'est pas aléatoire, mais qu'elle est plutôt structurée en plusieurs dimensions selon le contexte et le locuteur. Cette structure pourrait réduire la variabilité à laquelle les auditeurs font face en temps réel lors des interactions et nous aide à mieux comprendre comment le changement de son se produit au fil des générations.

# Contents

# List of Tables

# List of Figures

# Acknowledgements

It is a great pleasure to have this opportunity to thank so many special people without whom I could have never come this far.

I am deeply thankful to my amazing supervisors, Morgan Sonderegger and Meghan Clayards, for always having faith in me, even in those times I doubted myself, and for your encouraging words when I was feeling overwhelmed. Your teaching, mentoring, encouragement, and support has always been and will be invaluable to me. Thank you to all of the faculty at the McGill Linguistics Department, especially to Heather Goad for supervising my first Evaluation Paper and for offering advice and support throughout my years at McGill. Thank you to Yoonjung Kang at University of Toronto for supervising my second Evaluation Paper and for being a part of my thesis committee, sharing your expertise with me, and for being incredibly encouraging and caring. Thank you to Taejin Yoon at Sungshin Women's University for kindly sharing your aligned data with me, and for giving me advice about academia and jobs.

Thank you to my fellow students at the department–Donghyun Kim, Sepideh Mortazavinia, James Tanner, Henrison Hsieh, Jeffrey Lamontagne, Oriana Kilbourn-Ceron, Yuliya Manyakina, Yeong Woo Park, Gouming Martens, Jiajia Su, Sarah Colby, and all the others – for the conversations, stimulating discussions, and mutual support. I especially thank Jeffrey Lamontagne, James Tanner and Henrison Hsieh for being my

consultants for English grammar and computational linguistics. A special thank you to Jeffrey Lamontagne for translating my thesis abstract into French. Thank you to Andrew Hyunmin Lee for all the good and fun times around the campus. I am also thankful to several undergraduate research assistants at the Speech Learning Lab: Sara Perillo, Claire Suh, Alissa Azzimmaturo, and Charlene Alcena for their time and efforts with data annotations and assistance in my experiments.

A very special thanks goes to my dearest son, Jae Hyoung Hur, and my husband, Junho Hur, for their love and support in everything I have ever done. A huge thanks to my parents as well for their unwavering love.

# Contribution of Authors

The two studies presented in this thesis were prepared as manuscripts for publication elsewhere. I am the primary author of each manuscript. Chapter 2 has been accepted for publication in *Journal of Phonetics* as a co-authored article entitled 'The emergence, progress, and impact of sound change in progress in Seoul Korean: Implications for mechanisms of tonogenesis' with Prof with Prof. Morgan Sonderegger, Prof. Yoonjung Kang, Prof. Meghan Clayards, and Prof. Taejin Yoon. I was responsible for the original conception of the research questions, data preparation, and doing the bulk of writing for the paper. Prof. Kang and Prof. Yoon contributed manually annotated corpus data for training the automatic VOT aligner. I was in charge of training the aligner as well as data cleaning. The statistical analyses were prepared in collaboration with Prof. Sonderegger. Prof. Kang and Prof. Morgan Sonderegger helped further refine the research questions. All the coauthors, including Prof. Clayards, helped develop interpretations of the results and took part in several revisions of the manuscript.

The abstract for Chapter 3 was submitted to a special issue of *Journal of Phonetics* and has been selected for submission as a full paper co-authored with Prof with Prof. Morgan Sonderegger and Prof. Meghan Clayards. I was responsible for corpus data selection, developing research questions, performing statistical analyses, and writing the paper. I developed statistical analyses in consultation with Prof. Sonderegger. The

research questions and interpretation of the results were refined in conjunction with Prof. Clayards and Prof. Sonderegger, who also provided feedback on paper drafts.

# Chapter 1

# Introduction

Every day we produce, transmit, and perceive speech. The speech signal contains a multitude of spectral and temporal information that is highly complex and variable, due to many factors. In speech science, each piece of information that signals a contrast between sounds is called a *cue*. One fundamental challenge for theories of speech processing is to describe how listeners successfully retrieve a common set of linguistic units from highly variable acoustic cues.[1] Much is known so far about how listeners do this: listeners are sensitive to variability in the speech signal (McMurray, Clayards, Tanenhaus & Aslin, 2008; Miller & Volaitis, 1989; Pind, 1995; Summerfield, 1981); multiple cues are used in making categorical judgments (Abramson & Lisker, 1985; Bailey & Summerfield, 1980; Lisker, 1975; Repp, 1982; Summerfield & Haggard, 1977); in categorical judgements, some cues are weighted more highly than others (Abramson & Lisker, 1985; Lisker, Liberman, Erickson, Dechovitz & Mandler, 1977); less important cues have greater weights when the information given by more important cues is ambiguous, a phenomenon called *trading*

[1]The term 'acoustic cues' is often restricted to acoustic correlates of a contrast that have also been shown to signal the contrast for the listener. Here we use the term more broadly on the assumption that any acoustic correlate that is used by talkers to consistently mark a contrast will be used by listeners who have experience with the language.

*relations* (e.g. Repp, 1982). However, much is still not known about how cue weighting in perception is modulated by different sources of signal variability, for example whether cue weights also trade off across phonetic contexts. Given these trade-offs in perception, an important question is: "do tendancies towards cue trade-offs in perception come from trade-offs between cues in production?" Identifying *what variation in cue weights is present in the signal* is one focus of this dissertation.

Understanding cue (co)variation is also important for understanding how and why the way speech sounds are pronounced changes over generations in a speech community, or *sound change*. One type of sound change where these questions are commonly addressed is change in the primacy of different cues to the same contrast over time (across speakers in a community). Previous studies have suggested that this type of change is triggered by the presence in the signal of systematic cue covariability in productions of the same contrast (Hombert, Ohala & Ewan, 1979; Kingston, 2011), and speakers' selection of certain variants from this pool of varaibility (Lindblom, Guion, Hura, Moon & Willerman, 1995). However, the mechanisms by which cue covariability leads to a new pronunciation norm in a speech community are far from being understood. Another focus of this dissertation is to clarify this issue by examining a particular case of sound change, where the relative role of different cues to stop voicing categories (e.g. /b/ versus /p/) changes over time from a consonantal cue (e.g. VOT) to a vocalic cue (e.g. f0).

I begin in the current chapter (Chapter 1) by summarizing previous findings on cue variability in speech production, including known sources of cue variability, covariation of multiple cues and their effects on speech perception, and the contribution of cue covariation to sound change. In Chapter 2, I report the design, method, and results of Study 1, where I examined how change in cue primacy (i.e. *transphonologization*) orig-

inates in a language (Seoul Korean), how it progresses across speakers, contexts, and words, and how it affects other aspects of the language's sound system. In Chapter 3, I report Study 2, which examines how variability in multiple acoustic cues to the same contrast is structured, and what structure is common across languages and could serve as a precondition to sound change. Both studies use large corpus datasets and conservative statistical methods, to strengthen their results. The findings of these two studies should broaden our understanding of how human cognition copes with different sources of speech variability, as well as what elements of speech variability reshape how language users use speech cues in production and perception of speech sounds.

## 1.1 Background

The speech signal contains a multitude of acoustic cues that are used in speech categorization. These cues include spectral cues such as fundamental frequency (f0) and formants (F1, F2, F3, etc.), as well as temporal cues such as voice onset time (VOT) and segmental duration. In general, many cues (more than 1–2) provide integrated information for categorizing speech sounds, but all cues are not treated equally by speakers and listeners. For example, /t/ in 'tip' and /d/ in 'dip' in Engish are primarily differentiated by VOT, the interval between the release of a stop consonant and the onset of voicing, which is the most salient cue distinguishing between 'voiced' and 'voiceless' categories of stop sounds in many languages (Cho & Ladefoged, 1999; Lisker & Abramson, 1964). These sounds are additionally differentiated by many other cues including the f0 and F1 of the following vowel (Abramson & Lisker, 1985; Lisker, 1975; Whalen, Abramson, Lisker & Mody, 1993).

However, the information provided by acoustic cues is complicated by speech signal

variability in both the spectral and temporal domains. The sources of variability encompass many factors, including phonetic context (e.g. Liberman, Cooper, Shankweiler & Studdert-Kennedy, 1974), frequency of usage (e.g. Aylett & Turk, 2004), speaking rate (e.g. Miller, Green & Reeves, 1986), and inter-speaker differences caused by either physiological differences (Peterson & Barney, 1952) or social factors (Drager, 2011b; Hay & Drager, 2010; Hay, Warren & Drager, 2006).

This signal variability has been termed the 'lack of invariance problem' (Liberman et al., 1974) in the perception literature. Lack-of-invariance has been an important theme in speech perception research, because despite the massive variability, listeners generally succeed in retrieving the speaker's linguistic intention. Variabiliy in the signal is also closely linked to sound change, given that all change begins with some variation in how sounds are pronounced, and possibly how this variation is perceived by listeners (e.g. Baker, Archangeli & Mielke, 2011; Beddor, 2009; Lindblom et al., 1995). Therefore, knowing how variants are structured in the signal itself should provide insight into the mechanisms of both speech perception and sound change.

In the following subsections, I first discuss how some known sources of variability affect individual cues (Sec. 1.1.1), then how these same sources affect cue covariation (Sec. 1.1.2). I then move on to the relationship between cue covariation and the specific case of sound change considered in this thesis, that involves change in cue primacy (Sec. 1.1.3). In Sec. 1.1.4, I summarize the goals of this dissertation.

### 1.1.1 Cue variability

In this section, I discuss the known factors that cause variability in acoustic cues in speech production (Sec. 1.1.1.1–Sec. 1.1.1.3) and what roles these sources of variability play in

speech perception (Sec. 1.1.1.4).

#### 1.1.1.1 Phonetic context

One well-known factor that affects speech signals is phonetic context (Liberman et al., 1974). Phonetic context effects refer to the fact that an acoustic cue's value carries information about not only the target sound but also surrounding segments. Articulatorily, these effects are caused by speech gestures of one sound influencing the realization of other nearby sounds, a phenomenon referred to as *coarticulation*.

Common examples of coarticulatory effects on acoustic cues can be found in studies that examine the production of CV sequences. These studies have shown that consonantal place of articulation affects the transition of the second formant (F2) in the adjacent vowel (e.g. Sussman & Shore, 1996). For instance, the F2 of a high back vowel is greatly affected by the place and manner of the preceding consonant. This consonant-to-vowel coarticulatory effect is known to serve as a precondition to diachronic back vowel fronting (Harrington, Kleber & Reubold, 2008). Another example comes from the effect of lip rounding for the articulation of a round vowel affecting the spectral cues of a neighboring consonant. For example, the mean spectral energy (center of gravity: CoG) of /s/ is substantially lowered by an adjecent vowel's rounding gesture, due to its lengthening effect on the resonant cavity in front of the alveolar contriction (Fujisaki & O Kunisaki, 1976; Heinz & Stevens, 1961). As a result, talkers produce /s/ in English words such as 'Sue' with lip rounding in anticipation of the rounded vowel /u/. On the other hand, no rounding occurs during the production of /s/ in the word 'sea', which results in the CoG value of 'Sue' being lower than that of 'Sea'.

As such, coarticulatory effects primarily result from the shape and anatomy of human

speech organs. However, the effect of coarticulation on acoustic cues is not entirely constrained by the mechanical properties of speech production. Studies have found that the magnitude of coarticulation is language-specific (e.g. Solé, 2007). For example, American English speakers' lip rounding begins more or less within the temporal bounds of a syllable (Bell-Berti & Harris, 1979). In contrast, French speakers initiate lip rounding up to several sounds before producing the vowel, sometimes spanning word boundaries (Benguerel & Cowan, 1974). These inter-language differences in the magnitude of coarticulatory effects may relate to why the same coarticulatory phenomenon triggers a phonological change in some languages, but remains phonetic variability in others. I will return to this issue in Chapter 2, when I discuss the findings of Study 1 examining the path of transphonologization in Seoul Korean.

#### 1.1.1.2 Frequency of usage

Another source of variability in speech signals comes from frequency of usage, associated with how 'predictable' the meaning of a linguistic unit is in real-time speech processing. One typical example is lexical frequency, which influences a word's pronunciation along a hypo/hyper-articulation dimension (Lindblom, 1990), with high frequency words exhibiting typical properties of hypo-speech.

A number of studies have found evidence that words with high token frequency, or high probability in word sequences, undergo gestural reduction or lenition (Aylett & Turk, 2004, 2006; Bybee, 2002; Jurafsky, Bell, Gregory & Raymond, 2001; Phillips, 1984; Pierrehumbert, 2001). Acoustically, high-frequency words are shorter in duration (e.g. Gahl, 2008), exhibit more frequent deletion of sounds, and have vowels produced as more centralized in vowel space (Aylett & Turk, 2006; Munson & Solomon, 2004) than low-

frequency words. A well-known example comes from a study of homophones (Gahl, 2008) . Gahl (2008) found that the homophones that differ in token frequency systematically differ in their acoustic realization. For example, for the homophones 'time' and 'thyme', the lower frequency word (thyme) is longer in duration than the higher frequency word (time). This finding is consistent with the view that there is an inverse relationship between token probability and acoustic redundancy (Aylett & Turk, 2004, 2006).

As for the effect of word frequency on acoustic cues, most previous findings address vowel formants and VOT. Vowels are more centralized in F1/F2 space when they occur in high frequency words (Aylett & Turk, 2006; Munson & Solomon, 2004). VOT for voiceless stops is shorter for more frequent words than less frequent words (Sonderegger, Bane & Graff, 2017; Stuart-Smith, Sonderegger & Rathcke, 2015; Yao, 2009). Concerning the effect of word frequency on f0, previous work is limited to the use of f0 in tonal contrats, rather than f0 as one of the cues signaling a segmental contrast such as a voicing contrast (Zhao & Jurafsky, 2009). Furthermore, nothing is known about how word frequency affects cues in multiple dimensions. The two studies conducted in this dissertation examine both synchronic and diachronic effects of word frequency on multiple cues, in particular VOT and f0. The results of these studies should contribute to a better understanding of the effect of word frequency on cue covariation.

#### 1.1.1.3 Gender and Talkers

Besides linguistic effects that contribute to signal variability, differences among individuals are also a well-known source of this variability. One important source of talker variability is differences in the morphology of individuals' vocal tracts, such as males having longer and thicker vocal folds and larger vocal tracts than female speakers, which

is largely responsible for gender differences between males and females in certain acoustic cues. For example, male speakers have lower f0 and formant values than female speakers because f0 is primarily determined by the length and mass of the vocal folds, and formant frequencies by the size and configuration of the vocal tract (Stevens & House, 1955; Titze, 1994. Fant (1966) reported that a difference in the ratio of front and back cavity sizes is one major source of sex differences in vowel formants (see Simpson, 2009 for details and more examples).

Gender differences were also found in temporal cues, such as VOT for word-initial stops. For English voiceless stops (e.g. Koenig, 2000; Robb, Gilbert & Lerman, 2005; Ryalls, Zipprer & Baldauff, 1997; Swartz, 1992) and voiced stops (Koenig, 2000; Robb et al., 2005; Ryalls et al., 1997; Swartz, 1992), studies have reported that that female speakers tend to produce longer VOTs than male speakers. The tendency of female speakers to produce both long- and short-lag stops with longer VOTs than male speakers suggests that the gender differences in VOT are at least partly due to differences in vocal tract structures between men and women; women's smaller vocal tracts may delay glottal vibration, increasing VOT. However, other studies have reported conflicting results for the production of short-lag stops. Whiteside & Irving (1998) and Smith (1978) reported longer VOTs for male speakers than for female speakers when considering the short-lag category alone. The tendency of female speakers to produce long-lag stops with a longer VOT and short-lag stops with a shorter VOT than male speakers suggests the possibility that female speakers produce more 'prototypical' stops than male speakers, due to women's tendency to speak more clearly (i.e. hyperarticulate) more than men (Ferguson, 2004).

As such, gender differences in speech acoustics cannot be accounted for solely by

physiological differences between men and women. Another piece of evidence comes from a classic modelling study that found that tube size differences between male and female speakers were not sufficient to model adult male and female acoustic vowel systems (Fant, 1989), which is a modified view from his earlier studies that reported physiologically-grounded gender differences (e.g. Fant, 1966). This finding, as Fant points out, suggests that some portion of gender-specific patterns in speech cues is due to male and female speakers learning to speak according to gender norms (Eckert, 2000; Labov, 1990).

Talker variability in the production of VOTs (beyond gender effects) has also been observed, even after speaking rate and dialectal differences were controlled for (Allen, Miller & DeSteno, 2003; Chodroff & Wilson, 2017; Newman, Clouse & Burnham, 2001; Scobbie, 2006). Talker variability has also been observed in other cues. For example, the production of CoG to signal the place of articulation of English sibilant fricatives exhibited high individual variability that cannot be solely attributed to gender differences (Newman et al., 2001). Bang & Clayards (2016) find that talkers who produce English long-lag stops with a greater VOT tend to produce English /s/ with a higher CoG, suggesting that individual differences in speech production are systematic across sound contrasts and across temporal and spectral cues. Furthermore, perception studies have shown that listeners are sensitive to talker variability in the use of a cue and able to learn and adjust to this variability in speech perception tasks (Theodore & Miller, 2010; Theodore, Myers & Lomibao, 2015).

In summary, previous studies suggest that men and women as well as individuals may differ in the degree of hyperarticulation (Lindblom, 1990), likely associated with talkers' idiosyncratic differences in the way they maneuver their articulators (Johnson, Ladefoged & Lindau, 1993).

The effects of gender on *multiple cues* are, however, not known to date. The two studies conducted in this dissertation include gender as one potential factor affecting multiple cues to sound contrasts.

#### 1.1.1.4 Variability in perception

As discussed in the previous sections, the acoustics of speech signals are characterized by massive variability due to many sources, including phonetic context, word frequency, talker gender, and individual differences. Despite this variability, listeners can easily recognize speech. Thus, theories of speech perception have sought to explain how listeners cope with signal variability (Pisoni & Luce, 1987; Summerfield, 1981).

In fact, the sources of variability are not ignored, but have been found to affect listeners' labelling behavior. Due to this context-dependent perception, the same value of a cue can be judged as one category (e.g. /t/) or another (e.g. /d/) depending on other factors. For example, listeners' judgements in categorical speech perception tasks can vary depending on whether a talker speaks fast or slowly (McMurray et al., 2008; Miller & Volaitis, 1989; Pind, 1995; Summerfield, 1981). A common finding of these studies is that listeners estimate speaking rate from multiple sources of information, including the rate of the preceding sentence, the duration of the preceding syllable, and/or the length of the following vowel. The perception of slow speech rate shifts perception such that a longer VOT is required to perceive a voiceless stop.

Listeners also compensate for phonetic context effects (Mann, 1980, 1981; Mann & Repp, 1980). Mann (1981) reported that when listeners were exposed to the [ta]-[ka] continuum preceded by /s/-like noise and /ʃ/-like noise, more [k] identification occurred in the /s/ context. This bias is, according to the authors, due to listeners' tacit knowedge

about fronted /k/ constriction during articulation in the context of /s/. This knowledge makes listeners perceptually compensate for the fronted tongue gesture. Another example can be found in the effect of a round vowel context on the preceding consonant. Studies have found that listeners are biased towards /s/ when a sibilant noise is ambiguous between /s/ and /ʃ/ in the /u/ context but not in the /a/ context (Mann & Repp, 1980; Mitterer, 2006). This bias is arguably due to listeners taking into account the lowered CoG of /s/ as a result of lip rounding for /u/. Similarly, a preceding liquid is found to influence listeners' perception of a stop's place of articulation. Listeners are biased towards /g/ when the preceding context is [al], relative to when the preceding context is [ar]. Similarly, listeners are biased towards hearing a nasal vowel as an oral vowel in the context of nasal consonants, as they attribute the nasality in the signal to the consonant (Kawasaki, 1986).

Listeners also can compensate for talker variability. For example, listeners can compute each speaker's mean f0 (pitch) after a very short exposure to voice samples (Honorof & Whalen, 2005). This f0 estimate enables listeners to make an inference about talker gender, and to use this information to compensate for gender differences in perceiving vowel formants (Strange, 1989). Nearey (1989) also showed that estimates of the size of the vowel space could be an effective approach to speaker normalization.

There is also evidence that listeners are sensitive to socio-indexical information in the speech signal (Drager, 2011a; Hay et al., 2006; Johnson, Strand & D'Imperio, 1999; Strand & Johnson, 1996). Listeners' judgements on categorical boundaries on a single cue dimension can shift depending on the speaker's sex (Johnson:1999, Strand:1999), age (Drager, 2011a), and socioeconomic status (Hay et al., 2006). For example, in New Zealand English where a merger between the NEAR and SQUARE diphthongs is in

progress, the accuracy of identifying tokens as distinct words increases if the perceived voice is older or from a higher social class. New Zealand English is also undergoing a chain shift in the vowel system. Related to this change in progress, Drager (2011) found that for resynthesized vowel tokens ambiguous between the TRAP and DRESS vowels, voices perceived as 'young' triggered bias towards the TRAP vowel for older listeners. In an experiment where the perception of gender was manipulated, Strand (1999) found that listeners tended to shift the categorical boundary between /s/ and /ʃ/ based on whether the speaker's gender was male or female. Johnson et al. (1999) similarly reported the effect of perceived gender on the categorization of /ʊ/ vs. /ʌ/.

The findings of these laboratory studies are consistent with the results of several computational studies. McMurray & Jongman (2011) found that among various perception models with different assumptions, the model in which compensation for variability was accounted for performed the most closely to human listeners' performance. In an earlier statistical modelling study of vowel categorization, Cole, Linebaugh, Munson & McMurray (2010) found that partialing out coarticulatory and speaker-specific information from vowel formant values led to better categorization performance.

Together, the studies reviewed above suggest that variability in the signal in fact provides useful information to the listener rather than hindering speech perception, indicating that acoustic cue variability must be structured in a way that represents essential information for effective categorization of speech sounds.

### 1.1.2 Cue covariation

#### 1.1.2.1 Cue covariation in production

Another important property to consider in understanding speech variability is the multi-dimensionality of acoustic cues to speech sounds. Many studies have found that phonological contrasts, such as consonantal place of articulation (Bailey & Summerfield, 1980; Dorman, Studdert-Kennedy & Raphael, 1977; Harris, Hoffman, Liberman, Delattre & Cooper, 1958; Mann & Repp, 1980), manner of articulation (Dorman, Raphael & Liberman, 1979; Miller & Liberman, 1979), and stop voicing (Kingston & Diehl, 1994; Liberman, Delattre & Cooper, 1958; Port & Dalby, 1982), are signalled by *multiple* cues.

For example, in speech production, the place and manner of articulation of fricatives are marked by spectral properties of the frication noise (Forrest, Weismer, Milenkovic & Dougall, 1988; Hughes & Halle, 1956), F2 formant transitions at the CV boundary (Jongman, Wayland & Wong, 2000; Maniwa, Jongman & Wade, 2009; Sussman & Shore, 1996), duration and amplitude of frication noise (Jongman et al., 2000; Maniwa et al., 2009), and many other cues. Jongman et al. (2000) found that the manner of articulation of different fricatives is well distinguished by spectral peak location, spectral moments (mean, variance, skewness, and kurtosis), noise duration, normalized amplitude (noise amplitude minus vowel amplitude) both averaged across frequency regions and in the F3 region, F2 onset frequency, and the amplitude of frication noise relative to the vowel amplitude.

As for the voicing of word-initial stops in speech production, contrasts are signalled by VOT (Lisker & Abramson, 1964, 1970), f0 at the onset of the following vowel (Hombert, 1976), spectral tilt (Cho, Jun & Ladefoged, 2002), and the first spectral moment (spectral

mean) in the burst spectrum (Chodroff & Wilson, 2014). Many perception studies have also shown that these cues affect listeners' behavior in categorical judgement tasks (e.g. Abramson & Lisker, 1985; Whalen et al., 1993). Besides these cues, the length of the following vowel (McMurray et al., 2008; Summerfield, 1981), F1 onset and transition (Liberman et al., 1958), f0 trajectory (Haggard, Ambler & Callow, 1970), and aspiration amplitude (Repp, 1979) also affect listeners' perception of consonantal voicing.

Due to the multidimensionality of acoustic cues in speech, the sources of variability discussed in Sec. 1.1.1 affect not only the primary cue but also other cues that signal the same phonetic contrast. Some of these cues covary on a token-by-token basis, meaning that change in the value of one cue affects the value of another at every production. This type of token-level cue covariation can occur in case where multiple cues are intrinsically correlated because they are constrained by a single articulatory gesture or by articulatory contingency of multiple speech gestures. An example of cue covariation due to a single articulatory gesture can be found in the inverse relationship between the duration of a nasal vowel and the duration of the following nasal consonant in English VN sequences, that results from a single velum lowering gesture (Beddor, 2009). Another example is that vowels with high F1 values are longer in duration because the production of low vowels requires a jaw lowering gesture of greater magnitude than higher vowels, requiring a longer time to reach the target position (Lehiste, 1976). Cues can also intrinsically covary due to articulatory contingency, when an articulatory gesture required for one sound affects a gesture required for an adjacent sound, due mainly to muscular linkages. One example can be found in the cross-linguistic /u/ fronting phenomenon in alveolar stop contexts, which is caused by muscular linkages between tongue tip and body such that the fronting gesture of the tongue tip for the alveolar constriction prevents the tongue

body from moving to the target position for the back vowel (Harrington, 2012).

Cue covariation, however, is not limited to such automatic effects. Multiple cues to the same contrast can also covary due to reasons including communicative needs (Kingston & Diehl, 1994) and individual differences in cue use. For instance, word predictability affects duration and formants of vowels (Aylett & Turk, 2006), and speech style affects formant transitions, durations, and spectral peak energy of fricatives (Maniwa et al., 2009). With respect to individual differences, recent studies have reported that individual talkers differ in the use of VOT and the onset f0 of the following vowel for English stop voicing categorization (Clayards, 2018; Kirby, 2016; Shultz, Francis & Llanos, 2012). Some of these studies found that speakers who put more weight on one cue tend to put less weight on f0 and vice versa, suggesting that talkers may avoid providing excessive cue information to listeners, but stylistically differ in how they integrate multiple cues in speech production. However, more recently, Clayards (2018) reported an opposite pattern. This study found that talkers who place greater weight on VOT showed a trend (not statistically significant) towards placing greater weight on f0, indicating that talkers may differ along the hypo- and hyper-articulation continuum. That is, talkers on the hyper-articulation side enhance all cues together, and thus provide more robust acoustic information to listeners, compared talkers on the hypo-articulation side. This argument was further supported by the finding that the talkers who produce prototypical vowel duration for /p/ tend to produce it with prototypical VOTs.

As introduced above, gender is another factor that contributes to cue variability. In Sec. 1.1.1.3, I discussed how male and female speakers differ in the use of a single cue. However, gender differences are not limited to variability of a single cue. Male and female speakers have been found to differ in the use of *multiple* cues, although the findings to

date are limited to the case of VOT and f0 as cues to contrasting Seoul Korean stops. Variability in the use of multiple cues in Seoul Korean has been of particular interest because the language is undergoing a sound change, where the primary cue that distinguishes between aspirated (traditionally long-lag) and lenis (traditionally short-lag) stops is shifting from VOT to f0 (Kang, 2014). For example, in a corpus study, Kang (2014) found that Seoul Korean female speakers are ahead in this ongoing change, contrasting these categories with a greater f0 difference and smaller VOT difference compared to male speakers. This result suggests that Seoul Korean speakers of different genders exhibit a trade-off in the use of VOT and f0 during the sound change. However, it is not clear whether the effect of gender in the use of multiple cues is due entirely to the sound change in progress, or is indeed a cross-linguistically common way of socio-phonetically marking gender identity—that is accentuated during the sound change. The current dissertation addresses this issue in Study 2 (Chapter 3).

To sum up, multiple cues in the speech signal covary due to mechanical properties, communicative needs, individual differences, or a combination of these factors. These aspects of cue covariation may be closely related to why listeners change the relative weight of in categorization tasks depending on the contextual environment. The current dissertation examines cue (co)variation in speech production in a cross-linguistic setting. Our focus is on the *use* of multiple cues across different sources of variability, as well as its effect on sound changes where the relative importance of cues changes over time ('transphonologization').

#### 1.1.2.2 Cue covariation in perception

Multiple cues to the same phonetic contrast are also important in speech perception. Previous studies have shown that when listeners make a categorical choice between sounds, they integrate multiple acoustic cues by assigning a different weight to each cue relative to others. A *trading relation* is a phenomenon found in laboratory experiments that test listeners' judgements on sound contrasts upon exposure to (co)varying cue continua. It describes the situation where a change in the value of one cue can be offset by a change in the value of the other in the opposite direction "so as to maintain the original phonetic percept" (Repp, 1982, p. 87). Evidence for trading relations has been found in categorical judgements of stop voicing (Abramson & Lisker, 1985; Lisker, 1975; Summerfield & Haggard, 1977), stop place of articulation (Bailey & Summerfield, 1980), and manner of articulation (Dorman, Raphael & Isenberg, 1980), among other contrasts. A trading relation was also observed between silence duration and F1 onset frequency for a continuum between the two English words 'say' and 'stay' (Morrongiello, Robson, Best & Clifton, 1984).

The existence of trading relations provides evidence that multiple cues influence speech perception. Furthermore, studies of multiple cues have consistently found that there is a *primary cue* to categorization for a given speech contrast, and that less important cues have the strongest influence on perception when the information given by the primary cue is ambiguous (Whalen, Abramson, Lisker & Mody, 1990; Whalen et al., 1993). For instance, in an English voiceless and voiced stop categorization task when VOT values lie midway between /b/ and /p/ on a VOT continuum, higher f0 onset (e.g. Abramson & Lisker, 1985), higher F1 onset (e.g. Summerfield, 1981), and shorter F1 transition duration (Lisker et al., 1977) caused a bias towards voiceless categories. A similar result

was observed in listeners' use of VOT and f0 in the /zi/-/si/ categorization (Massaro & Cohen, 1976).

Individual listeners differ in the weights of cues for sound categorization. For example, Haggard et al. (1970) observed that some listeners assign greater weights to f0 than to VOT in stop voicing categorization tasks. Stevens & Klatt (1974) reported individual variability in the use of VOT and F1 onset frequency in stop voicing contrasts. Idemaru, Holt & Seltman (2012) found large individual differences in the perceptual weighting of two durational cues that distinguish Japanese singleton and geminate stop categories.

The findings that trading relations between multiple cues systematically operate in speech perception, and that listeners are sensitive to cue variability in multiple dimensions, suggest that multiple cues may be structured in the signal itself in a way that facilitates speech perception. This idea motivates one goal of the studies in this dissertation, to identify the structure of cue covariability in voicing contrasts as a function of different known sources of variability.

#### 1.1.2.3 Summary

This section has described what is known about variability in the speech signal. To summarize, the signal is highly variable across linguistic and social factors (and talkers); the variability comes from from multiple cue dimensions; listeners are sensitive to different sources of variability, and compensate for them in making categorical judgements. Taken together, listeners' use of variability in categorical judgements may be complementary to the variability found in the signal. Therefore, it seems that different sources of variability is structured in multiple acoustic cues to facilitate speech perception.

Identifying the structure of acoustic cue variation is also crucial for understanding why

and how the pronunciation of certain sounds changes over generations (i.e. transphonologization). Previous studies have shown that the shifts in how different cues are used involved in such cases of sound change are not random, but exhibit more or less predictable directions cross-linguistically. This similarity suggests that these sound changes progress by exaggerating existing structures of cue variability that are shared across languages. These aspects of sound change will be discussed further in the next sections, where I will discuss previous findings on transphonologization as well as different theories of sound change.

### 1.1.3 Sound Change

In the following subsections, I first introduce the specific types of sound change considered in this dissertation (Sec. 1.1.3.1–Sec. 1.1.3.2), then review relevant aspects of the sound change literature (Sec. 1.1.3.3–Sec. 1.1.3.5), introduce the sound change in progress in Seoul Korean (Sec. 1.1.3.6), and conclude by summarizing (Sec. 1.1.3.7).

#### 1.1.3.1 Transphonologization

*Sound change* refers to a long-term change in the perception and production target for a speech sound shared by a speech community. Some sound changes directly impact the size of the sound inventory used in a language (e.g. neutralization or tone split), while in other sound changes the pronunciation target for a set of speech segments shifts without changing the number of existing contrasts of the language (e.g. vowel chain shifts). Transphonologization is an example of the latter type of sound change, where traditionally redundant phonetic cues that signal phonetic categories become exaggerated and transphonologized as the amount of information from the traditionally primary

cue decreases and eventually reaches zero (Hyman, 1976). The cues of interest in the current dissertation are f0 and VOT, which covary to signal consonantal voicing in many languages. The transphonologization of f0 is of particular interest because this cue is strongly affected by articulatory contingency (Hombert et al., 1979; Löfqvist, Baer, McGarr & Story, 1989; but see Kingston & Diehl, 1994 for a different view), and is at the core of one typologically common type of transphonologization: the emergence of tones in a previously non-tonal language, a process termed *tonogenesis* (Matisoff, 1973).

#### 1.1.3.2 Tonogenesis

Tonogenesis is one type of transphonologization, which is relatively common across many genetically unrelated languages (e.g. Athabaskan, Germanic, Austroasiatic, Sino-Tibetan: Hombert et al., 1979; Kingston, 2011). Tonogenesis is known to originate from different phonetic sources: the influence of adjacent consonants on the f0 of adjacent vowels, intrinsic pitch differences between vowels with different heights, and the effects of prosodic factors such as stress and intonation on f0 (Kingston, 2011). The most well studied source is the first one, where contrastive tone develops as the redundant phonetic properties of consonantal voicing contrast are exaggerated and phonologized over time (Hombert, 1977). In this type of tonogenesis, high tones emerge on vowels adjacent to phonologically voiceless consonants and low tones emerge adjacent to phonologically voiced consonants.

Such cases of consonant-triggered tonogenesis are of two mostly-distinct types. In one type, tone develops from coda consonants in association with the loss of the laryngeal contrast in coda position. Languages where lexical tones emerged through this route include Vietnamese (Haudricourt, 1954), some Athabaskan languages (Kingston, 2005a), and the Uto-Aztecan language Hopi (Manaster Ramer, 1986). In the second type, tone

develops from onset consonants. This is a common type of change cross-linguistically (Hombert, 1978), which is responsible for tone in languages including many Mon-Khmer languages, the Chadic subfamily of Afroasiatic languages, and other African languages (Svantesson & House, 2006; Wolff, 1987).

Although many aspects of the development of contrastive tones via exaggerating redundant attributes of consonant articulation are well understood, the exact source and pathway of tonogenesis remains unclear: what synchronic variability is first selected, and through what route the variability is exaggerated. The overarching goal of the current dissertation is to examine these aspects of transphonologization, while contributing to a better understanding of the link between synchronic variability and diachronic change in cue primacy. Synchronic variation in the signal is indeed a necessary precursor to the emergence of longer-term sound change, as noted by Weinreich, Labov & Herzog (1968). In the next section, I introduce four of the key problems raised by the framework of Weinreich et al. (WLH).

#### 1.1.3.3 The WLH framework

An influential approach to understanding language change more generally is the framework of Weinreich et al. (1968), a foundational work in the modern study of linguistic variation and change. Weinreich et al. framed the study of language change in terms of a series of 'problems'—*actuation*, *constraints*, *transition*, *embedding*, and *evaluation*—which they argue must be addressed through empirical investigation from a perspective taking the social context of language into account. These problems fundamentally ask 'how' and 'why' diachronic change occurs, and how it is related to synchronic variation with respect to linguistic and social factors. I introduce four of the 'problems' in WLH's

framework, and how they apply to transphonologization and tonogenesis in particular.

1. The constraints problem: What are the phonetic precondition(s) to sound change—including tonogenesis? That is, what sources of f0 perturbations can be phonologized over time?

2. The actuation problem: Why does a certain change take hold in one language, but not in other languages, given that similar phonetic preconditions to the change are present in many languages? In the case of transphonologization of f0, cue covariability between f0 and VOT is present in many languages due to articulatory contingency. However, relatively few languages have undergone transphonologization and developed tones. Why does transphonologization rarely occur?

3. The transition problem: How does a language shift from the traditional pre-change state (redundant f0) to the new state (contrastive f0)? That is, which variant is selected and how does it spread across words and speakers?

4. The embedding problem: How is the innovative tonal system embedded into other aspects of the language during tonogenesis? That is, what influences does the phonologization of a traditionally redundant f0 cue exert on other properties of the linguistic system?

When framed in this way, previous work on tonogenesis can be characterized as largely focused on the constraints problem, seeking to uncover the phonetic sources of tonogenesis through articulatory (e.g. Löfqvist et al., 1989), acoustic (e.g. Hombert et al., 1979), or theoretical studies (e.g. Kingston, 2011). However, relatively little is known about the other aspects of tonogenesis: what properties of cue variability contribute to its initiation, how the change propagates through the speech community and the language,

and how it impacts other aspects of the linguistic system while transphonologization of f0 is underway.

With respect to sound change more generally, there has been a long theoretical debate concerning why sound changes begin (i.e. 'actuate'). One view emphasis the role of listeners, while another view emphasizes the role of speakers interacting with listeners. I summarize these views in the following section.

#### 1.1.3.4 Sound change originating from misparsing

One influential model of sound change is due to Ohala (1981; 1993a). According to Ohala's listener-oriented approach, sound change may arise when a redundant phonetic difference is reanalyzed as a primary contrast by the listener, which can lead to phonologization of the redundant cue over time. This approach proposes three main mechanisms of sound change: hypo-correction (i.e. failure to compensate for coarticulation), hyper-correction (i.e. overcorrection of coarticulation), and confusability of acoustically similar sounds. According to Ohala, listeners normally succeeded in compensating for the sources of variability.

Then, when and why does the listener 'misparse' variability in the speech signal differently from what the speaker intended? Ohala (1981) suggests that listeners in general succeed in compensating for contextual variation. For example, listeners can account for consonantal context effects in the case of fronted tongue position and higher F2 for /u/ in alveolar context, compared to /u/ in other contexts (Mann, 1980; Mitterer, 2006). However, on some occasions listeners may for whatever reason fail to attribute the phonetic effect (Beddor, 2009) (e.g. raised F2) to its source (e.g. /t/) without any 'teleological' intent (Ohala, 1981, 1993a). In the case of historical /u/ fronting,

listeners may accidentally interpret the high F2 of /u/ in the alveolar context as a gesture specified for the vowel, and would update their mental representation with this 'hypo-corrected' token. According to Ohala, a mini 'mini-sound change' has then occurred in an individual's mental grammar. However, these mini-changes only seldom eventually lead to community-level change.

More recently, Blevins (2004)'s Evolutionary Phonology theory of sound change also emphasizes the role of the listener, while the role of articulatory detail is also considered. Instead of 'hypo-' and 'hyper-correction' used in Ohala's model, Blevins considers 'change', 'chance', and 'choice' as the three fundamental mechanisms of sound change. The first two terms, 'change' and 'chance' are relatively similar to Ohala's terms (but without a one-to-one match), in that a new phonological representation develops as a result of the listener misparsing the speech signal. However, while 'change' refers to misperception of acoustic signals (similar to Ohala's view), Blevins' 'chance' differs from Ohala's view in referring to the mismatch between what the listener hears and how they parse the target form (e.g. a sound in a particular position), due to the intrinsic ambiguity of the form. Blevins' third term, 'choice', refers to the situation where there is no misparsing involved but the listener acquires a prototype which differs from the phonological form in the speaker's grammar by selecting from existing variants in the language, which the listener has stored in memory.

Later studies have elaborated these perception-oriented models. For example, Yu (2013) and Yu, Abrego-Collier & Sonderegger (2013) investigated the role of individuals in sound change, by determining what traits of listeners are responsible for listeners' perception of coarticulation. This line of research has focused on differences in individual listeners' cognitive processing styles, personalities, and social profiles. One interesting

finding is that there is a positive relationship between a listener's Autism Quotient score and their probability of correctly compensating for coarticulation, which indicates that listeners with less autistic traits may be the pioneers of a sound change (Yu, 2013; but see Kingston, Rich, Shen & Sered, 2015 for a different finding). Yu's findings are consistent with the Ohalian view that the failure of perceptual compensation is a key trigger of sound change, and listeners who are less likely to compensate for contextual effects than others may be the innovators of sound change.

Importantly, the focus of this view is on the role of listeners' parsing errors, while deemphasizing or ignoring the role of speakers or the sources of misparsing in speaker-listener interactions. In a competing view, the role of phonetic bias factors which are widely prevalent in spoken languages is emphasized and interpreted as a crucial cause of listeners' misparsing (Beddor, 2009; Browman & Goldstein, 1991; Lindblom et al., 1995). I provide more detail on this approach in the next section.

#### 1.1.3.5 Sound change due to production variants

Views emphasizing the role of phonetic bias in sound change stem from Lindblom's "H & H" theory of speech production (Lindblom, 1990). According to this theory, speakers tune their own speech production along a hypo- and hyper-speech continuum, balancing efficiency of articulation with listeners' needs, because speakers tacitly know how much information is needed for listeners to correctly perceive the speech signal (Lindblom et al., 1995). This view relates the mechanisms of sound change to variants in the signal, articulatory variability in the production system, and listeners' perception of such variability.

Lindblom et al. (1995) argue that phonetically-motivated sound changes tend to be triggered by reduction or lenition of speech gestures, which are characteristic of casual speech production due to greater coarticulation compared to careful speech. In this view, the 'selection' of variants at the initial stage of sound change is more likely to take place when listeners attend to 'how' a form is produced rather than 'what' is being said. Crucially, the 'how' mode operates more effectively when retrieval of semantic meaning from the signal is easier, which is associated with 'hypospeech' (Lindblom, 1990). When this mode operates and listeners update their own production to reflect what they hear (differing from the speaker's intent), sound change may occur.

The degree of hypospeech is a function of a word's predictability in a given context (Browman & Goldstein, 1991). For example, frequently used words are semantically highly predictable to the listener and therefore, speakers' productions of such words are biased towards hypoarticulation, increasing the likelihood of listeners making a parsing error (i.e. the listener's interpretation of the signal differs from the speaker's intention). This view is supported by the results of a recent articulatory study on the degree of /l/-vocalization as a function of word frequency. Lin, Beddor & Coetzee (2014) found greater reduction in the apical gesture for /l/ production in high frequency words such as 'milk' and 'help' compared to low frequency words such as 'whelp' and 'ilk'. This reduction caused substantial acoustic changes, possibly large enough to cause perception as a different phone (i.e. a rounded back vowel). This pattern of synchronic variation, they argue, may be a cause of /l/ vocalization, a common diachronic change.

Other studies have reported that the rate of misparsing increases in hypoarticulated speech or specific contexts where the source of coarticuation and its effect become ambiguous (Beddor, 2009; Harrington, Kleber & Stevens, 2016). Harrington et al. (2016)

found that the rate of misparsing associated with lip rounding increases in hypoarticulated signals, which may be a cause of diachronic /u/-fronting in standard southern British English (e.g. Harrington et al., 2008).

Context effects on speech variability are also known to play a crucial role in sound change. For example, the English allophonic s-retraction phenomenon, where /s/ is pronounced closer to [ʃ], varies greatly between phonetic contexts and speakers. Baker et al. (2011) found that s-retraction is most pronounced before /tɹ/ in onset clusters (as in 'street'), which is also the context with the most interspeaker variability. Another well-known example comes from the historical development of contrastive nasal vowels (Maddieson, 1984), which has been suggested to originate from anticipatory coarticulation of a nasal consonant in VNC structures. Beddor (2009) argues that such historical nasalization may have been triggered by listeners' experience with contexts where consonantal duration is shorter and (simultaneously) the nasalized portion of the vowel is longer. Earlier studies similarly found evidence on the effect of vowel context, such as vowel height and length, on the development of contrastive nasalization (Hajek, 1997; Hajek & Maeda, 2000).

In sum, much previous work focuses on variants present in the speech signal to explain long-term sound changes, and suggests that particular kinds of variability are closely linked to particular types of sound change. However, exactly what aspects of speech variability underlie the mechanism of transphonologization are not well-understood. This dissertation address this outstanding question using a large speech corpus from Seoul Korean, where a transphonologization sound change is in progress. In the following section, I introduce the laryngeal system of Seoul Korean and how this system is undergoing sound change.

#### 1.1.3.6 Seoul Korean

Seoul Korean has a typologically unusual three-way laryngeal contrast of voiceless stops (and affricates) in phrase-initial position; the stop series are commonly termed *tense* (or *fortis*), *lax* (or *lenis*), and *aspirated*. The three categories are signalled by a combination of acoustic cues, realized on the consonant or vowel: primarily VOT and f0, also closure duration, $F_1$ trajectory, and H1−H2 (breathiness) (Cho et al., 2002; Cho & Keating, 2001; Hardcastle, 1973; Jun, 1996; Kang & Guion, 2008; Lee & Jongman, 2012; Lisker & Abramson, 1964; Park, 2002).

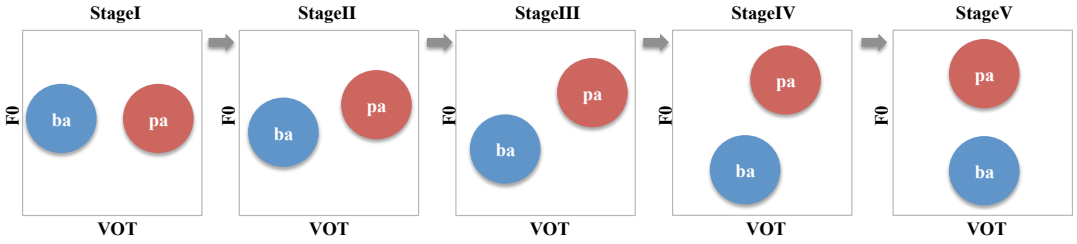


**Figure 1.1:** Five stages of tonogenesis adopted from Kang (2014) (originally Maran (1973)): Stage I: exclusive based on VOT, absence of f0 distinction, Stage II: emergence of redundant f0 distinction, Stage III: increased redundancy, Stage IV: reduced VOT distinction and further enhanced f0 distinction, and Stage V: a full substitution of VOT by f0

In the traditional system, the contrast was realized primarily using VOT while f0 only played a secondary role (Han & Weitzman, 1967; Kang & Nagy, 2016a) in phrase-initial position. However, more recent studies have reported a change in both VOT and f0 for lax and aspirated stops: the contrast in VOT between the categories has weakened, while the f0 contrast has been enhanced (Beckman, Li & Kong, 2014; Kang, 2014; Silva, 2006; Wright, 2007). The change underway in this language shows certain characteristics of a tonogenesis-like transphonologization as illustrated in Stages 3–5 in Figure 1.1. Further evidence for this sound change comes from perception studies, which found evidence that the dominant perceptual cue for distinguishing aspirated and lax stops is shifting from

VOT to f0 in recent years (Kim, 2004; Kong, Beckman & Edwards, 2011; Lee, Politzer-Ahles & Jongman, 2013).

Evidence for this sound change in progress has come especially from studies using an apparent-time sample (as this study does as well), where the realization of aspirated and lax stops are compared among different age groups of speakers of Seoul Korean in a variety of settings (Kang & Guion, 2008; Kang, 2014; Kang & Nagy, 2016a; Silva, 2006; Wright, 2007). All previous studies have found that the VOT contrast between Korean aspirated and lax stops is being reduced in younger speakers' speech, while some (but not all) also found that the f0 contrast is at the same time increasing in younger speakers' speech. Women have been found to be more advanced in the sound change, for VOT alone (Oh, 2011), or both VOT and f0 (Kang, 2014).

Kang (2014) is the first corpus study that examined sound change in progress in Seoul Korean in an apparent-time setting involving a large number of speakers. The focus of the study was to examine how gender and age modulate VOT contrast loss and f0 phonologization of Seoul Korean over time. The study found a significant gender effect: the mean VOT between aspirated and lax stops was greater for male speakers than female speakers while the mean f0 difference was greater for female speakers than male speakers. This result suggests that the decrease in VOT use and the increase in f0 use appears to be progressing concurrently across men and women. Further, based on the finding that /h/ behaves similarly to the aspirated stops (high pitch) and /n/ patterns together with the lax stops (low pitch), the study further suggests that the change is not only applied to the categories whose contrasts are at risk but targets more abstract levels of the language.

This ongoing sound change in Seoul Korean has been also simulated in a computational modelling study. Kirby (2013) performed an agent-based simulation of the Korean

sound change using data that represent the time points of 1960s and 2000s. The acoustic cues adopted in the study were ones that have been reported to be relevant to the perceptual categorization of the Korean aspirated-lax stop contrast, which include VOT, f0, duration of the following vowel, the difference in amplitude between the first two formants of the following vowel, and the amplitude of the burst. The results showed that the presence of both enhancement and bias (but neither in isolation) predicted the ongoing transphonologization rather than merger. The findings further suggested that the loss of contrast distinction of the primary cue (i.e. VOT) and the presence of a covert contrast in the most informative redundant cue (i.e. f0) could be a crucial factor in inducing transphonologization over a long time-scale.

Due to the nature of the change in progress as reported in the previous studies mentioned above, Seoul Korean should provide an ideal case study for examining the pathway and mechanism of transphonologization. Following the view that emphasizes the role of speech variants in sound change, this dissertation draws particular attention to the effects of the sources of variability on the ongoing change in Seoul Korean.

#### 1.1.3.7 Summary

Speech sounds are differentiated from one another by a multitude of acoustic cues in the speech signal. The speech signal is highly variable due to a number of factors and this variability affects multiple cues, not only the primary cue. Listeners are well aware of the variability and multidimensionality of acoustic cues and make use of this knowledge actively in speech categorization. The listener's behaviour suggests that the variability in the signal is not random but structured in a way to facilitate speech perception by providing linguistic and nonlinguistic information to the listener. A cross-linguistic study

is therefore necessary to understand how variability is structured in human sound systems and how it potentially facilitates speech production and perception.

The covariation of cues also likely has a close link to long-term changes in the pronunciation norm within a speech community over time. Among the various sound changes identified in many of the world's languages, transphonologization is of particular interest because it involves a shift in cue primacy in signalling contrasts.

Seoul Korean should provide an ideal case study for examining how phonetic variability relates to the pathway and mechanism of transphonologization because the language is undergoing a change that has many of the properties of transphonologization. Closely comparing the results from Seoul Korean with those from other languages that are not undergoing such change will therefore broaden our knowledge on the mechanism of sound change (in particular on transphonologization).

### 1.1.4 Goal of the dissertation

This dissertation focuses on identifying structures in the covariation of multiple cues across different factors and in particular how the role of one cue changes relative to other cues over time. For individuals, knowledge on the structure of variability is important for understanding what properties of speech production help listeners deal with variability across talkers, contexts, and styles to successfully recover phonetic (i.e. acoustic, articulatory) and phonological representations intended by the speaker. For communities, this should help us understand why and how the way people within a community pronounce a set of sounds changes over generations as the relative role of multiple acoustic cues changes.

In Study 1, I addressed three questions: what triggers transphonologization; how is

change progressing across the language and speech community; and what impact does the enhancement of a traditionally redundant cue exert on other aspects of the sound system of the language. To find answers to the questions, I examined how VOT and f0 covary in signaling Seoul Korean aspirated/lax stop categories across linguistic (i.e. word frequency, vowel height) and social (e.g. gender) factors over the time-course of change.

In Study 2, I aimed to identify the structures of variability in multiple acoustic dimensions in the speech signal to better understand what are the mechanisms and processes by which certain variability becomes a new pronunciation norm in a society. In this study, I investigated VOT and f0 covariation in stop voicing categorization across the factors observed in Study 1, using corpus datasets from three languages: German, English, and Korean. I compared synchronic and diachronic variation in the same factors by comparing two languages (German and English) that are not undergoing change with Seoul Korean which is. The goal of the study was to tease apart which cue correlations are common cross-linguistically (possibly as preconditions to change) and which are unique to ongoing sound change.

# Chapter 2

# Study 1

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

## 2.1 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 et al., 1979; Kingston, 2011). Tonogenesis has its origins in various phonetic sources (Kingston, 2011) but the most common and well-documented source of tonogenesis is the f0 differences in vowels adjacent to consonants with different laryngeal settings developing into contrastive tone (Hombert, 1977; Hombert et al., 1979; Löfqvist et al., 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, 2005b; 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, f0 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 f0. 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 f0 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 f0 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 2.5.4). 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$^{h}$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.[1] Despite the fact that f0 cannot be used to mark arbitrary syllables as H/L in Seoul Korean, meaning that lexical tones have not developed, 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 essentially exhibits the same type of transphonologization we find in cases of 'tonogenesis' reported in the literature (e.g. Khmer, Afrikaans), where f0 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

[1]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.

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 f0 on the following vowel, and also closure duration, $F_1$ trajectory, and breathiness (Cho et al., 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 f0 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 f0 playing a secondary role (Han & Weitzman, 1967, 1965; Hardcastle, 1973; Kang & Han, 2013; Kim, 1965). For example, Han & Weitzman (1967) found that f0 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 f0 difference has increased (Beckman et al., 2014; Silva, 2002). Jun (1996) reported both large VOT and f0 differences between lax and aspirated stops, recorded from speakers in their 30s in the 1990s.[2] f0 is also the primary perceptual cue to the lax/aspirated stop contrast in present-day Seoul Korean (Kim, Beddor & Horrocks, 2002; Kong et al., 2011; Lee et al., 2013).

More direct evidence for this sound change has come from apparent-time studies (Bailey, Wikle & Tillery, 1993; Weinreich et al., 1968) that map out the diachronic change

[2]The information of the age of speakers comes from John Kingston's personal communication with the author. I thank John Kingston for sharing this information.

by comparing the realization of aspirated and lax stops among different age groups of Seoul Korean speakers (Kang & Guion, 2008; Kang, 2014; Kang & Nagy, 2016b; 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 f0 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 f0 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 f0 (Kang, 2014). This gender effect is mirrored in perception, with listeners relying on f0 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 f0 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, 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.

## 2.2 Background

### 2.2.1 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. Berry & Moyle, 2011; Bybee, 2012; Bybee & Hopper, 2001; Hooper, 1976; Ogura, 2012; Phillips, 1984)—to which we now turn.

### 2.2.2 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 et al., 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; Berry & Moyle, 2011).

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, 1981, 1993a), 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 f0 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 et al., 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 et al., 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 2.1.

### 2.2.3 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, meaning that the size of the VOT difference between the two categories shrinks and that of f0 difference increases over time. 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 continouous 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.

#### 2.2.3.1 Predictions: Word frequency

If the sound change originates in production bias we expect to observe the pattern in Figure 2.1 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 2.1 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:

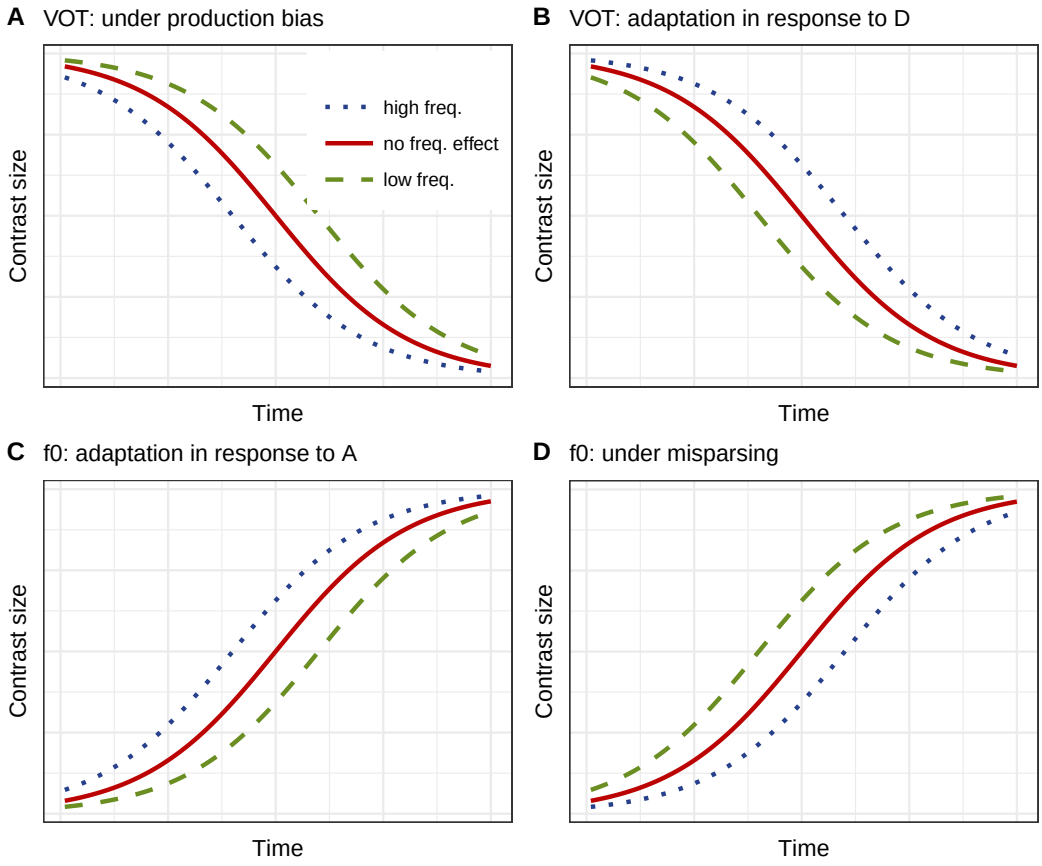


**Figure 2.1:** Hypothesized effects of word frequency on sound change in Seoul Korean: The S-curves illustrate change over time in the importance of VOT (A, B) and f0 (C, D) in contrasting aspirated and lax stop series. The solid lines represent the expected pattern if there were no frequency effect. 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 2.2. 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.

The patterns in (A) and (C) of Figure 2.2, 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; Berry & Moyle, 2011; 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

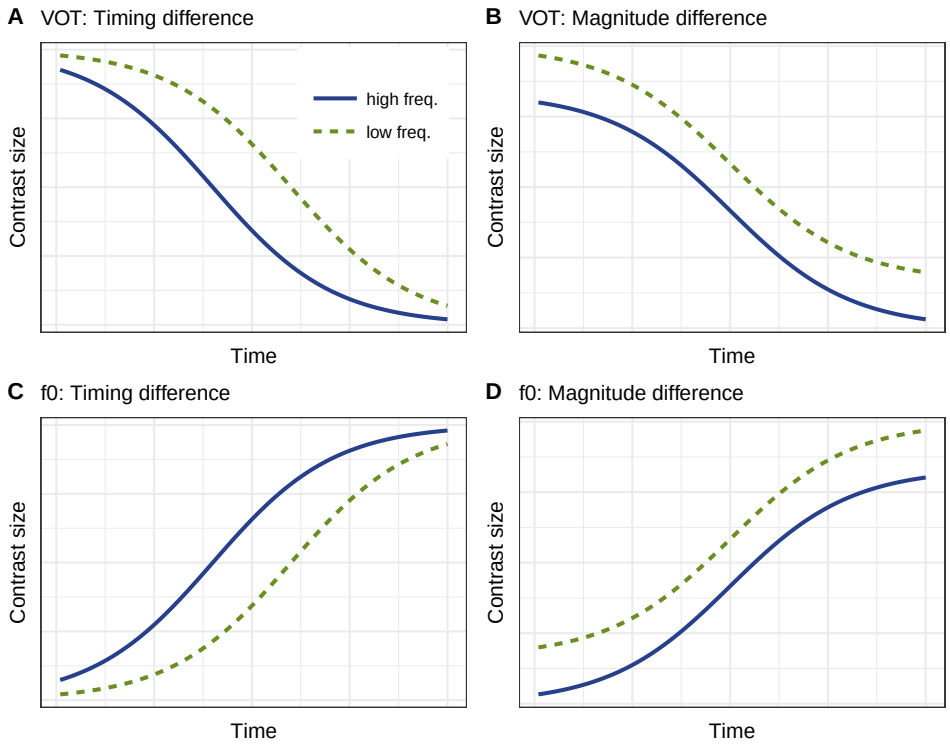


**Figure 2.2:** Schematic of effects of word frequency on sound change that would result from timing effects (A, C) versus magnitude effects (B, D). The solid (high frequency) and dotted (low frequency) lines represent the expected trajectories for words with high and low frequency. (A) and (C) are expected if the change is caused by production bias in VOT and an adaptive link to f0, as predicted in Scenario 2 (see text).

or analogy-driven changes (see Sec. 2.2.2) 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.

#### 2.2.3.2 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. 2.2.4). 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. 2.5.2.) 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 2.1, 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)?*

### 2.2.4 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, f0 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, f0 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 f0 (Honda, 1983), which is thought to be responsible for the cross-linguistic tendency of high vowels to have higher f0 than non-high vowels (i.e. *intrinsic f0* effects: IF0 effects; Lehiste, 1976; Whalen & Levitt, 1995; Whalen et al., 1995). Thus, variation in f0 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 f0 in tonogenetic sound change could affect non-contrastive variation in f0. 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 f0 within a single language, where other variables are held constant. Because f0 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 f0 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 f0 change differs by stop in Seoul Korean—f0 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 f0, 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 f0 for signaling phonological contrasts of a particular stop category.

Our third research question is: *does the IF0 effect in Seoul Korean change as contrastive f0 emerges, and does the magnitude of change in the IF0 effect differ by stop?*

## 2.3 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.

### 2.3.1 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, 2015; 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 ({alveolar, bilabial, velar} $\times$ {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 considered 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.

### 2.3.2 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 et al., 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 2.1.

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

**Table 2.1:** 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 | $p^h$ | 84.31 | 18.87 | 161.78 | 34.05 | 37 |
| | | $t^h$ | 53.06 | 20.26 | 148.28 | 37.46 | 71 |
| | | $k^h$ | 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 | $p^h$ | 72.77 | 22.65 | 201.92 | 58.49 | 126 |
| | | $t^h$ | 52.60 | 16.46 | 199.12 | 62.25 | 251 |
| | | $k^h$ | 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 | $p^h$ | 57.05 | 17.30 | 260.19 | 54.70 | 160 |
| | | $t^h$ | 40.81 | 13.67 | 265.01 | 56.43 | 335 |
| | | $k^h$ | 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 | $p^h$ | 61.94 | 18.82 | 213.94 | 97.70 | 62 |
| | | $t^h$ | 43.66 | 11.68 | 214.14 | 104.03 | 126 |
| | | $k^h$ | 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 | $p^h$ | 51.17 | 15.69 | 200.26 | 59.14 | 176 |
| | | $t^h$ | 34.45 | 11.76 | 197.83 | 63.73 | 415 |
| | | $k^h$ | 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 | $p^h$ | 48.72 | 20.98 | 231.48 | 72.27 | 66 |
| | | $t^h$ | 32.20 | 10.14 | 239.02 | 74.41 | 143 |
| | | $k^h$ | 62.95 | 14.76 | 217.37 | 74.06 | 26 |

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.

### 2.3.3 Statistical models

#### 2.3.3.1 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 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 2.3 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.[3]

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

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).

[3]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.

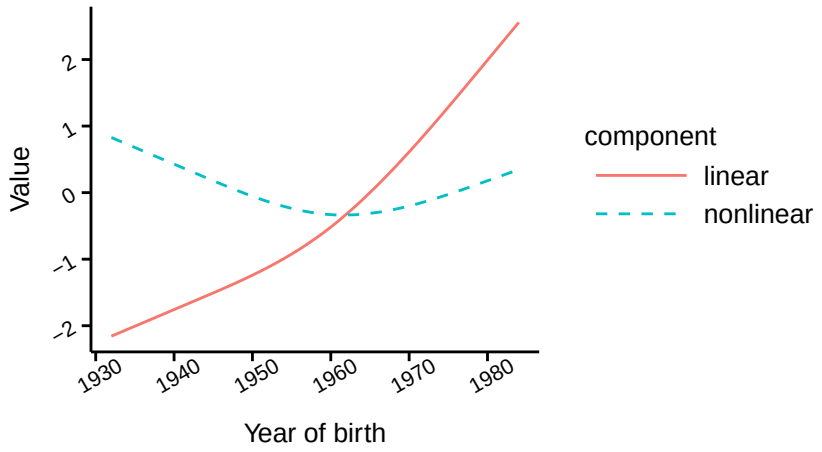


**Figure 2.3:** Values of the first ('linear') and second ('nonlinear') components of the restricted cubic spline coding of YOB, for the range of years of birth represented in the dataset.

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 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, 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 (≥ 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. 2.2.2), 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.

#### 2.3.3.2 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 (Bates, Mächler, Bolker & Walker, 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:LARYGNEAL:FREQUENCY and YOB:LARYGNEAL:HEIGHT interactions assess whether word frequency and vowel height tease apart synchronic magnitude effects and diachronic timing effects. The YOB:LARYGNEAL: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:LARYGNEAL: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. $p < 0.05$).

## 2.4 Results

The fixed effects for the statistical models of VOT and f0 are summarized in Table 2.2: 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. 2.4.1) addressed in previous work; we then turn to the effects of word frequency and following vowel height (word-level variables: Sec. 2.4.2), which are the foci of our research questions; and briefly discuss the effects of other variables included as controls (Sec. 2.4.3). 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),[4] 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

[4] 95% confidence intervals for model predictions in Figures 2.4–2.8 were calculated using the variance-covariance matrix of the fixed-effect terms.

**Table 2.2:** 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.

| | f0 | | | | | VOT | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| FULL MODELS | 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 |

other variables which are part of interactions involving $X$ (holding continuous variables at average values; averaging over categorical variables).

### 2.4.1 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 2.4 shows the empirical distributions and the model predictions of f0 and VOT by stop category, speaker year of birth, and gender.

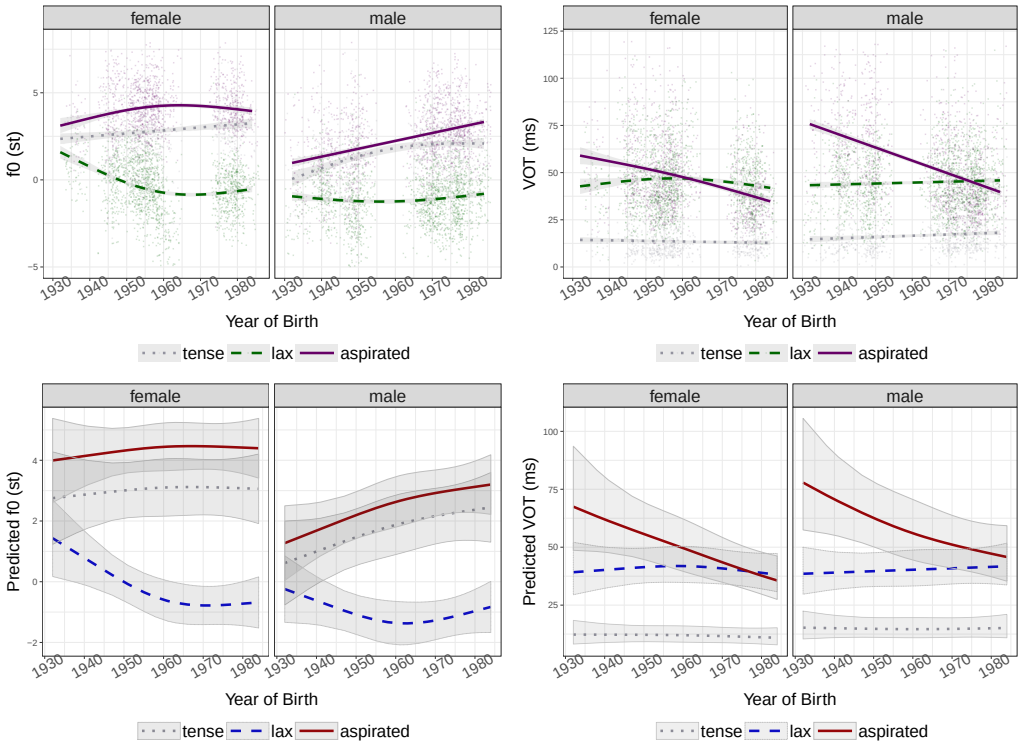


**Figure 2.4:** Empirical plots (top) and model prediction plots (bottom) for f0 (left) and for VOT (right) of three laryngeal categories for female and male speakers as a function of speaker year of birth: Lines show a quadratic smooth to empirical data or the model-predicted effect; shadings are 95% confidence intervals (CIs).

#### 2.4.1.1 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 2.4 (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 2.4, 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 as seen in Figure 2.4.

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 2.4, lower-left).

#### 2.4.1.2 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 2.4 (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. More specifically, Figure 2.4 suggests that it is the aspirated stops whose VOTs have shortened considerably over time such that they are no longer distinct from lenis stops in VOT, while lenis stops do not change very much. 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 2.4, 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 2.4 (right panels), as the sound

change being more advanced for female speakers.[5] 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 aspirated/lax VOT difference for a handful of young speakers.

#### 2.4.1.3 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.

### 2.4.2 Change across words

#### 2.4.2.1 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 2.2) by terms for the main effect of FREQUENCY and its interactions with LARYNGEAL2 and YOB. Three-way

[5]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).

interactions will be discussed in Sec. 2.4.2.3. Figures 2.5–2.6 show the empirical and model-predicted effects of word frequency on VOT and f0.

#### 2.4.2.1.1 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).

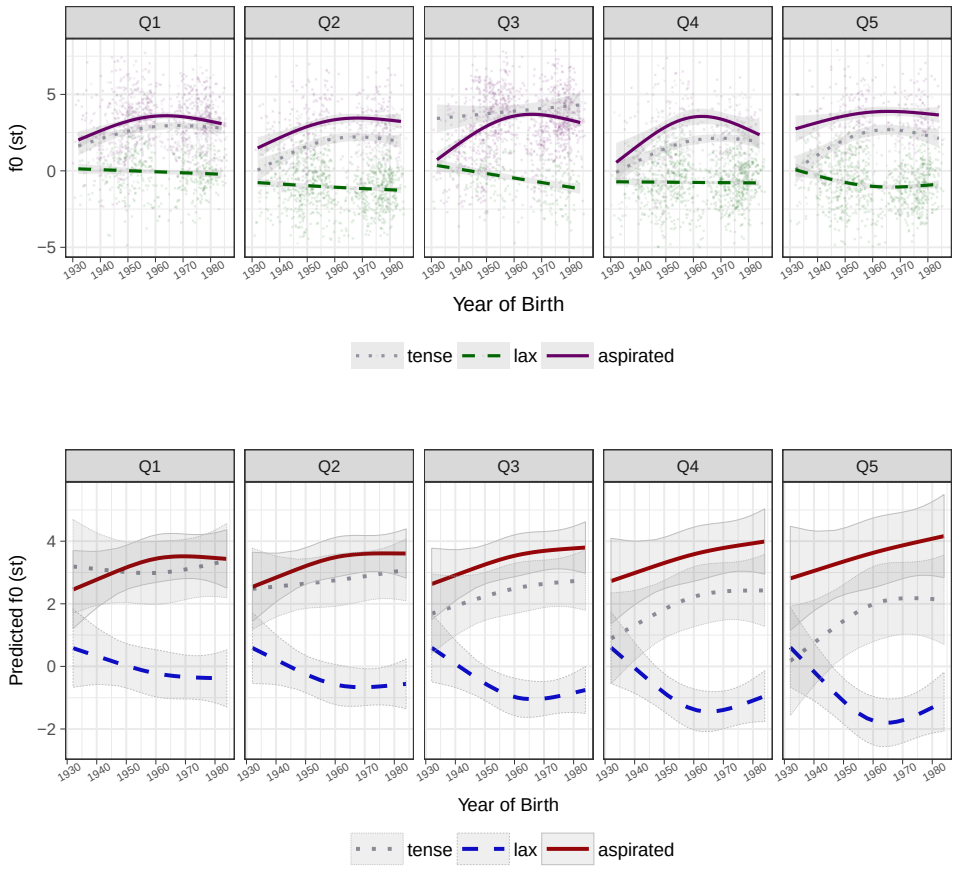


**Figure 2.5:** Empirical plots (top) and model prediction plots (bottom) of f0 as a function of word frequency & laryngeal category. Lines and shadings as in Figure 2.4. Q1–Q5 refer to word frequency quantiles from lowest (Q1) to highest (Q5).

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 2.5 as an increasing distance between the lines corresponding to aspirated and lax stops, as frequency increases.[6] 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 2.5 for tense stops, for which f0 appears to be increasing over time along with aspirated stops, as a member of the same natural class (as proposed by Kang, 2014). However, for tense stops, the change in f0 is more advanced before *lower* frequency words. This pattern makes sense if f0 in tense stops is changing by analogy with aspirated stops—since low-frequency words are expected to lead analogical sound changes.[7]

#### 2.4.2.1.2 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-

[6] A reviewer notes discrepancies between the empirical trends and model fits in word frequency effects on both VOT and f0 (in Figure 2.5 and Figure 2.6). 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.

[7] We thank an anonymous reviewer for this suggestion.

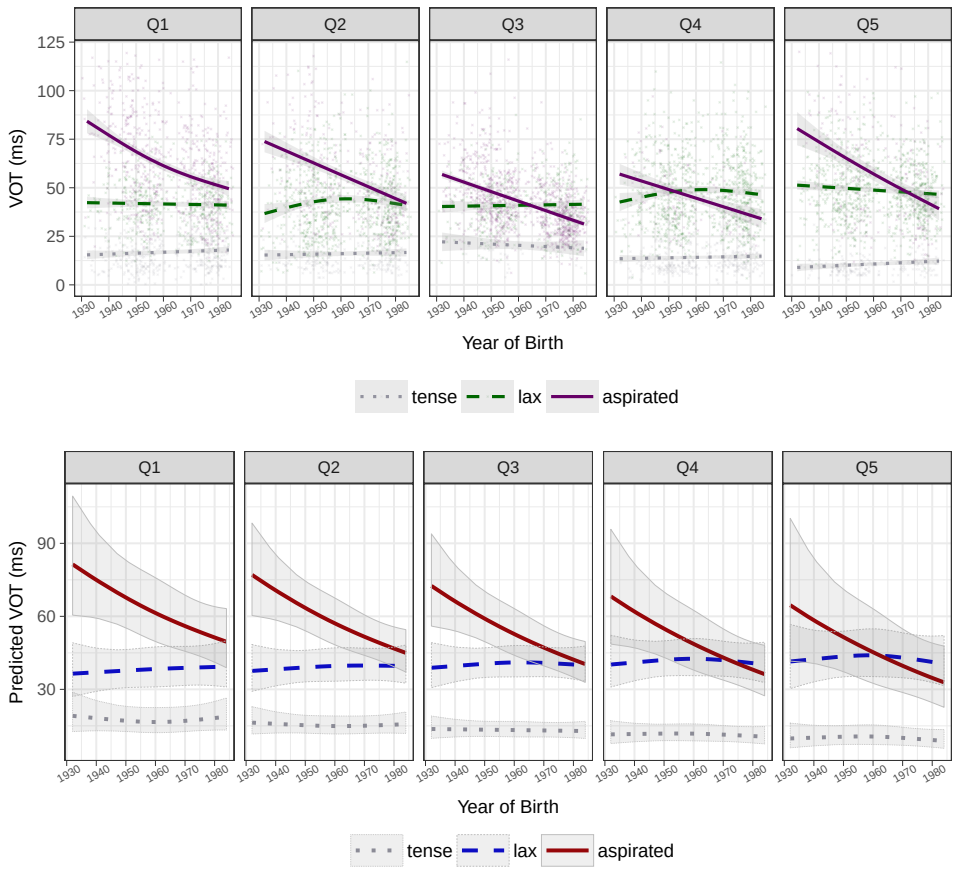


**Figure 2.6:** Empirical plots (top) and model prediction plots (bottom) of VOT as a function of word frequency & laryngeal category. Lines and shadings as in Figure 2.4. Q1–Q5 refer to word frequency quantiles from lowest (Q1) to highest (Q5).

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 2.6 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 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 2.6 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.

#### 2.4.2.2 Vowel Height

We turn to the effect of vowel height on VOT and f0, 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 f0? We examine the directionality of any vowel height effect, whether this directionality is the same or different for VOT and f0, and whether the IF0 effect varies over time and across stop categories.

The relationship between vowel height and each cue (VOT, f0), and how it changes over time, are captured in the models (Table 2.2) 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 2.7 shows the empirical and model-predicted effects of vowel height on VOT and f0, and Figure 2.8 shows the diachronic development of this effect for each stop category.

##### 2.4.2.2.1 f0: across vowel context

Concerning our second research question, the f0 difference between aspirated and lax stops is modulated by vowel height. The difference in f0 between aspirated and lax

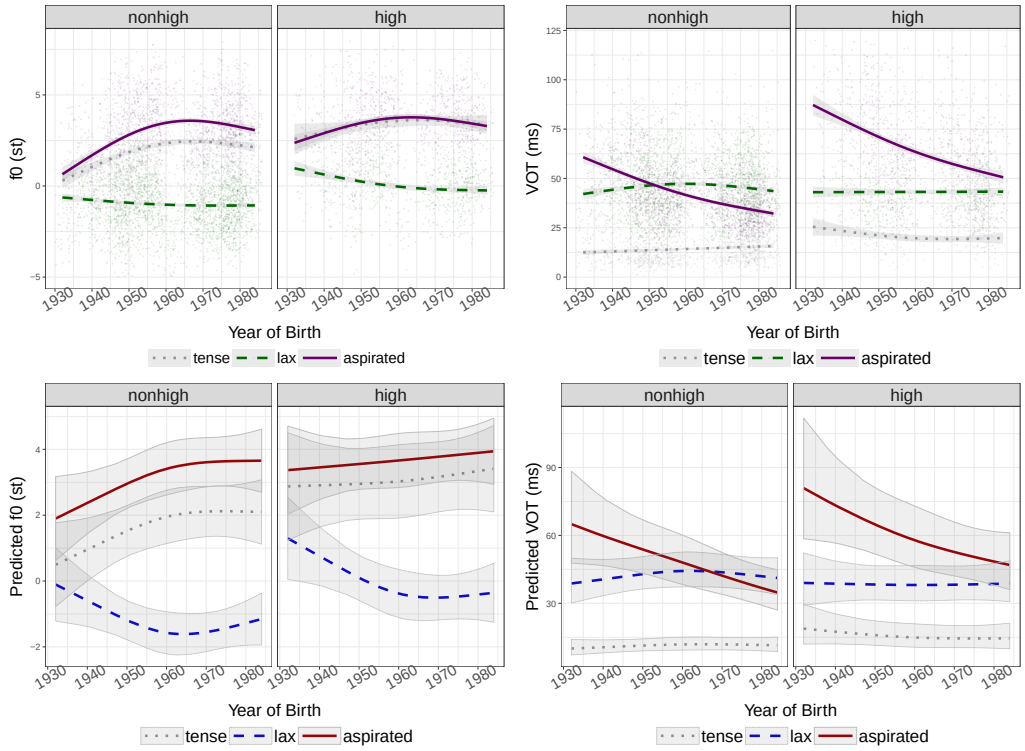


**Figure 2.7:** Empirical plots (top) and model prediction plots (bottom) of f0 (left) and VOT (right), as a function of vowel height and laryngeal category. Lines and shadings as in Figure 2.4.

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 f0 over time is more advanced in nonhigh vowel context than in high vowel context. Figure 2.7 shows that for the oldest speakers, the f0 differences between aspirated and lenis stops are about equally small in nonhigh and high vowel contexts, but the VOT differences are considerably smaller for nonhigh than high vowels. The f0 differences increase over time for both nonhigh and high contexts, as the VOT differences shrink. However, due to the VOT contrasts being greater for high than nonhigh vowels, it appears that the contrast is not being fully shifted from VOT to f0 for high vowels, even for the youngest speakers.[8]

[8]I thank John Kingston for this comment.

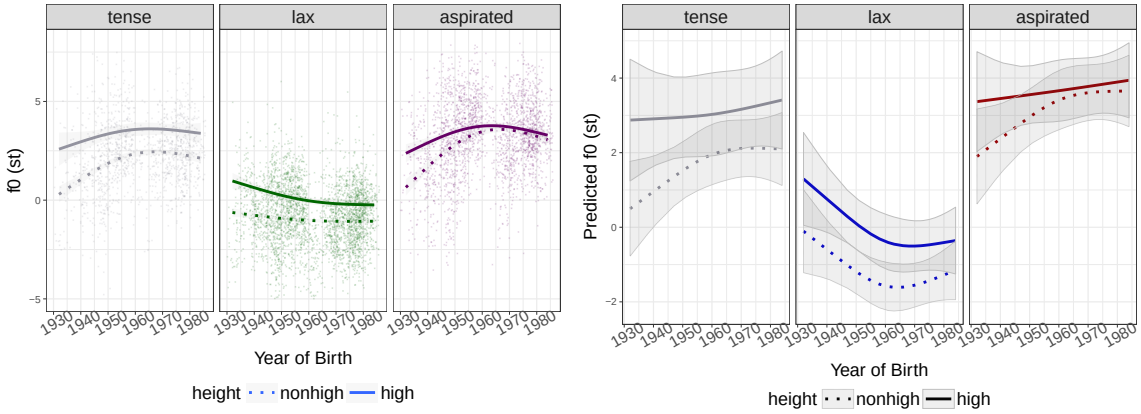


**Figure 2.8:** Empirical plots (left) and model prediction plots (right) showing a change in the size of IF0 effects over time by each laryngeal category. Lines and shadings as in Figure 2.4.

#### 2.4.2.2.2 IF0 effects

There is a significant main effect of HEIGHT: as expected (Sec. 2.2.4), high vowels have intrinsically higher f0 than low vowels ($\hat{\beta} = 0.96$, $p < 0.001$). More importantly, as illustrated in Figure 2.8, we find a significant interaction between YOB and HEIGHT: the linear term suggests that the intrinsic difference in f0 between high and nonhigh vowels is attenuated over time as contrastive f0 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 YOB″:LARYNGEAL2 effects suggest that IF0 attenuation is decelerating as the sound change is nearing completion in phrase-initial position.

#### 2.4.2.2.3 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.

#### 2.4.2.3 Magnitude versus timing effects

In Sec. 2.2.3.1 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. 2.2.3.1, 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 2.9, which show the model-predicted difference in VOT and f0 between aspirated and lax stops over time, for words with dif-

ferent frequencies and with different vowel heights (with other variables held constant, as above).[9]

We first consider three-way interactions with frequency, the evidence for which was mixed. For f0, 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 2.9(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 2.9(d).

Turning to the three-way interaction terms with vowel height: for f0, 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 2.9(a): the enhancement of the f0 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 2.9(b)).

In sum, the three-way interactions (between LARYNGEAL2, year of birth, and fre-

[9]Model predictions and 95% prediction intervals were approximated by simulation. For each model (VOT and f0), $n = 10000$ 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 2.9 and Figure 2.10.

quency/vowel height) for f0 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 f0 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 f0 contrasts over time, for different classes of words in Figure 2.9, which can be compared directly to the trajectories that were predicted under magnitude versus timing effects in Figure 2.2.

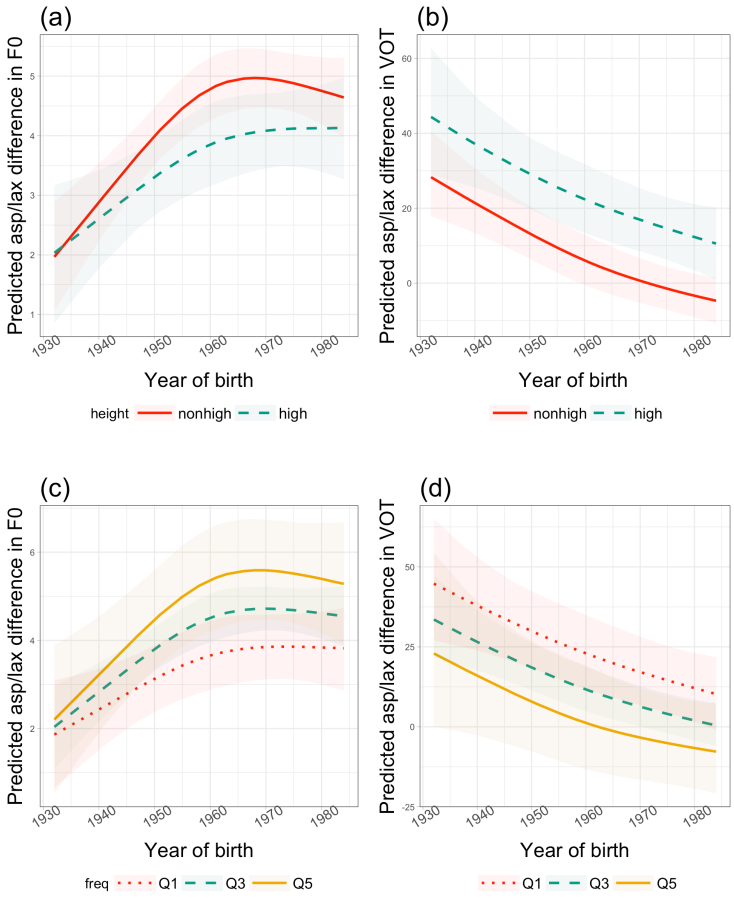


**Figure 2.9:** Model-predicted differences between aspirated and lax stop VOT and f0 over time, for different vowel heights (top row) and word frequencies (bottom row). Lines and ribbons are median model predictions and 95% prediction intervals calculated by simulation from the model posterior. Q1–Q5 refer to word frequency quantiles from lowest (Q1) to highest (Q5).

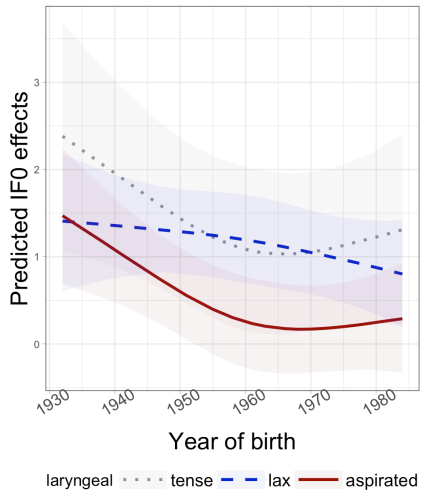


**Figure 2.10:** Model-predicted IF0 effect (f0 difference between high and non-high vowels) over time, for each class of stops. Lines and ribbons are as in Figure 2.9.

Crucially, the slopes for f0 difference seem to vary across words with different frequency (Figure 2.9(c)) and stops in different vowel contexts (Figure 2.9(a)) until the change becomes stabilized (compare to Figure 2.2(c)). In contrast, the slopes for VOT differences do not exhibit noticeable differences across words (Figure 2.9(d)) and vowel contexts (Figure 2.9(b)). Thus, the null effects for f0 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 f0 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' (f0 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 f0 model are summarized in Table A1 in the appendix. We omit discussion of most results of this model, which largely overlap with our previous f0 model, and report only the two three-way interaction terms of interest: 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 2.10, which shows the model-predicted IF0 effect (f0 for high minus non-high vowels) of stops over time (with other variables held constant).[10] 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. 2.5.3).

[10]These model predictions and prediction intervals are calculated using the same simulation-based method as for the aspirated/lax differences.

#### 2.4.2.4 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. 2.3.3.2).[11] 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. 2.5.2), in the context of what each one would mean for our research questions.

[11] Indeed, removing all terms for PLACE lowered the $p$-values of all frequency effects.

#### 2.4.2.5 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 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.

### 2.4.3 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 after a longer pause than after 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 after 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): $\hat{\beta} = 0.314$, $p < 0.001$).

## 2.5 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 coupled with the language's goal of maintaining the number of sound contrasts 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 f0 contrast enhancement are greater in the same conditions. Third, the vowel intrinsic f0 difference between high and non-high vowels is attenuated as contrastive f0 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.

### 2.5.1 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 f0 difference and smaller VOT difference between lax and aspirated stops. Note that the effect sizes of word frequency are rather small compared to other effects identified in the data. A further discussion about the small effect sizes of word frequency is provided in Sec. 2.4.2.4. 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. 2.2.2). 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 et al., 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- and low-frequency words should show different *rates* of change. We included terms in our statistical models to test the possibility that the pattern of f0 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 f0 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. This finding is in contrast with a previous result in (Jun, 1996) where the opposite timing was observed: f0 differences became large before VOT differences shrank. The difference may partly come from differences in the nature of the data analyzed in each study; Jun's data were elicited rather than spontaneous. [12]

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.

[12] I thank John Kingston for sharing this idea.

### 2.5.2 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 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 2.1) due to independent synchronic pressures operating on each cue, combined with an ongoing sound change progressing at the community level. Similarly, it might be surprising to find that there is lenition in the most prosodically strong position. One might argue instead the f0 was enhanced in the prosodically strong position, which then allowed the lenition of VOT. However this interpretation does not explain why the f0 would be enhanced in high frequency words. The pattern we observe therefore (A and C in Figure 2.1), suggests that the sound change occurring in the phrase boundaries 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 f0 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 f0 contrasts to laryngeal status are reduced requires future cross-linguistic work.

Our findings on parallel change in VOT and f0 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 f0 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. f0 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, f0). This type of compensatory link may also underly the inverse relationship between VOT and f0 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. 2.4.2.4), 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 et al., 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 f0, due to production bias coupled with an adaptive response.

### 2.5.3 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 2.4). 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.

### 2.5.4 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 as-

pirated 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.[13] 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.

[13]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).

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 et al. (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 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 f0 by a combination of prosodic and seg-

mental 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 f0 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.

## 2.6 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 f0 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 (f0), plausibly to avoid merger. We also found evidence that the vowel intrinsic pitch difference is attenuated as contrastive f0 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.

# Appendices

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

| | $F_0$ | | | | |
|---|---|---|---|---|---|
| F0 MODEL | Estimate | SE | df | $t$ | $P(>t)$ |
| 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 |

# Preface to Chapter 3

Using a dataset from a large apparent-time corpus of Seoul Korean, Chapter 2 examined 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. The focus was on word frequency, vowel height, gender, and individual speaker effects. We found that both VOT contrast reduction and f0 contrast enhancement are more advanced in high-frequency words, in stops before non-high vowels, and in young female speakers' speech, showing cue trade-offs. These findings indicate that the change is spreading across words, phonetic contexts, and speakers in parallel. Interestingly, the change is more advanced in the conditions which are presumably prone to articulatory lenition linked to decreased acoustic contrast, indicating that the change may have been triggered by phonetic bias.

Even though the findings from Seoul Korean shed light on the mechanism of transphonologization that involve cue trade-offs in detail, it is still challenging to tease apart synchronic cue trade-offs that may exist as preconditions to sound change and diachronic trade-offs that may be unique cue structures that occur during transphonologization. This motivates Study 2.

Building upon the findings of Chapter 2, in Chapter 3, I address two fundamental questions regarding the variability in the use of VOT and f0 to signal stop voicing contrasts in a cross-linguistic setting. Using speech corpus datasets from German, English, and Korean, I first seek to better describe the relationship between VOT use and f0 use in production across the same factors considered in Chapter 2. Second, by comparing the structure of the cue covariation in a language undergoing transphonologization

(Korean) and languages that are not undergoing such change (German and English), I aim to identify the extent to which the structure of cue covariation is a precursor for transphononogization versus a unique characteristic of such process. The comparison of Korean versus German and English provides a suitable ground for addressing these two questions.

# Chapter 3

# Study 2

## Structures of multiple cues to stop voicing and their impact on sound change

## 3.1 Introduction

The speech signal of a phonological contrast is comprised of a "constellation of cues"(Raphael, 2004, p. 15). These acoustic cues serve as cues to the perceptual identification of various phonological contrasts such as consonantal place of articulation (Dorman et al., 1977; Harris et al., 1958; Mann & Repp, 1980), manner of articulation (Dorman et al., 1979; Miller & Liberman, 1979), and stop voicing (Liberman et al., 1958; Port & Dalby, 1982).

In the signal, cues are highly variable due to a range of factors, such as phonetic contexts (Bang, 2017; Heinz & Stevens, 1961; Liberman et al., 1974), speaking rate (Gay, 1978; Miller et al., 1986; Pind, 1995), frequency of usage or predictability of meaning (Aylett & Turk, 2004, 2006; Jurafsky et al., 2001), as well as inter-speaker differences caused by physiological differences (Peterson & Barney, 1952) and the social groups they are associated with (Drager, 2011b; Horvath, 1985; Scobbie, 2006; Stuart-Smith, 2007).

Despite the "lack of invariance" (Liberman et al., 1974) in the signal, listeners quickly extract discrete linguistic units. One possible property that facilitates speech perception is that listeners compensate for the sources of variability, for example, by adjusting durational cues to the speaking rate (Summerfield, 1981) or by adjusting to talker variability in mean pitch (Magnuson & Nusbaum, 2007). Another property that helps speech perception is the mutlidimensional nature of the signal itself. In fact, studies have observed that certain cues such as f0 and F1 (e.g. Lehiste, 1976), CoG (or spectral mean) and f2 onset frequency (e.g. Jongman et al., 2000), and VOT and f0 (e.g. Lisker & Abramson, 1964) covary across categories, potentially providing additional acoustic information about speech categories for the listener (Kingston, 2007; Kingston & Diehl, 1994). In fact, a number of perception studies have found that listeners attend to a secondary or redundant cue, for instance, f0 in stop voicing categorization tasks to disambiguate when the primary cue, VOT, is ambiguous (Bailey & Summerfield, 1980; Dorman et al., 1979; Lisker, 1975; Repp, 1982; Summerfield & Haggard, 1977) or when the conditions are less-than-ideal (Gordon, Eberhardt & Rueckl, 1993), for example, where speech is masked by noise (Winn, Chatterjee & Idsardi, 2013). Other studies have found that listeners will stop attending to f0 when the pattern of covariation with VOT is altered (Idemaru & Holt, 2011), indicating the importance of cue covariation in perception; and that listeners track the pattern of cue covariation by context (Idemaru & Holt, 2014).

A well-known source of cue covariation is when multiple cues arise from articulatory contingency, for example the covariation between the duration of vowel nasality and the duration of the following nasal consonant as a result of the timing of velum lowering (Beddor, 2009), the covariation between vowel duration and F1 as a result of jaw raising/lowering (Lehiste, 1976 but see Solé & Ohala, 2010), and the covariation between

VOT and the f0 of the following vowel due to stiff/slack glottis (Hombert et al., 1979; Löfqvist et al., 1989 but see Kingston & Diehl, 1994). While arising from articulatory contingencies, it is also thought that the role of secondary cues can be strengthened in particular languages through phonologization (Hagège & Haudricourt, 1978; Maran, 1973). In one particular type of sound change, the role of the primary cue in signaling a contrast is eclipsed by the secondary cue, a phenomenon termed 'transphonologization'.

Recent studies suggest that during such sound change, the relative role of co-varying cues depends on speech style (Kirby, 2014), phonetic context (Bang, Sonderegger, Kang, Clayards & Yoon, 2018; Beddor, 2009);gender (Bang et al., 2018; Kang, 2014), and word frequency (Bang, Sonderegger, Kang, Clayards & Yoon, 2015; Bang et al., 2018), such that, across these sources of variability, one cue plays a larger role and the other a smaller one.

Patterns of covariation between cues in production influence listeners in speech perception tasks, indicating that listeners are aware of the covariations across cues, and may shift attention between them in a context-dependent manner. Furthermore, certain sound changes also seem to exploit covariation across cues, gradually shifting how contrasts are signaled from one cue to the other across contexts over time. Both these facts suggest that these covariations are important for both listeners and talkers. It also suggests that there may be more structure to these covariations in languages not undergoing change than has yet been described. For example there may be patterns within any language across contexts and talkers like the ones found in Seoul Korean by Bang et al., (in press; i.e. Chapter 2) and Kang (2014).

Using corpus data from three languages, Korean, English, and German, the current study examines how VOT/f0 cue weights that signal stop voicing are affected by lin-

guistic factors such as word frequency and vowel height as well as socio-linguistic factors such as gender and individual (stylistic) variations. Our first goal is to provide a better understanding of how covariation of multiple cues signaling phonetic categories is structured. By comparing the structure of the cue covariation in a language undergoing transphonologization (Korean) and languages that are not undergoing such change (German and English), our second goal is to identify the extent to which the structure of cue covariation is a precursor for transphonologization versus a unique property of this process. We believe that the comparison of Korean versus German and English provides a suitable ground for teasing apart synchronic structures and diachronic structures in cue covariation.

The following section lays out the background literature related to the research questions of our study. Sec. 3.2.1 introduces different views on the mechanism of cue covariation in association with its implications for transphonologization. In Sec. 3.2.2, we provide a summary of cross-linguistic findings on transphonologization. In Sec. 3.2.3, we review recent findings on the structure of VOT and f0 covariation across speakers. Finally, in Sec. 3.2.4, we summarize our hypotheses and predictions for VOT and f0 covariation across multiple sources of variability.

## 3.2 Background

### 3.2.1 Cue covariation across categories

It has been well attested that multiple phonetic cues covary across sound categories. However, to what extent they are intrinsically linked (Halle & Stevens, 1971; Hombert et al., 1979; Kirby & Ladd, 2016; Löfqvist et al., 1989) by an articulatory source and to what ex-

tent they are controlled for enhanced informativity (Dmitrieva, Llanos, Shultz & Francis, 2015; Kingston, 2007; Kingston & Diehl, 1994) is still a topic of debate. One well-known case of cues that covary across categories is f0 and F1, which are negatively correlated in signalling vowel categories (Lehiste, 1976; Whalen & Levitt, 1995). Cross-linguistically, high vowels such as /i/ and /u/ are acoustically signalled by lower F1 and higher f0 than non-high vowels such as /a/ and /æ/. Some researchers have suggested that the relationship between f0 and F1 is the result of muscular linkages between the tongue body and vocal folds (Honda, 1983; Whalen, Gick, Kumada & Honda, 1999), which is known as an *intrinsic* f0 difference between high and non-high vowels. An opposing view has argued that this covariation is the result of speakers' controlling articulations independently for F1 and f0 in order to mutually enhance a specific auditory property (Fahey & Diehl, 1996; Kingston, 1992; Kingston & Diehl, 1994). According to the latter view, speakers deliberately make robust f0 differences between vowels to enhance phonological contrasts.

These two opposing views are also the topic of controversy on covariation between VOT of the stop and f0 at the onset of the following vowel (henceforth, onset f0) across stop voicing categories. Within a language, a stop category with the longest VOT is associated with a higher f0 while a stop category with a shorter VOT or voicing lead has lower onset f0 (Dmitrieva et al., 2015; House & Fairbanks, 1953; Lisker & Abramson, 1964). Many researchers, though not agreeing on what exact physiological mechanism accounts for this cue covariation (see Kirby & Ladd, 2016 for detail), attribute physiological contingencies such as differences in vocal fold tension or stiffness (e.g. Halle & Stevens, 1971; Hombert et al., 1979; Löfqvist et al., 1989) and/or in aerodynamic properties of the stop closure release (e.g. Kohler, 1985) across categories to this relationship. However, similarly to the case of f0/F1 covariation across vowels, the auditory perspective

(e.g. Kingston & Diehl, 1994) argues that the speech gestures for VOT and onset f0 are independent of each other and their covariation is instead related to auditory perceptual goals of speakers.

Recent experimental studies also diverge in their views as to the mechanism behind VOT/onset f0 covariation across stop voicing contrasts. In an acoustic study, Dmitrieva et al. (2015) addressed this issue by comparing VOT/onset f0 covariation within and across voicing categories in two languages that have a phonologically similar structure of voiceless and voiced stop contrasts but are phonetically realized in a language-specific way (i.e. short-lag VOTs vs. voicing lead in Spanish; long- vs. short-lag VOTs in English). Across categories, the study found that VOT and f0 covariations are phonologically-determined, meaning that the actual phonetic realization of stop voicing categories does not affect f0 at vowel onset. Furthermore, no strong within-category correlations were observed in any category in any of the languages. These results taken to support the view that VOT/onset f0 correlation is a consequence of an intended strategy by the speaker, whose target is perceptual enhancement of a phonological distinction, consistent with the auditory view (Kingston, 2007; Kingston & Diehl, 1994).

However, a more recent acoustic study suggests that VOT and onset f0 correlations are fundamentally 'phonetic' or 'automatic' (Kirby & Ladd, 2016). The study examined VOT and f0 correlation in Italian and French, languages in which voiced stops are phonetically surfaced with voicing lead throughout the stop closure, and confirmed between-category covariations in these languages, similarly to the findings in Dmitrieva et al. (2015). However, by using /m/ as a reference condition (Hanson, 2009), the study further found acoustic evidence that the acoustic relationship is the consequence of phonetic gestures used to prevent glottal vibration for voiceless stop productions and those used to sustain

voicing for voiced stop productions. Regarding within-category correlations, as was the case in Dmitrieva et al. (2015), no effects or only very weak effects were found, which were not consistent across individuals.

Further, Kirby & Ladd (2016) add that even though VOT/f0 covariation across categories is automatic in nature, this does not necessarily exclude the possibility that speakers control or enhance the existing cue correlations. This argument is consistent with the findings of earlier articulatory studies (Hoole & Honda, 2011; Hoole et al., 2006) that directly addressed this issue. Hoole & Honda (2011) compared cricothyroid (CT) activity using EMG data with f0 acoustic data and found that German long-lag stops involve higher CT activity than short-lag stops consistently across speakers. However, Hoole & Honda's findings can be also interpreted as evidence of speakers' control of gestures intended to raise f0 for long-lag stops.[1] Further, significant variability between speakers and contexts was observed: some speakers extended the consonantal CT activity further into the vowel in certain contexts. Hoole & Honda argue that the speaker and context variability provides evidence that speakers can deliberately enhance the phonetic effects, possibly to enhance stop contrasts, by 'going with the flow' (Hoole et al., 2006, p. 360). This controlled aspect of speech—coupled with the idiosyncratic vocal tract configuration of each speaker— may explain the lack of within-category covariation found in the previous studies discussed above (Dmitrieva et al., 2015; Kirby & Ladd, 2016).

Thus there is not yet any agreement on the basis of the VOT/f0 covariation across stop voicing categories—whether it is 'controlled', 'automatic', or involves both mechanisms—in languages not undergoing change. However there is general consensus in the literature on sound change that the covariation between VOT and f0 is the source of the process whereby a VOT distinction becomes 'transphonologized' into a pitch contrast, and that

[1]Thanks to John Kingston for suggesting this alternative interpretation of Hoole & Honda's findings.

during this process, the f0 variation must become 'controlled' (Garrett & Johnson, 2013; Lindblom et al., 1995). For example, some have argued that speakers adaptively enhance the covaried cue (i.e. f0) in the face of reduced informativity of the primary cue (e.g. VOT) during transphonologization, and this adaptivity is what results in trade-off between the cues (Bang et al., in press; Bang et al., 2015; Kirby, 2013).

The finding that transphonologization progresses via cue trade-offs across different linguistic and non-linguistic conditions (Bang et al., in press; Kang, 2014) raises the question of whether adaptivity is present as synchronic variability. If so, transphonologization progresses by strengthening this synchronic structure in cue variability, which serves as a precursor to change. The current study addresses this question by examining the structure of VOT and f0 covariation used to signal stop voicing contrasts across linguistic and non linguistic conditions in three languages: phonological context, lexical properties (e.g. word frequency), social factors (e.g. gender), and individual differences. We then discuss the implications of our results for understanding diachronic change in the way multiple cues signal phonetic categories.

In the next section, we elaborate in detail the empirical findings on the relationship between sources of cue variability and transphonologization.

### 3.2.2 Transphonologization

A full review of the literature on transphonologization is beyond the scope of this chapter (see Bang et al. in press, i.e. Chapter 2, as well as Kirby, 2010). Here, I review recent studies of languages undergoing transphonologization, including sound changes in progress in Afrikaans, Phnom Penh Khmer, and Seoul Korean.

Using production and perception experiments, Coetzee (cited in Beddor 2015) showed

that Afrikaans, which traditionally had a contrast between a prevoiced ([b]-[d]) stop series and a short-lag ([p]-[t]) stop series in word-initial position, is undergoing transphonologization of f0. As a result, in present-day Afrikaans there is speaker variability across different age groups in the degree of category overlap in VOT, which is consistent with sound change in progress. For younger speakers,the stop categories are differentiated primarily by an f0 difference, with very little VOT difference, while some older speakers still show dependence on VOT.

Another relevant experimental study, though not on transphonologization of VOT and f0, examined an ongoing change in Phnom Penh Khmer where a consonantal distinction is being transphonologized into a voice quality distinction involving breathiness and f0 (Kirby, 2014). In this change, /r/ in /CrV/ forms is being replaced by vocalic acoustic cues on the following vowel. The sound change is conditioned by speaking style, as the innovative cues are more salient in hypoarticulated (i.e. colloquial) speech than in read speech.

Of most relevance to the current study is recent work examining the relationship between the use of VOT and f0 during the change in progress in Seoul Korean. Seoul Korean has a three-way stop distinction between aspirated, lenis, and tense stops A number of studies have reported that Seoul Korean aspirated (or long-lag) and lenis (or short-lag) stops are no longer differentiated by longer and shorter VOT values in phrase-initial position among young speakers in production, but rather that f0 has taken the primary role (Bang et al., 2015; Beckman et al., 2014; Kang, 2014; Kim et al., 2002; Silva, 2006; Wright, 2007). Further, several perception studies have suggested that the change in production is reflected in cue weights in perceptual categorization, as the dominant perceptual cue for distinguishing aspirated and lenis stops is shifting from VOT to f0

(Kim, 2004; Kim et al., 2002; Kong et al., 2011; Lee et al., 2013).

Two studies tracked the ongoing change in Seoul Korean using apparent time data (speakers of different age groups recorded done at the same point in time; Bailey et al., 1993) from a large speech corpus. Kang (2014) found trade-offs between the size of the VOT difference and the size of the f0 difference between aspirated and lenis stops in speech production across speakers of different ages and genders. Using a much larger dataset derived from the same speech corpus as Kang, Bang et al. (in press) further found a tight inverse relationship between the use of VOT and that of f0 for the aspirated/lenis stop categorization across words of different frequencies and phonetic contexts (different vowel heights). Importantly, this study found that the change in Seoul Korean is more advanced in non-high vowel contexts and high frequency words, which are presumably the conditions prone to contrast reduction compared to low frequency words and high vowel contexts.

The findings from Seoul Korean further suggest that cue trade-offs are likely driven by speakers' adaptive control of both cues as they attempt to maintain phonological contrasts. The question yet to be addressed is where this knowledge of cue trade-offs originates from. To what extent do the trade-offs between VOT and f0 signalling stop voicing categories exist as pronunciation variants—thus, forming a precursor to sound change in the input signal—versus being unique characteristics of a language undergoing transphonologization? This question is the focus of the current study.

The relationship between the strengths or weights of VOT and f0 in the context of stop voicing contrasts across speakers is the subject of recent work (Clayards, 2018; Kirby, 2016; Shultz et al., 2012). However, studies so far have reported inconsistent results, which we will review in the following section. Furthermore, VOT and f0 trade-offs across other

factors such as lexical (e.g. word frequency), phonetic (e.g. vowel contexts), sociophonetic (e.g. gender), or stylistic (or idiosyncratic) properties have not been examined outside of the studies in Seoul Korean just reviewed. The current study, to the best of our knowledge, is the first to examine VOT and f0 tradeoffs across such factors in speech production in languages not undergoing sound change (English and German). Before we turn to our research questions, we briefly review previous findings on how speakers differ in how they use VOT and f0 to signal stop voicing.

### 3.2.3 VOT/f0 weights across speakers

Although the relationship between VOT and f0 across stop voicing categories has attracted much attention, there has only recently been interest in the relationship between VOT and f0 weights which has so far been limited to examining patterns across *speakers* (Clayards, 2018; Kirby, 2016; Shultz et al., 2012).

In a study of American English, Shultz et al. (2012) examined the relative weighting of onset f0 and VOT in both production and perception of stop voicing. In production they used coefficients from linear discriminant function analysis (LDA) and found a negative correlation between onset f0 and VOT weights. Shultz et al. conclude, based on their results, that the primary cue to voicing for all speakers is VOT and that speakers who put more weight on VOT than others tend to put less weight on f0 and vice versa. In contrast, Clayards (2018) found a positive correlation between the strengths or weights of VOT and f0 among English speakers, using Cohen's $d$ and LDA, which is the opposite pattern to what was found by Shultz et al. (2012). In ongoing research, using LDA, Kirby has examined VOT and f0 cue weights across speakers in various languages including English, French, Italian, Khmer, Thai, and Vietnamese (Kirby, 2016). This work has found that

speaker-level trade-offs are only observed in the languages with a long-lag VOT category. Furthermore, in the English data, a trade-off was found only in citation forms and not in natural speech, suggesting that cue covariation may be task-specific.

The inconsistency between previous studies of the use of VOT and f0 for signalling stop voicing across speakers in English could be due to small sample size (8 speakers in Kirby, 2016; 20 in Clayards, 2018; and 25 in Shultz et al., 2012). The current study examines a much larger number of speakers in order to increase statistical power and obtain more reliable results. Furthermore, previous studies investigated cue covariation across speakers only. As discussed in the previous studies on Korean sound change, cue trade-offs are observed not only across speakers, but also across various factors including word frequency, vowel context, and gender, which we will focus on in the present research.

To maximize comparability across languages, we perform the same analyses for the data from each of German, English, and Korean, all languages where stop "voicing" is phonetically characterized by long-and short-lag VOTs in word (English and German) or phrase (Korean) initial position. We consider only the categories of aspirated and lenis stops from Korean here, analogous to English / German long- and short-lag categories.

Before beginning our analyses, we consider possible patterns of cue covariation.

### 3.2.4 Possible patterns of cue (co)variation

Cross-linguistically, the weights of cues across categories could covary in multiple ways and their covariation may depend on a number of different factors. We identified four possible patterns regarding how cues can co-vary across word frequency, vowel context, gender, and individual speakers, schematized in Figure 1. We expect negative correlations across talkers and contexts for Seoul Korean, which we previously argued is due at least

in part to the change in progress (i.e. it is diachronic variation, Bang et al. in press). Comparing both the direction and strength of the correlation to English and German will help us determine if this is the case. Negatively correlated VOT and f0 weights across either speakers or contexts (C in Fig. 1) could mean that 'overall informativity' is maintained but distributed differently between the two cues for different talkers or contexts (Kirby, 2013). There is some evidence for this type of correlation for VOT and f0 across speakers of different languages, even though there is inconsistency in previous work as discussed in Sec. 3.2.3. There is also evidence in favor of negative cue correlations from studies of other contrasts. Fridland, Kendall & Farrington (2014) examined the low back vowel pair /ɑ/-/ɔ/ across three dialects of American English which differ in the degree of their spectral merger and found that the smaller the spectral difference, the larger the duration difference between the vowel pair. Another piece of evidence comes from a study of nasality in English VNC syllables where a negative correlation was found between the duration of nasality in the vowel and the duration of the nasal consonant (Beddor, 2009). Interestingly, this relationship was found across speakers and contexts (i.e. the voicing of the final obstruent) which is argued to be crucially linked to historical vowel nasalization. If these negative correlations are observed between VOT and f0 in all three languages in our data, these patterns may be a general property of cue-covariation across (all) languages. This finding would also suggest this this type of variability may be linked to sound change, as a precursor to transphonologization.

On the other hand, VOT and f0 weights may be positively correlated as observed for VOT and vowel length across speakers in Clayards (2018). This type of correlation could indicate that speakers and words differ along a hypo– and hyper-articulation continuum as some speakers articulate more clearly than others and some words are articulated more

clearly than others (Lindblom, 1990). Support for this hypothesis comes from studies of a single cue dimension within a voicing category (Chodroff & Wilson, 2017; Scobbie, 2006), showing a positive cross-speaker correlation between English voiced and voiceless stop mean VOT values, possibly in part as a function of social factors (Scobbie, 2006). Positive correlations are also predicted by the cross-linguistic observation that women tend to hyperarticulate more than men (Bang, Clayards & Goad, 2017; Byrd, 1994; Diehl, Lindblom, Hoemeke & Fahey, 1996; Ferguson, 2004; Hillenbrand, Getty, Clark & Wheeler, 1995; Labov, 1990; Simpson, 2009; Smiljanic & Bradlow, 2009), and this may extend to multiple cues for a single contrast. Furthermore, there is considerable evidence that high frequency words are more reduced (Aylett & Turk, 2004; Baker & Bradlow, 2009; Bell et al., 2003; Bybee, 2000)

Alternatively, VOT and f0 across stop voicing categories may not covary; they may vary randomly, or in a structured way only in a single dimension. If variation is random, for example, across speakers or words, it would mean that the use of VOT and the use of f0 in signalling stop voicing contrasts are not systematically related to each other across speakers in the speech community or across words with different frequencies.

Alternatively, a single cue may be systematically related to another variable, such as word frequency or vowel height—for example due to hypo- and hyper-articulation —while the other cue remains relatively constant.

Figure 3.1 illustrates these four possible patterns. Here, we refer to the levels of each of factor of interest as a 'group'. 'Gender' has two groups ('male', 'female'), as does 'vowel context' ('high', 'non-high'). For variation across individual speakers, every speaker is one group. Finally, word frequency has multiple groups, ranging from 'low' to 'high'.

- Pattern A: There is no difference in the use of the cues across groups (Figure 3.1-A).

For instance, male and female speakers do not differ in the way they weight VOT and F0 in signalling stop voicing categories.

- Pattern B: Groups differ in one cue dimension but not in the other (Figure 3.1-B). For instance, male speakers rely more on VOT than female speakers but the two groups do not differ in the use of F0.

- Pattern C: Groups which weigh one cue more rely the other one less (Figure 3.1-C). For example, male speakers rely more on VOT and less on F0 than female speakers.

- Pattern D: Groups which weigh one cue more weigh the other one more (Figure 3.1-D). For instance, male speakers rely more on VOT and f0 than female speakers.

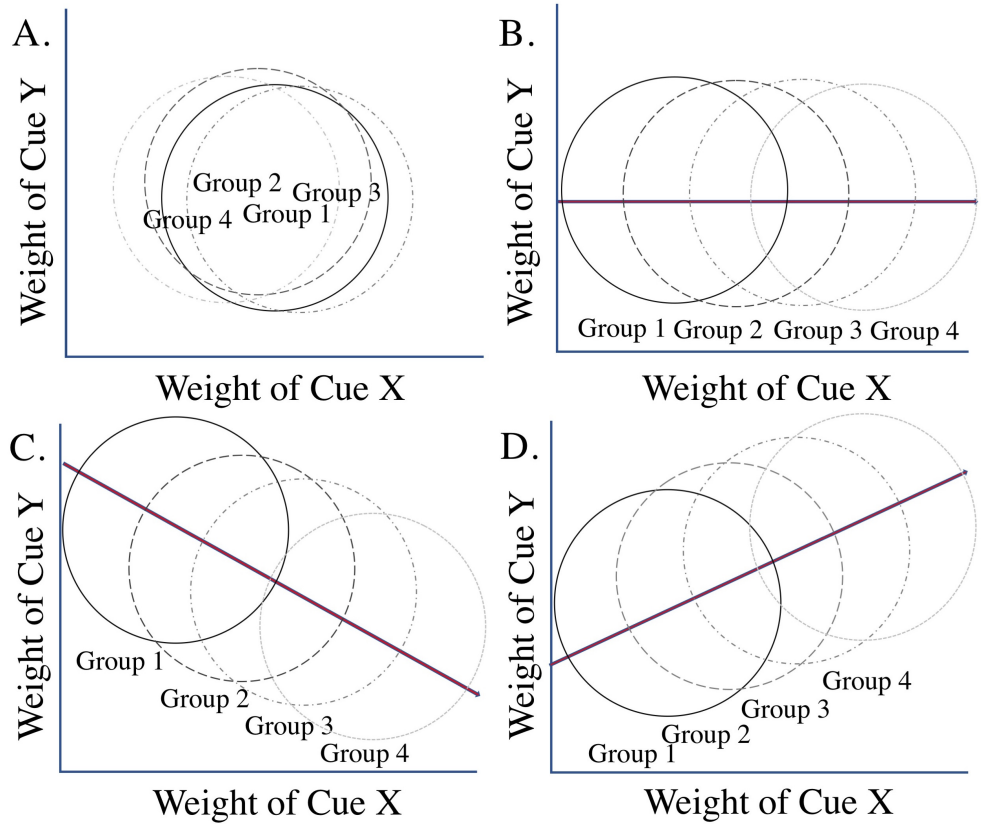


**Figure 3.1:** The four predictions on cue covariation. Each axis represents a cue weight which signals category distinctions (e.g. VOT and F0).

As reviewed above, Pattern C has been shown in earlier studies on Seoul Korean, where cue trade-offs between the use of VOT and f0 have been observed across speakers

of different ages and genders as well as in stops that are followed by vowels with different heights and words with different frequencies. If the change is progressing in the language through strengthening existing cue correlations due to language adaptivity, we would expect to see the same patterns in (at least) some of the conditions in languages that are not undergoing change, German and English. Synchronically, this cue covariation would mean that the value of one cue would predict that of the other and negatively correlate with it and the total of the informativity would be more or less maintained across groups.

If we observe any pattern other than C across groups for a given factor, we could tentatively conclude that trade-offs are a unique characteristic of transphonologization for that factor. Pattern A would suggest that cue weights do not covary in multiple dimensions and that each individual cue's weight does not vary across groups. Pattern B would indicate that the total informativity of the cues varies across groups due to between-group variability in one cue but not the other. Pattern D would suggest that both cues accumulate. As such, the value of one cue would predict that of the other and positively correlate with it. The total informativity would vary across groups accordingly. For both Patterns B and D, groups would differ along a hypo- and hyper-articulation dimension in terms of total cue informativity, but the patterns differ in whether groups differ in only one cue (B) or covary in both cues (D). Any of patterns A–D observed in this study will provide insight into how variability of multiple cues that signal phonetic categories is structured, and into the relationship between this variability and diachronic sound change.

## 3.3 Data and methods

### 3.3.1 Corpus data

The structure of VOT/f0 covariation across stop voicing categories is examined using speech corpora of three languages. In order to control for the effect of speech style such as acoustic differences between casual speech versus careful speech (e.g. Kang & Guion, 2008), our analysis considered read speech only. The details of each corpus are introduced below.

**English:** English tokens were extracted from the corpus of West Point Company G3 American English Speech, which contains short sentences read by 53 female and 56 male speakers. The data were collected from 2000 to 2001 by the U.S. Military Academy's Department of Foreign Languages (Morgan, LaRocca, Bellinger & Ruscelli, 2005). Among the list of 185 utterances that are designed to cover most possible American English syllable structures, each speaker read approximately 104 sentences. Each sound file lasts about 4 seconds. The corpus additionally includes recordings of 30 male speakers from an earlier project sponsored by the U.S. Army Research Laboratory where each speaker read 50 short sentences that contain military terms or traffic messages. All the recordings were made at a sampling rate of 22050 Hz. The current study analyzed the recordings from both projects.

**German:** The German data come from the PhonDat1 corpus (Draxler, 1995). The corpus contains a total of 21681 read sentences recorded by 201 different speakers (100 males and 101 females) at four different sites in Germany. Each speaker read a subcorpus of about 450 short sentences. The recording was done at a sampling rate of 48000 Hz and downsampled to 16000 Hz.

**Korean:** The Korean data are from The Speech Corpus of Reading-Style Standard Korean collected in 2003 (The National Institute of the Korean Language, 2005), the same corpus used in Study 1. The corpus contains recordings of essays and children's stories consisting of 930 sentences in total. The stories were read by 120 Seoul dialect speakers (60 males and 60 females) in the age range of aged 19–71 years old and recorded at the sampling rate of 16000 Hz. The version of the corpus used in the current study is force-aligned at the word and segment level (Yoon, 2015).

**Data:** The current study examines how the strengths (i.e. weights) of VOT and f0 that signal stop voicing contrasts vary as a function of word- and speaker-level variables. To this end, we extracted word-initial stops, read by most speakers in a given language, from as similar a prosodic context as possible. Only pre-tonic stops were extracted for stress languages (English and German) in order to avoid the effects of pitch accents on the vowels and target stops in stressed syllables, and we sought tokens that were from utterance-initial position as much as possible. However, there were not enough utterance-initial tokens, mainly due to aspects of the syntactic structure and phonotactic constraints of each language which affect the distribution of stops. Many German and English sentence-initial words are articles; most German articles begin with a short-lag stop; and English articles do not begin with a stop consonant. Korean word-initial stops are biased towards lenis stops and this bias is exaggerated for sentence-initial stops, because many sentence-initial words are conjunctions (e.g. 'Geureonde', 'Geurigo') beginning with a velar lenis stop. Therefore, token selection was extended to other positions beyond the beginning of sentences. For German, tokens were extracted from the first or second words of sentences. For English, there were not enough stop-initial words even when words that were second in the sentence were allowed, making it necessary to obtain

tokens from other positions. Given that each sentence in the English corpus is short enough to be comprised of only one intonational phrase, tokens from final and pre-final words were avoided in order to minimize the effect of final intonation rising or falling, which would affect f0. Korean stops were taken from intonational phrase-initial positions where a force-aligned pre-pause was longer than 30 ms, given that intonational phrases are very often signalled by some pause (Ladd, 1996). All intonational pitch contours were further auditorily checked. Note that we do not consider Korean tense stops in this study because they do not have an analogous stop category in German or English in terms of articulation, and are not undergoing sound change. The possible effect of sentence position on VOT and f0 was statistically controlled for the three datasets.

All the sound files of the selected tokens were checked by the first author for alignment errors, mispronunciations, and in case of Korean, also for intonational-phrase boundaries. All stops with voicing lead were discarded (German: 58; English: 86). The final dataset consist of 2660 tokens from 79 words and 118 speakers for German, 4208 tokens from 76 words and 126 speakers for English, and 5559 tokens from 60 words and 118 speakers for Korean.

### 3.3.2 Measurements

For each token in this dataset, we measured VOT and f0, and extracted information about place of articulation, following vowel height, word frequency, sentence position, and speaking rate. For VOT measurements, we used a semi-automatic method. First, VOT boundaries were generated by the software package 'AutoVOT' (Keshet, Sonderegger & Knowles, 2014) which uses a supervised learning algorithm trained on a small set of hand-annotated tokens (See Sonderegger & Keshet, 2012 for more detail). Subsequently,

all the automatically positioned boundaries were manually-checked and hand-corrected if misaligned. For both the training and final datasets, VOT onset was determined at the onset of the burst and VOT offset at the time of the first periodic cycle in the waveform that signals the onset of voicing.

For each token, f0 was extracted at 2 ms into vowel onset for English and German and at the vowel midpoint for Korean. We took vowel measures from different time points in different languages based on previous findings that consonant-induced f0 perturbation is generally limited to voicing onset for languages like English and German (Hombert, 1976), while the f0 distinction extends far beyond voicing onset as a result of the sound change in Seoul Korean (Jun, 1996; Silva, 2006).[2] The f0 values were first automatically taken using Praat's autocorrelation function with a 25 ms analysis window and f0 range set at 80–450 Hz. Then, we created histograms by gender and stop category for German and English and by gender, decade of birth, and stop category for Korean. We checked values at the histogram edges, and manually corrected pitch-tracking errors if necessary. All f0 measures were then normalized by converting hertz to semitones relative to each speaker's mean f0 to allow for comparison of f0 values across gender (Nolan, 2003). We describe measures extracted other than VOT and f0 in Sec. 3.4.1 below.

## 3.4 Results

We performed two separate analyses to examine the structure of variation in VOT and f0 across stop categories. In Analysis 1, we investigated whether and how VOT/f0 weights in production (co)vary across words, vowel contexts, and genders. We constructed two

[2]If the different points of f0 extraction between languages affect the results, it would be the case that this methodological choice renders weaker f0 effects in Korean, which would have the effect of *strengthening* the pattern we will actually see in the data (stronger effects in Korean than in English or German).

linear mixed effects regression models separately for VOT and f0 for each language. The same model structure was used for each cue in each language, for comparability of results across languages.

In Analysis 2, we examined how individuals and languages differ in the relative weights of VOT and f0, using three approaches: (1) comparing the degree of separation in means between stop categories in VOT with that of f0 using Cohen's $d$; (2) comparing the cue weights of VOT and f0 obtained using two classification methods, linear discriminant analysis (LDA) and support vector machines (SVM), that seek optimal weights for separating stop categories; (3) applying the same classification methods to both raw and 'normalized' data where the effects of the sources of variability on each cue are parsed out.

Before proceeding, we contextualize the analytical choices made in Analysis 2 which take into account that it is not clear how best to analyze VOT/f0 covariation, especially in corpus data. Our use of three approaches contrasts with most previous work comparing VOT and f0 weights, which has used LDA (Kirby, 2016; Schertz, Cho, Lotto & Warner, 2015; Shultz et al., 2012) or LDA and Cohen's $d$ (Clayards, 2018). LDA takes into account multiple parameters of the two bivariate distributions (e.g. means, standard deviations, covariance), corresponding to the two categories—which together determine the VOT and f0 weights. The motivation for employing multiple approaches in our study is to evaluate both the contribution of separation between the means, which estimates how well each cue in isolation discriminates stop categories (using $d$), and how well each cue discriminates stop categories relative to the other cue when both cues are simultaneously considered (using LDA and SVM). We use both LDA and SVM to minimize the possibility that our findigns are an artefact of assumptions of a given classification method, given that

the two methods differ in assumptions about the cue distributions, as described further below.

Second, in contrast to previous approaches which examined raw data, Analysis 2 investigates the structure of cue covariation at the speaker level (across gender and individual speakers) using both raw data and the residuals of regression models in which the effects of various factors on VOT and f0 (from Analysis 1) *besides* stop category are factored out. These two types of data correspond to two different ideas from the speech perception literature about how listeners process acoustic cues in categorization, given that these cues themselves (e.g. VOT) are affected by factors such as speaking rate (Summerfield, 1981), talker (e.g. Nearey, 1978), and neighbouring sounds (Liberman et al., 1974). There are different views on normalization, as extensively discussed by Magnuson & Nusbaum (2007). One view suggests that given the rich acoustic input in multiple dimensions, speech categories may be linearly separable in multiple acoustic dimensions, which may make normalizaton redundant (Nearey, 1997). Another view suggests that cues are subject to compensation or normalization prior to categorization (Cole et al., 2010; McMurray & Jongman, 2011; Summerfield, 1981) and therefore, a better model of speech categorization is computing linear boundaries on the data where context and talker information have been parsed out (see McMurray & Jongman, 2011 for more detail). It should be noted that our study does not directly address which perception model predicts better listeners' performance but rather aims to identify cue structures in multiple dimensions that listeners may have knowledge of and make use of in speech categorization. We perform the same analyses on the two types of data ('raw' and 'normalized') to check whether the patterns of structured cue variability we observe are robust to whether or not normalization is applied. We leave to future work a comparison of these models in

terms of actual prediction accuracy.

More details on each of these approaches and classification methods will be provided in the following sections.

### 3.4.1 Analysis 1: word frequency, vowel height, and gender

#### 3.4.1.1 Model structure

In Analysis 1, we constructed VOT and f0 models as a function of word frequency, vowel height, and gender using linear mixed-effects models, fitted using the `lmer` function from the `lme4` package (Bates et al., 2015) in R. In our models, we controlled for speaking rate, sentence position, and place of articulation. Both speaker-level and token-level speech rate were included as predictors in the models (Clayards, 2018; Stuart-Smith et al., 2015). These predictors separate the possible effect of a speaker's average speech rate from the effect of 'slower' and 'faster' speech within a given speaker, because both independently affect VOT (Sonderegger et al., 2017; Stuart-Smith et al., 2015) and possibly f0 as well. Sentence position was included in order to control for prosodic effects as utterance position is known to affect both VOT and f0 (Cho & Keating, 2001; Jun, 1998; Keating et al., 2003; Ladd, 1996). Finally, place of articulation was controlled for due to its robust effect on VOT, which tends to vary as velar > alveolar > bilabial (Cho & Ladefoged, 1999; Lisker & Abramson, 1964) cross-linguistically (but see Docherty & Foulkes 1999).

We fitted two identical models per language separately, using VOT and f0 as response variables. Using the same model structure across cues and languages allows us to assess the extent to which the size of VOT and f0 weights covary across words, phonetic contexts, speakers, and languages, without the confound of accounting for different variables for different cues/languages. We discuss the terms of the statistical models below. Terms

included in the model as fixed effects are indicated in SMALL CAPS.

- Stop category (CATEGORY) is a categorical variable and has two levels of 'long-lag' and 'short-lag' stops.

- Vowel height (HEIGHT) is a categorical variable that has two levels of 'high' and 'non-high' vowels. The levels were defined based on the phonemic transcription of the following vowel segment in each corpus.

- Log-transformed word frequency (FREQUENCY) is a continuous variable obtained from the log SUBTLEX frequency for English and German (English: Brysbaert & New 2009; German: Brysbaert, Buchmeier, Conrad, Jacobs, Bölte & Böhl 2011). For Korean, raw wordform frequency information was taken from the KAIST Concordance program (KAIST, 1999) and log-transformed.

- Place of articulation (POA) of the word-initial stop in stressed syllables is a categorical variable with three levels ('bilabial', 'alveolar', 'velar')

- Gender (GENDER) is a categorical variable with two levels ('female', 'male').

- Sentence position (POSITION) is a categorical variable with two levels, 'initial' and 'medial', depending on whether the stop occurs in sentence initial position.

- Speaker mean speaking rate (MEAN RATE) is a continuous variable. In order to obtain the values, we first calculated a token's speaking rate as syllables per second for each sound file that contains one short sentence, then calculated the average value of each speaker.

- The difference between the speaker mean rate and a token's speaking rate (RATE DEV) was calculated and included in the models as a continuous variable.

Before being fitted to the models, CATEGORY, HEIGHT, GENDER, and POSITION were coded using sum contrasts (CATEGORY: short < long; HEIGHT: non-high < high; GENDER: female < male; POSITION: initial < medial). POA was coded using Helmert contrasts, corresponding to labial vs. non-labial stops (POA1) and alveolar vs. velar stops (POA2). Three continuous variables, FREQUENCY, MEAN RATE, and RATE DEV were standardized (centered and divided by two standard deviations). All these variables were included as main-effect terms. Interaction terms were chosen to address our research questions with regard to how word frequency, vowel height, and gender affect the contrast size (or strength) of each cue relative to the other in contrasting stop voicing categories. The interaction of CATEGORY with RATE DEV was included to control for its potential effect based on the previous findings that faster speech decreases VOT for stops with long-lag VOTs, while its effect on short-lag categories is small or absent (Kessinger & Blumstein, 1997; Miller et al., 1986).

The models were constructed with "maximal" random effect structure (Barr et al., 2013) by including all possible by-word and by-speaker random intercepts and slopes in the models. This way, our models account for variability across speakers and words in VOT and f0 beyond the effects of variables included in the models, as well as variability across speakers and words in the effects of these variables. We did not include correlations between random-effect terms, to facilitate model convergence.

#### 3.4.1.2 Results

The goal of this analysis is to examine the extent to which word frequency, vowel height, and gender affect the directions and strengths of the VOT and f0 contrast, across three languages. Therefore, more attention is paid to how HEIGHT, FREQUENCY, and GENDER

**Table 3.1:** English: Summary of all fixed-effect coefficients for the models of logVOT (left) and f0 (right); coefficient estimates, standard errors, degrees of freedom (df), $t$-values, and significances.

| | VOT | | | | | f0 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| FULL MODELS | Estimate | SE | df | $t$ | $P(>t)$ | Estimate | SE | df | $t$ | $P(>t)$ |
| Intercept | 38.492 | 0.915 | 112.146 | 42.07 | $< 0.001$ | -0.021 | 0.092 | 63.884 | -0.224 | 0.823 |
| CATEGORY (short-lag vs. long-lag) | 38.081 | 1.658 | 75.069 | 22.963 | $< 0.001$ | 1.17 | 0.189 | 66.12 | 6.183 | $< 0.001$ |
| HEIGHT (nonhigh vs. high) | 0.553 | 2.585 | 57.446 | 0.214 | 0.831 | 1.169 | 0.317 | 66.04 | 3.69 | $< 0.001$ |
| FREQUENCY | 0.178 | 1.396 | 59.8 | 0.128 | 0.899 | 0.078 | 0.168 | 65.992 | 0.467 | 0.642 |
| GENDER (female vs. male) | -0.716 | 1.394 | 161.018 | -0.514 | 0.608 | -0.009 | 0.054 | 46.041 | -0.165 | 0.87 |
| RATE DEV. | -1.661 | 0.469 | 328.726 | -3.539 | $< 0.001$ | -0.04 | 0.049 | 374.171 | -0.82 | 0.413 |
| POSITION (initial vs. medial) | 3.002 | 1.927 | 259.793 | 1.558 | 0.12 | -0.389 | 0.195 | 352.125 | -1.989 | 0.048 |
| PLACE1 (labial vs. non-labial) | 7.28 | 1.603 | 57.57 | 4.542 | $< 0.001$ | 0.057 | 0.194 | 64.01 | 0.296 | 0.768 |
| PLACE2 (alveolar vs. velar) | 6.824 | 1.961 | 58.772 | 3.479 | 0.001 | 0.002 | 0.237 | 65.002 | 0.009 | 0.993 |
| MEAN RATE | -2.392 | 1.236 | 126.733 | -1.935 | 0.055 | -0.035 | 0.05 | 60.382 | -0.703 | 0.485 |
| CATEGORY:HEIGHT | -0.588 | 4.949 | 55.648 | -0.119 | 0.906 | 1.454 | 0.606 | 64.155 | 2.399 | 0.019 |
| CATEGORY:FREQUENCY | 0.693 | 2.826 | 61.137 | 0.245 | 0.807 | -0.535 | 0.341 | 67.031 | -1.57 | 0.121 |
| CATEGORY:GENDER | -4.378 | 2 | 148.544 | -2.189 | 0.03 | 0.31 | 0.131 | 70.34 | 2.36 | 0.021 |
| CATEGORY:RATE DEV. | -2.495 | 0.948 | 3385.832 | -2.631 | 0.009 | -0.116 | 0.099 | 305.014 | -1.171 | 0.242 |

terms interact with CATEGORY than other terms in reporting the results. We first present the results from English and German, then discuss the results for these languages in comparison with the results from Korean from our previous study (Bang, in press, also Chapter 2 of this thesis) as summarized in Sec. 3.2.2. This comparision will allow us to identify the relationship between synchronic variability and diachronic change in the use of multiple cues to stop voicing categories.

Table 3.1 and Table 3.2 summarize the fixed-effect coefficients for both VOT and f0 models. Coefficient significances were computed using the Satterthwaite approximation as implemented in the `lmerTest` package (Kuznetsova, Brockhoff & Christensen, 2015).

#### 3.4.1.2.1 English

**VOT:** There is a significant main effect of CATEGORY, indicating that long-lag stops have longer VOTs than short-lag stops, averaged across all speakers and other variables ($\hat{\beta}$ = 38.081, $p < 0.001$). We did not find strong evidence that VOT differs by vowel height, sentence position, gender, and words with different frequency ($p > 0.12$). On the other hand, slower speakers (MEAN RATE: $\hat{\beta}$ = -2.392, $p = 0.055$) and slower speech

(RATE DEV: $\hat{\beta}$ = -1.661, $p < 0.001$) result in longer VOTs. For the effects of place of articulation on VOT, our result is consistent with the typical cross-linguistic order of bilabial < alveolar < velar (POA1: $\hat{\beta}$ = 7.28, $p < 0.001$; POA2: $\hat{\beta}$ = 6.824, $p = 0.001$).

Turning to the interaction effects with CATEGORY, which are crucial terms to address our research questions, we did not find strong evidence that vowel height and word frequency ($p > 0.807$) modulate the size of the VOT contrast in Engish. However, there is a significant gender effect on VOT contrast size, which is greater in female speech than in male speech ($\hat{\beta}$ = -4.378, $p = 0.03$). As speakers speak faster than their average rate, VOT contrast size decreases ($\hat{\beta}$ = -2.495, $p = 0.009$)—as expected given the asymmetric effect of speech rate on VOT for voiced and voiceless stops observed in previous work.

**f0:** There is a main effect of category (CATEGORY: $\hat{\beta}$ = 1.17, $p < 0.001$), indicating that voiceless stops have higher f0s than voiced stops averaged across speakers and other variables, as expected given previous work on onset f0 effects as reviewed in Sec. 3.2.1. f0 is higher when the following vowel is a high vowel than a low vowel averaged across stop categories, speakers, and other variables (HEIGHT: $\hat{\beta}$ = 1.169, $p < 0.001$), consistent with the well-documented 'intrinsic' effect of vowel height on pitch (e.g. Whalen & Levitt, 1995). Stops extracted from sentence-medial positions have on average lower f0 values compared to sentence-initial stops ($\hat{\beta}$ = -0.389, $p = 0.048$), probably due to declination of f0 over the sentence. We did not find strong evidence that f0 values (averaged across categories) are modulated by word frequency, gender, and speaking rate at either the speaker or token levels ($p > 0.413$).

On the other hand, the size of the f0 contrast across categories significantly differs by vowel height (CATEGORY:HEIGHT, $\hat{\beta}$ = 1.169, $p < 0.001$), indicating that the f0 contrast is greater for stops that precede a high vowel. Our results show a trend in the

expected direction that the size of the f0 contrast is reduced as word frequency increases (CATEGORY:FREQUENCY: $\hat{\beta}$ = -0.535, $p$ = 0.121). There is a significant effect of gender: male speakers distinguish the stop categories with greater f0 distinctions than female speakers (CATEGORY:GENDER: $\hat{\beta}$ = 0.31, $p$ = 0.021). The size of the f0 contrast does not significantly differ due to faster speech relative to speakers' mean speaking rate ($p$ = 0.413).

#### 3.4.1.2.2 German

**VOT:** There is a main effect of CATEGORY, indicating that German long-lag stops have longer VOTs than short-lag stops, averaged across speakers and other variables ($\hat{\beta}$ = 36.459, $p < 0.001$). Stops that precede a high vowel have longer VOTs than those that precede a non-high vowel (HEIGHT: $\hat{\beta}$ = 4.013, $p$ = 0.001), consistent with the cross-linguistically common pattern that VOTs are longer before a high vowel than before a low vowel (Higgins et al., 1998; Klatt, 1975; Weismer, 1979). High frequency words have shorter VOTs when averaged across categories (FREQUENCY: $\hat{\beta}$ = -2.621, $p$ = 0.016). Faster speech and sentence medial stops also have shorter VOTs compared to slower speech and sentence initial stops (RATE DEV.: $\hat{\beta}$ = -8.659, $p < 0.001$; POSITION: $\hat{\beta}$ = -8.659, $p < 0.001$). For the effect of place of articulation, nonlabial stops have longer VOTs than labial stops and velar stops have longer VOTs than alveolar stops (POA1: $\hat{\beta}$ = 8.433, $p < 0.001$, POA2: $\hat{\beta}$ = 5.663, $p < 0.001$), in line with the cross-linguistically common pattern. We did not find strong evidence that speaker's mean speaking rate and gender affect VOT values across stop categories ($p > 0.144$).

Turning to the size of cue contrasts, which is of primay interest: the VOT contrast is greater in high vowel contexts than in non-high vowel contexts (CATEGORY:HEIGHT:

**Table 3.2:** German: Summary of all fixed-effect coefficients for the models of logVOT (left) and f0 (right); coefficient estimates, standard errors, degrees of freedom (df), $t$-values, and significances.

| | VOT | | | | | f0 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| FULL MODELS | Estimate | SE | df | $t$ | $P(>t)$ | Estimate | SE | df | $t$ | $P(>t)$ |
| Intercept | 26.358 | 0.79 | 147.455 | 33.363 | $< 0.001$ | -0.002 | 0.139 | 69.747 | -0.015 | 0.988 |
| CATEGORY (short-lag vs. long-lag) | 36.459 | 1.525 | 132.688 | 23.911 | $< 0.001$ | 1.221 | 0.29 | 73.171 | 4.207 | $< 0.001$ |
| HEIGHT (nonhigh vs. high) | 4.013 | 1.218 | 75.971 | 3.295 | 0.001 | 1.627 | 0.279 | 73.56 | 5.835 | $< 0.001$ |
| FREQUENCY | -2.621 | 1.067 | 74.114 | -2.456 | 0.016 | 0.213 | 0.252 | 75.185 | 0.842 | 0.403 |
| GENDER (female vs. male) | -0.583 | 0.909 | 158.259 | -0.642 | 0.522 | 0.045 | 0.078 | 54.964 | 0.57 | 0.571 |
| RATE DEV. | -3.138 | 0.904 | 373.057 | -3.47 | 0.001 | 0.191 | 0.175 | 421.894 | 1.091 | 0.276 |
| POSITION (initial vs. medial) | -2.137 | 0.858 | 203.113 | -2.49 | 0.014 | 0.537 | 0.179 | 389.169 | 3.003 | 0.003 |
| PLACE1 (labial vs. non-labial) | 8.433 | 1.267 | 69.985 | 6.657 | $< 0.001$ | 0.434 | 0.294 | 67.133 | 1.475 | 0.145 |
| PLACE2 (alveolar vs. velar) | 5.663 | 1.425 | 71.682 | 3.973 | $< 0.001$ | 0.358 | 0.33 | 67.802 | 1.085 | 0.282 |
| MEAN RATE | -1.59 | 1.08 | 128.895 | -1.471 | 0.144 | -0.221 | 0.08 | 38.39 | -2.744 | 0.009 |
| CATEGORY:HEIGHT | 9.054 | 2.69 | 88.43 | 3.366 | 0.001 | -0.508 | 0.582 | 75.041 | -0.872 | 0.386 |
| CATEGORY:FREQUENCY | -8.659 | 2.356 | 93.55 | -3.676 | $< 0.001$ | -0.1 | 0.506 | 77.986 | -0.197 | 0.844 |
| CATEGORY:GENDER | -5.132 | 1.718 | 122.898 | -2.988 | 0.003 | 0.384 | 0.18 | 64.22 | 2.125 | 0.037 |
| CATEGORY:RATE DEV. | -9.259 | 2.136 | 344.328 | -4.336 | $< 0.001$ | -0.167 | 0.295 | 552.566 | -0.567 | 0.571 |

$\hat{\beta} = 9.054$, $p = 0.001$). Higher frequency words are produced with smaller contrast sizes than lower frequency words (CATEGORY:FREQUENCY: $\hat{\beta} = -8.659$, $p < 0.001$). There is also a significant gender effect on the contrast size; female speakers distinguish the stop categories with greater VOT distinctions than male speakers ($\hat{\beta} = -5.132$, $p = 0.003$). The effect of speaking rate shows that as speakers speak faster than their average speaking rate, the size of the VOT contrast becomes smaller (CATEGORY:RATE DEV: $\hat{\beta} = -9.259$, $p < 0.001$), which is expected given previous work (as discussed for the English data).

**f0:** Voiceless stops have higher f0 than voiced stops, averaging across speakers and other variables (CATEGORY: $\hat{\beta} = 1.221$, $p < 0.001$), which is expected onset f0 effect. A cross-linguistically common pattern is also observed for the effect of vowel height: high vowels on average have higher f0s than non-high vowels (HEIGHT: $\hat{\beta} = 1.627$, $p < 0.001$). F0 is higher for the vowels that follow a sentence medial stop than for the ones that follow a sentence initial stop (POSITION: $\hat{\beta} = 0.537$, $p = 0.003$), probably because German sentence initial stops are mostly function words which do not carry prosodic prominence. There is a main effect of MEAN RATE, indicating that speakers who speak faster than others produce the stops with lower f0 (MEAN RATE: $\hat{\beta} = -0.221$, $p = 0.009$).

Word frequency, gender, faster speech, and place of articulation did not significanlty modulate f0 values ($p > 0.145$).

Concerning interaction effects, word frequency, vowel height, and faster speech did not significantly modulate the size of the f0 contrast ($p > 0.386$). However, there is a strong effect of gender, indicating that male speakers produce the stops with greater f0 distinctions than female speakers ($\hat{\beta} = 0.384$, $p = 0.037$).

#### 3.4.1.2.3 Across English, German, and Korean

So far, we have examined how the variability of VOT and f0 that signal stop voicing categories is structured across words, vowel contexts, and speaker genders in German and English. In this section, we focus on cross-lingusitic similarities and differences by comparing the results from the two languages, and further by comparing the current findings to those from our previous study on Korean sound change (Bang. et al, in press). By doing so, we gain insight into the relationship between synchronic variability and diachronic change in the way multiple cues signal phonetic categories.

The results for the three languages are summarized and compared in Table 3.3. The top section shows how each cue is modulated by a given source of variability when the stops are *averaged* across categories, and the bottom section shows how the sizes of the cue contrasts *across* stop voicing categories are modulated by the same variables.

Concerning the values averaged across categories, for English stops, f0 but not VOT is greater when the following vowel is a high vowel compared to a non-high vowel. For German and Korean stops, both VOT and f0 are greater in high vowel contexts. These cross-linguistic findings on the effects of vowel height on the cue values are consistent with a long literature showing that high vowels are associated with greater VOTs and

**Table 3.3:** The effects of vowel height, words frequency, and gender on cue values averaged across categories and contrast sizes used to signal stop categories. The top section is main effects of the factors, while the bottom is of their interactions with laryngeal category. Asterisks indicate effects with $0.05 < p < 0.15$ whose directions are of interest, despite not reaching significance ($p < 0.05$).

| Average value | English | | German | | Korean | |
|---|---|---|---|---|---|---|
| | VOT | f0 | VOT | f0 | VOT | f0 |
| height | – | nonhigh < high | nonhigh < high | nonhigh < high | nonhigh < high | nonhigh < high |
| frequency | – | – | – | – | low > high* | low > high |
| gender | – | – | – | – | male > female | male < female |
| Contrast size | | | | | | |
| height | – | nonhigh < high | nonhigh < high | – | nonhigh < high | nonhigh > high |
| frequency | – | low > high* | low > high | – | low > high* | low < high* |
| gender | male < female | male > female | male < female | male > female | male > female | male < female |

higher f0s (Higgins et al., 1998; Honda, 1983; Honda et al., 1994; Klatt, 1975; Sapir, 1989; Weismer, 1979).

On the other hand, word frequency does not significantly modulate VOT and f0 values averaged across stop categories in English and German. This result is in contrast with the results from Seoul Korean where low frequency words have longer VOTs and higher f0s compared to high frequency words. While no significant gender difference was found in English and German, in Korean, males have greater VOTs and smaller f0s than females.

Turning to the size of the cue contrasts, the focus of our study: we did not find strong effects of vowel height on VOT in English or on f0 in German. However, English stops are contrasted with bigger f0 differences, and German stops with greater VOT differences in high vowel contexts than in low vowel contexts. These cross-linguistic differences in vowel effects are illustrated in Figure 3.2. These patterns in German and English suggest that the amount of informativity in the two cue dimensions is smaller in the nonhigh vowel context for both languages due to the noticeable variability in one cue dimension, matching our Pattern B (Figure 3.1-B). Crucially, the synchronic variability observed in German and English is different from the diachronic pattern found in Seoul Korean where a tradeoff between the use of VOT and f0 across vowel contexts is observed—our Pattern

C (Figure 3.1-C).

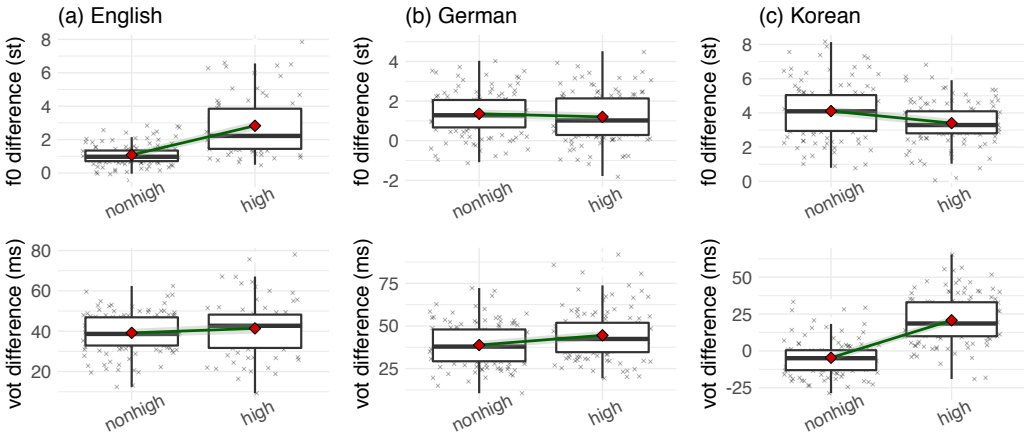


**Figure 3.2:** Speaker mean differences between long- and short-lag stops in f0 (top) and VOT (bottom) across vowel height contexts. Shading around the lines indicates 95 % confidence intervals.

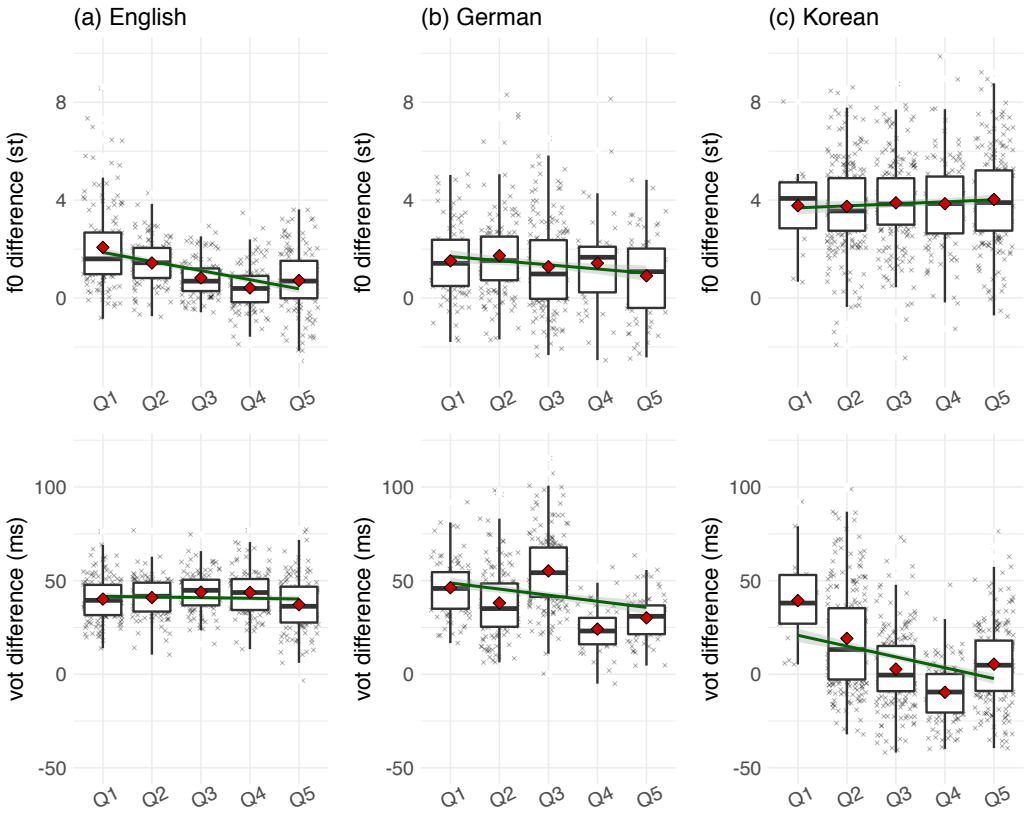


**Figure 3.3:** Speakers' mean f0 differences (top) and VOT differences (bottom) between long- and short-lag stops across words with different frequencies. Each point represents a speaker and Q1-Q5 refer to word frequency quantiles from lowest (Q1) to highest (Q5). Shading around the lines indicates 95 % confidence intervals.

Notably, high frequency words and low frequency words behave similarly to nonhigh

vowel contexts and high vowel contexts respectively. As was the case in the effects of vowel height, we did not find strong evidence that the size of the VOT contrast in English stops and the size of the f0 contrast in German stops differ across words with different frequency. However, the size of the f0 contrast in English and the size of VOT contrast in German tend to decrease with increasing word frequency as shown in (a) and (b) in Figure 3.3. This finding suggests that the overall informativity of the cues is reduced as word frequency increases, showing Pattern B (Figure 3.1-B). For Korean stops, on the other hand, the direction of the VOT contrast is negative and that of the f0 contrast is positive with increasing word frequency, exhibiting a trade-off between the two cues (Figure 3.1-C). This effect is shown in Figure 3.3-(c): as word frequency increases from Q1 to Q5, VOT difference decreases while f0 difference increases.

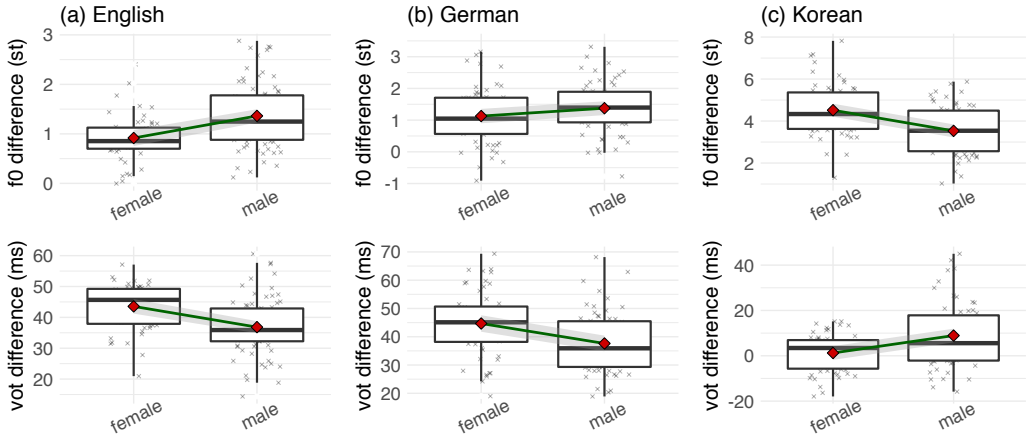


**Figure 3.4:** Speakers' mean f0 differences between long-and short- lag stops (top) and vot differences (bottom) between long- and short-lag stops across gender. Each point represents a speaker. Shadings indicate 95 % confidence intervals.

Compared to word frequency and vowel height, the effects of gender on contrast size shown in Figure 3.4 are clearer—significant for each cue in each language. For English and German, female spekaers produce greater VOTs and smaller f0s to cue stop categories than male speakers. However, for Korean, female speakers produce greater f0s and smaller VOTs, consistent with the finding that female speakers lead the ongoing sound change

(Kang, 2014). The gender effects suggest that cross-linguistically, there are trade-offs between the use of VOT and use of f0 between male and female speakers, following the pattern in Figure 3.1-C.

#### 3.4.1.3 Discussion

In this analysis, we have examined whether and how the sizes of the VOT contrast and f0 contrast that signal stop voicing are correlated across words and vowel contexts in the speech signal. While doing so, we focused on cross-linguistic similarities and differences in the structure of cue variability. Our results so far suggest that for the languages not undergoing sound change, cues may not be equally affected by the sources of variability. Furthermore, which cue is more susceptible to the variability relative to other cues may be language-specific: f0 in English and VOT in German. In German and English, the total cue informativity that signals stop voicing categories is greater in high vowel contexts and low frequency words than non-high vowel contexts and high frequency words, in line with Pattern B in Figure 3.1.[3]

These findings in German and English are different from the patterns observed in the language undergoing sound change. In Seoul Korean, both of the cues are modulated across the linguistic conditions such that if the strength of one cue is reduced in one condition then that of the other is enhanced for compensation, showing cue trade-offs as illustrated in (c) in Figure 3.1. As a result, in Seoul Korean, total cue informativity is maintained across these conditions during the change.

[3]Note that caution is needed in interpreting the null results ($0.05 < p < 0.15$) in English VOT and German f0. The null results in these cues may be due to a lack of power, rather than there not being an effect in reality. We are assuming here that the relevant effects are not *zero* (which is likely), and that our models have at least determined the correct effect *direction*. Future work with increased sample size should confrim our interpretation of the cross-linguistic differences. However, one thing that is crucial in our findings is that the pattern observed in Seoul Korean is different from English and German. There are clear cue trade-offs across the linguistic factors in Seoul Korean while clear trade-offs are not observed in the other languages.

Regarding gender, a crucial cross-linguistic pattern was observed; in all three languages, VOT and f0 contrasts *trade off* across gender (Figure 3.4), consistent with Figure 3.1-C. This gender effect may imply soico-phonetically motivated variation in the use of multiple cues: pronunciation variation signaling the social group(s) that a speaker identifies him- or herself with (e.g. male speech versus female speech). We discuss this possibility in more detail in Sec. 3.5.

In what follows, we test this prediction regarding individual speakers' use of VOT and f0 for stop voicing categorization while again focusing on cross-linguistic similarities and differences in the relative weights of the cues.

### 3.4.2 Analysis 2: speakers across languages

In this section, we turn to examining whether and how individual speakers differ in the use of VOT and f0 in categorizing stop voicing and what cross-linguistic similarities and differences are. While we compare results across languages, we focus in particular on the relevance of our results for the relationship between synchronic variability and diachronic change.

Speaker differences in the use of VOT and f0 were investigated based on two approaches, which make use of three methods for quantifying cue strengths. In the first approach, we estimated the relative strengths of cues by calculating how well each cue, in isolation, discriminates stop categories. Following Clayards (2018), we estimated cue strengths using Cohen's $d$ values computed for each cue and each speaker, as described further in Sec. 3.4.2.2.

The second approach involves computing optimal classification boundaries that separate stop categories in multiple acoustic dimensions. Within a language, two linear

classification procedures, linear discriminant analysis (LDA) and linear support vector machine (SVM) were performed for each speaker to estimate each cue's strength relative to another in stop voicing classification. 'Linear' here means that the boundary separating the two categories is assumed to be a linear combination of cues (i.e. a line in VOT/F0 space). LDA and SVM in particular were chosen because they make different underlying assumptions, which gives confidence that our results are not an artefact of assumptions of a particular classification method. More details of these two methods will be discussed in Sec. 3.4.2.3.

These three methods ($d$, LDA, SVM) were performed on both raw data and the residuals of mixed effects models where VOT and f0 values were predicted by the same variables controlled for in Analysis 1. This method should allow us to examine whether and how strongly the strengths of VOT and f0 covary across speakers in raw multidimensional cue space (Nearey, 1997), and in multidimensional 'residualized' cue space (McMurray & Jongman, 2011), where variability has been reduced.

#### 3.4.2.1 Dataset construction

To obtain residualized data, we first constructed mixed effects regression models in which each VOT and f0 value was predicted from:

1. The variables (except for CATEGORY) introduced in Analysis 1, which include vowel height, word frequency, gender, sentence position, place of articulation, and two speaking rate effects (at speaker- and utterance-level).

2. Interaction terms between CATEGORY and vowel height, word frequency, gender, and (utterance-level) speaking rate.

3. A by-speaker random intercept, by-word random intercept, and all possible by-

speaker random slopes corresponding to (1) and (2).

These terms parse out all factors *except* voicing category affecting VOT and f0. We expect that after this procedure, the information in the residuals is limited to variation that can be predicted by how speakers differ in signaling stop contrasts, as well as residual error.[4] After parsing out these sources of variability in VOT and f0 values, we computed measures of cue strength for stop categorization (predicting CATEGORY) in exactly the same way as for raw VOT and f0, but using the residualized cues, which we denote $\text{VOT}_r$ and $\text{f0}_r$.

#### 3.4.2.2 Contrast separability: $d$

First, we assess cue strengths using Cohen's $d$, which estimates the strengths of cues by comparing differences in category means normalized by the variance pooled across categories. Hence, if two speakers have same values in category mean differences for a given cue while they differ in its variance, the speaker with a smaller variance will have a higher $d$ than the other speaker. Also, if two speakers differ in mean differences but they have the same variance, $d$ will be larger for the speaker with a greater mean difference. $d$ as a measure of cue strength in production shares an important property with speech perception, in that listeners' degree of uncertainty in categorization is proportional to the variance of the cue when difference in category means is held constant (Clayards, Tanenhaus, Aslin & Jacobs, 2008).

$d$ values were estimated by calculating the mean difference between two stop categories and dividing the difference by the pooled standard deviation, using the denominator in

[4]There are known statistical issues with partialing out variables and using residuals as the variable of interest, causing the analyst to either find spurious effects or miss effects, compared to non-residualized data (Wurm & Fisicaro, 2014). As we will see below, the results of our analyses using residualized and non-residualized data are qualitatively similar, suggesting these statistical issues do not strongly influence our results. In addition, residualized cue values here actually correspond to a theoretical position (McMurray & Jongman, 2011), which motivates their use in our analyses.

the formula below (Cohen, 1988). Among various methods introduced to obtain pooled standard deviation for $d$ calculation (Lakens, 2013), the one used here is the version with Bessel's correction for bias in the estimate of data variance:

$$d = \frac{M_p - M_b}{\sqrt{\frac{(n_p-1)SD_p^2+(n_b-1)SD_b^2}{n_p+n_b-2}}} \tag{3.1}$$

where $M_p$ and $M_b$ refer to the means of the long-lag and short-lag categories respectively, $SD_p$ and $SD_b$ refer to the standard deviations of the two categories, and $n_p$ and $n_b$ are the number of observations in each stop category.

This correction should improve the estimate of the degree of overlap for our data, given dissimilar sample sizes across stop categories and speakers, which commonly occurs in unbalanced corpus data. Without correcting this bias, the estimate of $d$ may be inflated for speakers with a smaller sample size, due to the smaller pooled standard deviation associated with the smaller sample size. By weighting each category's standard deviation by the sample size, as in Equation (3.1), this upward bias will be reduced.

Using this formula, we obtain $d$ values separately for each cue and for each speaker and then measure how the strengths of VOT and f0 covary across speakers within each language using Pearson correlation coefficients ($r$).

When performed on raw data, negative correlations in $d$ values between VOT and f0 are significant across English speakers ($r = -0.24$, $p = 0.006$) and a stronger correlation is found across Korean speakers ($r = -0.5$, $p < .001$). For German speakers, even though the direction is the same as other languages, no significant correlation is observed ($r = -0.03$, $p = 0.7$). A cross-linguistic difference is apparent in terms of the primacy of cues. Figure 3.5 shows that for all German speakers and most English speakers, the $d$ values are much greater for VOT than f0 with greater speaker variability in VOT than

in f0. By contrast, in Korean the majority of speakers have $d$ values much higher for f0 than VOT and speaker variability is much greater in f0 than in VOT.

Korean and English still show significant negative correlations even when $d$ values are calculated using the residualized data. However, while the correlation has a similar value in English ($r = -0.21$, $p = 0.01$), in Korean the correlation becomes weaker in the residualized data ($r = -0.36$, $p < .001$) compared to the raw data. One interesting pattern found in the residualized data is that after other sources of variability on VOT and f0 are accounted for, the range of $d$ values across Korean speakers become greatly reduced for f0, making the distribution of $d$ in the f0 dimension more comparable across the three languages. The reduction in f0 variability could be explained by our previous finding involving the same Korean data (Bang et al., in press) that speakers differ in the vowel height effect on the size of the f0 contrast, as a function of how far along they are in the sound change. Therefore, when by-speaker random slopes for HEIGHT:CATEGORY are partialled out, the difference across speakers in f0 weights decreases.

Another interesting cross-linguistic difference is observed in the values of lower bounds of confidence intervals (CIs) of $d$ values, which tell us how many speakers in a language place very little weight on a given cue. For German and English speakers, almost none of the lower bounds of CIs overlap with zero for VOT while many of the Korean speakers' lower bounds do overlap with zero, indicating that unlike German and English speakers, some Seoul Korean speakers give little or no weight to VOT. On the other hand, the CIs overlap zero for f0 but not for VOT for English and German, and VOT but not f0 for Korean, indicating that for most Seoul Korean speakers, f0 has a higher weight compared to English/German speakers.

The results from Cohen's $d$ in English and Korean provide evidence that if a speaker

produces stop contrasts with less overlap in one cue, the speaker is likely to contrast the category with more overlap in another cue, exhibiting a trading relation in the use of cues to contrast stops. Further, this trading relation is stronger among Korean speakers who are undergoing change in cue primacy. However, for German it was not clear whether the same type of correlation exists.

So far, we have assessed how speakers and languages differ in the use of VOT and f0 in stop categorization by comparing measures of separability of a *single* cue (i.e. $d$) for VOT and f0, across speakers, where the measure for each cue was obtained independently of the other cue. In the next section, we address the same question using classification methods in multi-dimensional cue space by computing the strength of each cue relative to the other in forming a linear boundary that results in maximal category separability. Given that listeners perform categorization using multiple cues, the result of a classification method in multi-dimensional cue space may better represent human language processing than a method ($d$) which considers each cue independently.

#### 3.4.2.3 Classification: LDA and SVM

Applied to speech production data, each classification method finds a linear combination of cues (here, a line in VOT/F0 space) that results in optimal category separation. However, what this 'optimal' line is differs between classification methods, because each method makes different assumptions and applies a different algorithm. To reduce the risk that our results are an artefact of the details of a particular classification method, rather than representing a robust generalization that can be made based on the data, we perform the same analysis using two classification methods, LDA and SVM, that make different assumptions and determine the 'optimal' category boundary in different ways.

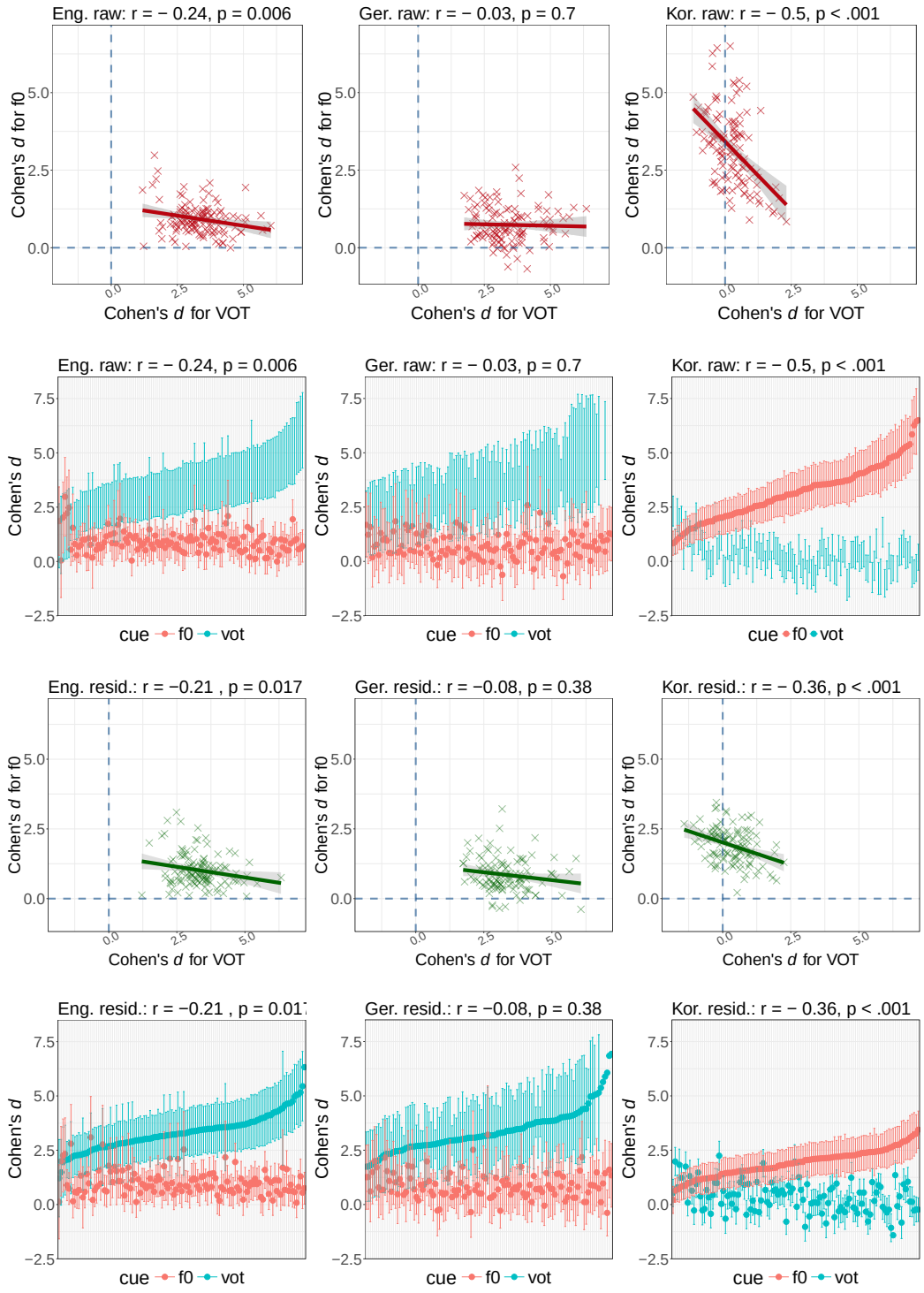


**Figure 3.5:** Cohen's $d$ values plotted for VOT against f0 across speakers. Top row: values computed on raw data; Bottom row: values computed on residualized data.

Applied to data from two stop categories in two-dimensional cue space, both LDA and SVM give linear boundaries that best separate data points belonging to the two categories. To achieve adequate discrimination between categories, LDA chooses the boundary that

gives the highest ratio of between-group variance to within-group variance. This turns out to mean that LDA performs best when the two category distributions are (multivariate) normal, with similar covariance matrices. When these assumptions are not met, the decision boundary is prone to being influenced by outliers (Croux & Joossens, 2005).

Based on visual inspection of our data, these two assumptions may not be met when applying LDA to data from individual speakers—as done here, and in previous work (Clayards, 2018; Kirby, 2016; Shultz et al., 2012). At least for some speakers, the covariance matrices for the long-lag and short-lag categories are quite different, which may cause bias when the pooled covariance estimate is used. The bottom panels of Figure 3.6 illustrate the problem: for this speaker with relatively few data points, the covariance matrices look very different, and are strongly affected by outliers, which likely strongly affects the location of the decision boundary.

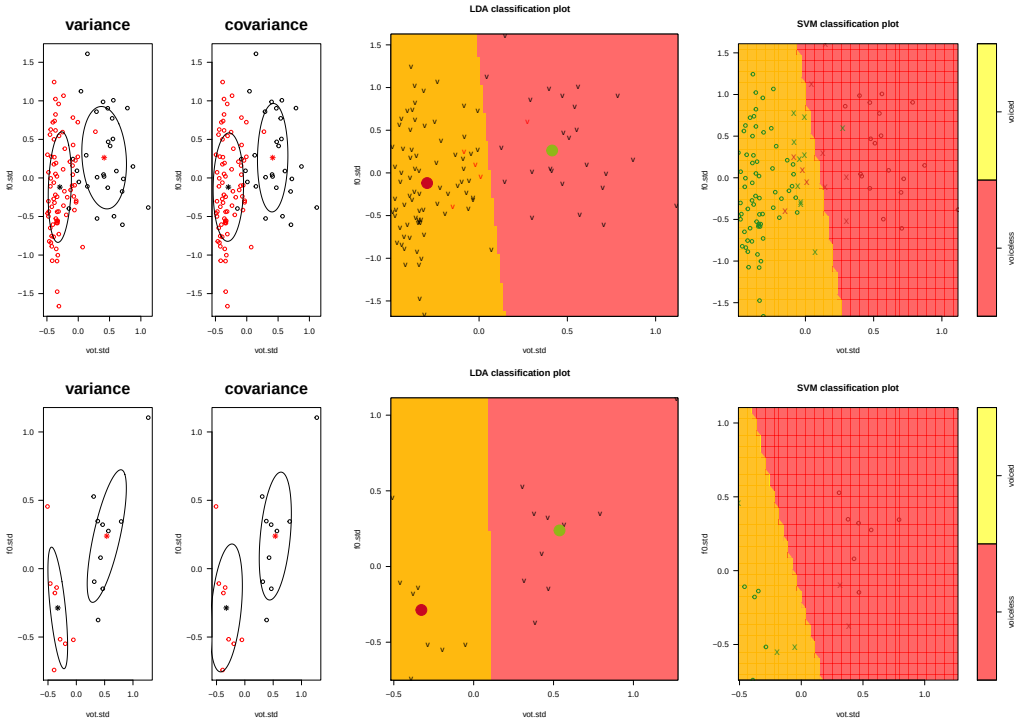


**Figure 3.6:** Data from two German speakers (top and bottom). From left to right: Covariance of VOT and f0 calculated separately for stop categories; pooled covariance; classification boundaries computed from LDA; classification boundaries computed from SVM.

On the other hand, SVM does not make particular assumptions about the distribution of data, making it a very flexible method. SVM computes an optimal linear classification boundary over a subset of the data (i.e. support vectors) that lie on the separating 'margin'. Unlike LDA, which searches for boundaries that maximize the ratio of between- to within-category variances, SVM searches for a boundary where the margin is maximized. These two classification methods are visualized in Figure 3.6 where the data points used as support vectors near the categorization boundary are indicated as 'X'. The optimal boundary obtained from SVM is less affected by sample size or outliers, compared to LDA.

If similar results are obtained from the two classification methods, we have more confidence in the computed coefficients (cue weights) in our data, and may be able to resolve the inconsistency of previous results for English data using LDA.

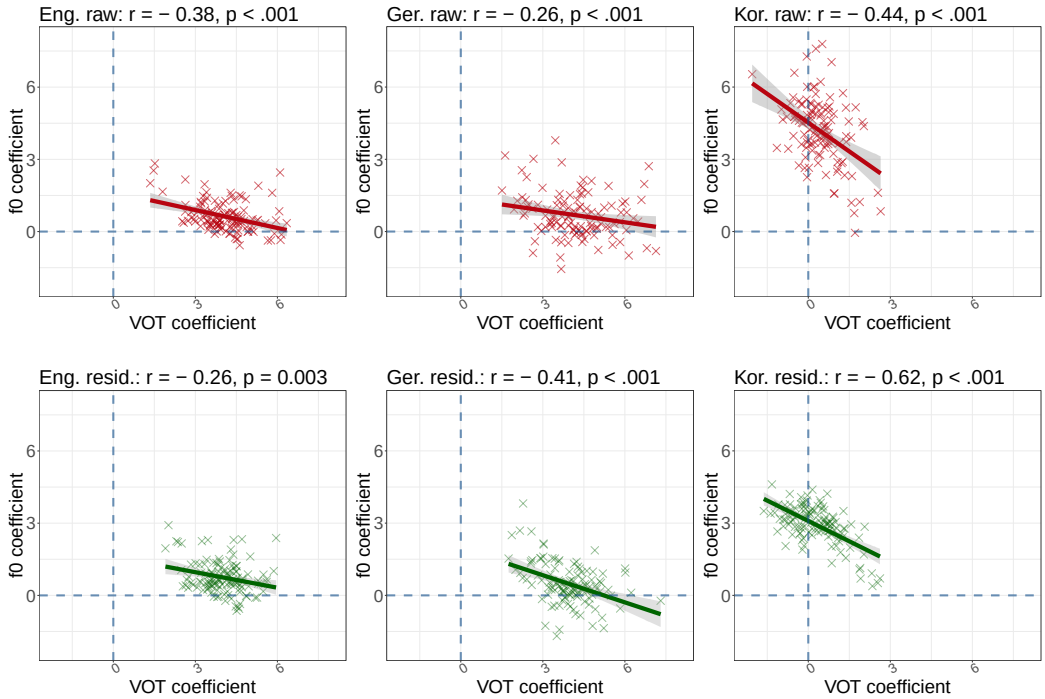


**Figure 3.7:** Coefficients computed from Linear Discriminant Analysis classification, performed on raw data (top row) and residuals (bottom row). In the plots, each point is one speaker and lines are linear fits, corresponding to the Pearson's $r$. Shadings are 95 % confidence intervals.

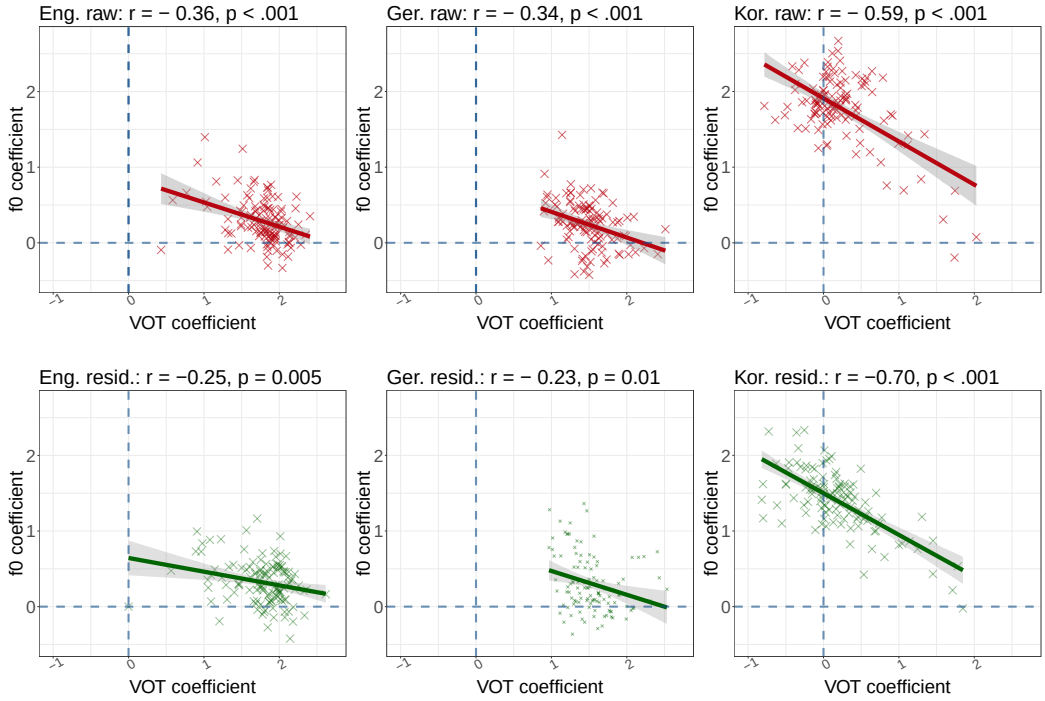


**Figure 3.8:** Coefficients obtained from Support Vector Machine classification, performed on raw data (top row) and residuals (bottom row). In the plots, each point is one speaker and the lines are linear fits, corresponding to the Pearson's $r$. Shadings are 95 % confidence intervals.

We performed both LDA and SVM classification on both the VOT (ms) and f0 (st) values, and the VOT$_r$ and f0$_r$ values, after standardizing each set of values (by transforming to z-scores) across the data from all speakers. For example, the full dataset of VOT (ms) values was z-transformed. By standardizing in this way, the VOT and f0 coefficients are comparable, and the coefficient for each cue provide a metric for how much each speaker weights a given dimension in production relative to the other.

#### 3.4.2.3.1 Linear discriminant analysis

The LDA coefficients for VOT and f0 were computed for each speaker using `lda()` in the `MASS` package in R (Venables & Ripley, 2002). We then compared how the VOT and f0 coeffiients (or cue weights) correlated across speakers in the three languages.

When performed on the raw data, significant negative correlations between VOT and f0 weights were found across speakers of all three languages. The correlation is

progresively stronger for German ($r = -0.26$, $p < .001$), English ($r = -0.38$, $p < .001$), and Korean ($r = -0.44$, $p < .001$).

The coefficients of VOT are positive for all German and English speakers, and most English speakers and the majority of German speakers also have positive coefficients for f0 in both raw and residualized data. Korean speakers have positive coefficients for f0 except for one older male speaker in his 60s. In terms of coefficients for VOT, about thirty-six Korean speakers have negative weights in VOT. Among them, twenty-four speakers produce Korean aspirated (traditionally long-lag) stops with shorter VOT than lenis (traditionally short-lag) stops, suggesting that negative weights do not necessarily mean reversed VOT values for aspirated and lenis stops but could be due to violation of the equal-covariance assumption of LDA. Compared to the speakers of German and English, Korean speakers have higher f0 coefficients and lower VOT coefficients. The results from LDA show opposite patterns for VOT and f0 in Korean versus English/German while the patterns between English and German speakers are more or less similar in VOT. One interesting difference between German and English is that the use of f0 seems to be stronger and less variable among English speakers compared to German speakers.

When the coefficients were computed on the residuals, the negative correlations are strengthened for Korean ($r = -0.62$, $p < .001$) and German ($r = -0.41$, $p < .001$) speakers but weakened for English speakers ($r = -0.26$, $p = 0.003$). Crucially, the negative correlations between the weights of VOT and f0 across speakers are preserved in the residualized data in all languages.

#### 3.4.2.3.2 Support vector machine

The SVM coefficients for VOT and f0 were computed on each speaker using the `e1071` package in R (Meyer, Dimitriadou, Hornik, Weingessel & Leisch, 2017). The patterns found in the weights of VOT and f0 computed by SVM are similar to the results from LDA: the VOT and f0 weights are negatively correlated in all three languages (English: $r = -0.36$, $p < .001$; German: $r = -0.34$, $p < .001$; Korean:$r = -0.59$, $p < .001$). Again, these correlations are maintained in the residualized data (English: $r = -0.25$, $p = 0.005$; German: $r = -0.23$, $p = 0.01$; Korean: $r = -0.7$, $p < .001$). Similarly to LDA, the results from SVM show that the strength of the correlation between the weights of VOT and f0 across speakers is in the order of German < English < Korean in both raw and residualized data.

Using both classification methods, we found that the use of VOT and f0 as cues to contrast stop categories negatively covary across speakers of the three languages. Furthermore, this cue covariation is stronger for Korean compared to English and German. We statistically tested this observation using linear regression models where VOT (or $\text{VOT}_r$) weights from LDA were predicted as a function of f0 (or $\text{f0}_r$) and languages (categorical variable with two levels: corresponding to Korean vs. non-Korean (LANGUAGE 1) and English vs. German (LANGUAGE 2)). We found significant interactions between LANGUAGE 1 and f0 cue weights in both the VOT data ($p = 0.022$) and $\text{VOT}_r$ data ($p = 0.017$), confiriming that the negative correlation between VOT weights and f0 weights across speakers is stronger in the language undergoing sound change than the other languages that are not. The cue weights from SVM returned similar results (VOT data: $p = 0.014$; $\text{VOT}_r$ data: $p < .0001$).

## 3.5 Discussion

The first goal of the current study was to examine whether and how multiple cues to phonetic categorization are structured across several sources of cue variability. This research question was motivated by previous findings in both speech perception and speech production. In speech perception, listeners use multiple cues to rapidly recover discrete linguistic units from a highly-variable signal. In speech production, cue covariation is linked to transphonologization, where the relative importance of two cues to the same contrast shift over time. The second goal of this study was to identify what structures in cue (co)variability could serve as precursors to sound change. This question was motivated by previous work showing for cases of synchronic variability (Beddor, 2009) and diachronic change (Bang et al., in press) that the informativity of the coarticulatory 'source' and that of the coaritculatory 'effect' in the signal are inversely related across some sources of variability (e.g. speakers, contexts). These previous findings suggest that synchronic trade-offs between the source and the effect of coarticulation may account for the phonologization of a traditionally redundant cue (Beddor, 2009; Bang, in press). The current study addressed this issue and further aimed to identity what specific structures in the variability serve as a catalyst for transphonologization.

We examined the structure of covariation in the use of VOT and f0 to signal stop voicing categories, across vowels of different heights, words with different frequencies, gender, and individual speakers, in three languages. We found interesting cross-lingusitic similarities and language-specific structures of this covariation. In German and English, languages that are not undergoing sound change, the total cue informativity is greater in high vowel contexts and low frequency words compared to non-high vowel contexts and

high frequency words, showing that one condition provides greater informativity than the other. Further, which cue is primarily affected by these linguistic sources of variability is language-specific, as discussed in more detail below (Sec. 3.5.1). Crucially, the conditions that are associated with less total cue informativity in German and English—non-high vowel contexts and high frequency words—coincide with those where transphonologization is further advanced in Seoul Korean. Assuming German and English can be taken to reflect what phonetic preconitions exist in a language not undergoing sound change (such as pre-change Seoul Korean), this pattern supports the idea in Bang et al. (in press) that the Seoul Korean sound change may have been triggered by decreased informativity in multiple cue dimensions in certain contexts, with enhancement of the redundant cue (f0) as the result of an adaptive mechanism to compensate for the weakened informativity of the primary cue (VOT) alone (Kirby, 2013). That is, prior to sound change in Seoul Korean, VOT may have been subject to variability in informativity across the linguistic souces of variability similarly to what we found in German. However, as observed in English data, it is not always the primary cue that is subject to variability. In some languages, it may be the secondary cue that is more subject to cue variability in certain conditions. Since the primary cue provides enough informativity in this type of language in the conditions where total informativity decreases, these languages are less likely to undergo transphonologization. This may partly accout for the actuation problem, with respect to transphonologization—why sound change is not more common, despite the existence of precursors to change in many languages (Sóskuthy, 2015; Weinreich et al., 1968).

In contrast to linguistic factors, the structure of cue covariation across gender and individual speakers were very similar across languages, even though there was a cross-

linguistic difference in the effect of gender. Female talkers placed more weight on VOT and less on f0 than male talkers in German and English, while the reverse pattern was found in Seoul Korean (see Sec. 3.5.2 for further discussion). The strengths of VOT and f0 are inversely correlated across speakers of different genders, and across individual speakers, while cue informativity is more or less maintained. We will discuss this speaker-level cue structure in more detail in Sec. 3.5.2, including the possibility that it is shaped by socio-phonetic and stylistic factors. The existence of cue trade-offs across gender and individual speakers in *all* languages supports the idea that language 'adaptivity' (Lindblom et al., 1995) operates to satisfy the speakers' goal of maintaining total cue informativity similar to others in the speech community. The resulting negative correlation in cue weights may serve as an important precursor to transphonologization.

We now turn to the discussion of our research questions in detail by presenting what the structures of cue covariation are in each of the sources of variability and what implications these structures have for transphonologization.

### 3.5.1 Linguistic factors

We first examined whether and how the weights of VOT and f0 for the stop voicing categorization in speech production (co)vary as a function of two factors: word frequency and vowel height. We found in both English and German that high-frequency words and non-high vowel contexts are associated with less distinction between voiced and voiceless stops (i.e., lower informativity), in VOT/f0 space. Less total informativity in these conditions is due to variability in the weights of a different cue, in each language: primarily variability in f0 in English, and primarily variability in VOT in German. Thus, cross-linguistically, the total cue informativity (in VOT/F0 space) varies as a function of

these lingustic factors, but which cue is primarily affected is language-specific.

Our findings in German and English, that the stop voicing contrast is signaled with reduced informativity in two-dimensional cue space (VOT/F0) for high-frequency words, parallels the findings of Munson & Solomon (2004) for English vowels, which are signaled with smaller spectral distinction in F1/F2 space in high frequency words. In addition, the dependence of f0 contrast strength on frequency for English words is consistent with the findings of Zhao & Jurafsky (2009) for a tonal language (Cantonese), where f0 distinctions across lexical tones decrease as word frequency increases. Our finding extends the effect of word frequency on pitch distinctions to redundant f0 distinctions in a non-tonal language. The effect of word frequency that primarily targets VOT in German can be partly explained by the previous findings where word probability such as word frequency (Rhodes, 1992, 1996) decreases word duration, likely associated with segmental and cue duration. Then, one possibility is that long- and short-lag categories do not undergo duration reduction to the same magnitude but word frequency affect long-lag stops substantially more than short-lag stops as observed in some recent studies (Sonderegger et al., 2017; Stuart-Smith et al., 2015). This asymmetry between long- and short-lag stops is in fact well-documented in the literature on how speaking rate affects VOT of different stop categories: in faster speech, the difference between long- and short-lag stops becomes smaller, largely due to the long-lag stop's VOT decreasing more than the short-lag stop's VOT cross-linguistically (Kessinger & Blumstein, 1997; Miller et al., 1986; Pind, 1995). If this asymmetry is applied to other reduction conditions, it will explain why the strength of VOT decreases in German high-frequency words.

As for vowel height, although it has repeatedly been shown that within-category VOTs for long-lag stops are longer and f0s are higher in high vowel contexts compared to non-

high vowel contexts (Esposito, 2002; Higgins et al., 1998; Klatt, 1975; Weismer, 1979), to our knowledge, there is no study exploring how the strength of cue contrasts between stop categories varies across different vowel contexts. Our study provided cross-linguistic evidence that it is the non-high vowel context that is associated with contrast reduction when the total of cue informativity in both dimensions is considered. However, the motivation behind the difference between high and non-high vowels in light of cue informativity is less clear than words with different frequencies where meaning predictability is known to be negatively correlated with acoustic redundancy (Aylett & Turk, 2004, 2006). We raise one possibilty here, related to perceptual integration between f0 and f1 in stop voicing perception: low f0 and low f1 enhance the auditory perception of the short-lag category, while high f0 and high f1 are associated with the perception of the long-lag category (Benkí, 2001; Kluender, 1991; Lisker, 1975; Summerfield & Haggard, 1977). We speculate that high vowels, which are characterized by low F1, would cause perceptual bias towards the short-lag category for post-vocalic stops when VOT is an ambiguous or less informative cue to voicing category—which is where perceptual trading relations are known to operate (Kingston, 1992; Repp, 1982). To compensate for this bias, speakers may make use of their prior knowledge about high vowels intrinsically having higher f0 and greater VOT and enhance either of these intrinsic patterns (Hoole & Honda, 2011; Hoole et al., 2006) for the voiceless category. This enhancement would result in a greater contrast in either VOT or f0 in high vowel contexts. This spectulation will be borne out if the f0 of voiceless stops are more affected by vowel height than voiced stops, a prediction to be tested in future work.

Evidence of vowel context effects on stop voicing judgements, even if indirect, can be found in two previous studies. Results reported by Fischer & Ohde (1990) suggest that

when steady-state F1 is higher, as in lower vowels, falling F1 during the transition serves as a cue to post-vocalic stop voicing in perceptual judgements.[5] Supporting evidence for the asymmetric effect of vowel height on pre-vocalic stop voicing judgements can be found in Mayo & Turk (2004) which examined stop voicing categorization along a /tV/∼/dV/ continuum, for V of different heights. In the study, listeners relied exclusively on VOT when V was a high vowel (/i/), but relied both on VOT and a secondary cue (formant transition) when V was a non-high vowel (/a/). One interpretation of Mayo & Turk's (English) results is that listeners may make use of the secondary cue more in this context *because* VOT is less informative, compared to high vowel context.

The effects of word frequency and vowel height on cue relationships in Seoul Korean exhibit similarities with and yet crucial differences from English and German. Similarly to German, Korean stops in high frequency words and non-high vowels undergo greater VOT reduction than those in low frequency words and high vowel contexts. However, unlike in English and German, f0 showed the opposite pattern to VOT: f0 was a stronger cue in exactly the contexts in which VOT was reduced, suggesting a pattern of 'compensatory enhancement'. That is, VOT strength and f0 strengths are negatively correlated across words and vowel contexts (as in Figure 3.1). Given that high frequency words and non-high vowel contexts provide reduced total cue informativity in the other languages, we argue that the trade-off between VOT and f0 strength found in Seoul Korean is an adaptive reaction used to maintain phonological contrasts in these particular environments, where the VOT contrast was already smaller in pre-change Seoul Korean.

[5] I would like to thank John Kingston for this insightful comment.

### 3.5.2 Across gender and individual speakers

We found that male and female speakers differ in how they use VOT and f0 to signal stop voicing categories, in each language. For English and German, female speakers signal the contrast using a greater VOT strength and a smaller f0 strength compared to male speakers. For Korean, the pattern was reversed: female speakers produced stop contrasts using smaller VOT strength and greater f0 strength compared to male speakers, consistent with previous work (Bang in press, Kang, 2014) with the same dataset showing that female speakers are leading the ongoing transphonologization sound change.

We found an interesting cross-linguistic pattern related to speaker gender when considering the primary cue in each language: VOT for German and English speakers, and f0 for (the majority of) Korean speakers. Women show a greater distinction for this cue, which can be interpreted as women hyperarticulating in the primary cue dimension. This interpretation is consistent with the common cross-linguistic pattern that women tend to speak more clearly than men (e.g. Bang et al., 2017; Byrd, 1994; Diehl et al., 1996; Ferguson, 2004; Hillenbrand et al., 1995; Simpson, 2009; Smiljanic & Bradlow, 2009). For example, female speakers tend to produce vowels with greater spectral distinctions (e.g. Hillenbrand et al., 1995), more acoustically canonical /s/ (e.g. Bang et al., 2017) and more frequent sentence-final stops with release, and fewer flapping of intervocalic stops (e.g. Byrd, 1994) than male speakers While most of the previous studies concern how men and women differ in the degree of hyperarticulation within a category, our results provide evidence that cross-linguistically, women contrast categories with a stronger primary cue than men.

Unlike the results considering only the primary cues, when multiple cues are considered together, it is not clear whether the variability observed across genders differs in the

degree of hyperarticulation. In fact, the existence of trade-offs between the strengths of VOT and f0 across genders in all three languages provides evidence that the total of cue informativity of the signal may be invariable across genders, all else being equal, consistent with Pattern C (Figure 3.1-C). This finding, together with greater total cue informativity in low-frequency words and high vowel contexts due to enhanced VOT in German and f0 in English, suggests that hyperarticulation is not the same speech phenomenon as increasing information of the primary cue alone. Rather it may be the case that different types of cue enhancement are required in different hyperspeech conditions (Martin, Utsugi & Mazukaac, 2014) and which cues are enhanced may also be constrained by language-specific, soiophonetic, and linguistic constraints. For example, Schertz (2013) examined acoustic changes in categorizing English tense /i/ and lenis /ɪ/ vowels in hypo- versus hyper-speech induced by 'mishearing' of the target segment and found that it was the secondary cue (durational differences) that was enhanced rather than the primary cue (formant differences) while in other clear speech conditions, it is often the case that formant contrasts (F1  F2 distances) among vowel categories are enhanced (Bradlow, Kraus & Hayes, 2003; Ferguson & Kewley-Port, 2007; Picheny, Durlach & Braida, 1986). We leave the issue of how individuals differ in the use of multiple cues across social groups and linguistic conditions, as well as across languages, to future work.

Across individual speakers, we also found a negative correlation between the weights of VOT and f0 in all three languages, while cross-linguistic differences are present in the strengths of the correlations. Korean speakers exhibited the strongest correlation compared to speakers of German and English, likely reflecting the ongoing sound change in Korean affecting the relative weight of cues that signal the aspirated/lenis stop contrast, and the tight correlation between speaker variability and speakers' year of birth (e.g.

Bang et al., 2015; Kang, 2014; Silva, 2006). In addition, we found that except for a handful of older male speakers who weight VOT more highly than f0, the majority of Korean speakers weight f0 more highly than VOT, with some speakers even showing a reversed aspirated/lenis VOT pattern. These findings are consistent with our previous work showing that the change in primary cue in phrase-initial position in this language, from VOT to f0, is nearing completion (Kang, 2014; Bang et al., in press).

Our results further showed that similarly to the findings for gender, there is great speaker variability within the primary cue in how much speakers make use of the cue in signalling stop voicing contrasts. In other words, speakers differ in how much they hyper-articulate within the primary cue (Bang & Clayards, 2016). However, when VOT and f0 are considered together, speakers who use more of one cue tend to use less of the other, again showing trade-offs between cues, consistent with the constant total informativity hypothesis across speakers. Importantly, these types of correlations were present even after the effect of gender (along with other sources of variation) was parsed out.

What explains these individual differences in how cues are used to signal voicing contrasts, beyond gender? One possibility is that other social variables, age, ethnicity, and social class may account for some variability in speakers' cue weights in production. Whether this kind of socially structured variation in cue *weights* is restricted to gender, or holds for other social variables, it could be useful for speech perception, in line with work showing that listeners use social information encoded in cue *values* (e.g. f0 for /p/, not its use in contrast) to resolve ambiguity in the speech signal (age: Drager (2011a); Hay et al. (2006); gender: Johnson et al. (1999); Strand & Johnson (1996); ethnicity: Staum-Casasanto (2009); social class: Hay et al. (2006)). For example, Drager (2011a) found that when listeners of New Zealand English, where /æ/ is undergoing raising and merger

with /e/, are exposed to resynthesized vowel tokens in a continuum between /æ/ and /e/, voices perceived as "young" associated with a paired photo bias perception towards /æ/ in some conditions. If the trade-offs in VOT and f0 strengths observed across speakers are in part due to social group differences, this could serve to simplify the listeners' job in speech categorization as it would reduce the range of cue distributions or number of combinatorial possibilities once a speaker's social identity factors, such as gender (Johnson et al., 1999; Strand & Johnson, 1996) and sexual orientation (Munson, 2007), are identified (Kleinschmidt & Jaeger, 2015; Sumner, 2014). In addition, if individuals may also differ in the degree of hyperarticulation in the primary cue dimension alone, the use of the rest of the cues may be constrained by speakers' attempt to maintain total cue informativity with other speakers. This structured individual variability would then facilitate speech perception because all listeners need to do would be to identify how a subset of speech cues operate for a given speaker within the intersection of their social groups and all the rest may come for free. Similar suggestions have been made in recent studies (Bang & Clayards, 2016; Chodroff & Wilson, 2017). Chodroff & Wilson (2017) examined speaker variability —albeit in one cue dimension (VOT)—and found correlations of speakers' mean VOT values across stops with different place of articulation and voicing. Bang & Clayards (2016) found correlations in talker variability across cues and contrasts. For example, speakers who produce longer VOT for voiceless stops produced /s/ with longer duration and higher CoG (centroid or spectral mean), and produced more peripheral vowels. Taken together, these studies and the current results suggest that variability across speakers may be structured in multi-dimensional space in a way that simplifies speech perception.

As for what properties of cue correlations cause a language to undergo change in cue

primacy, our findings from the three languages together suggest a relationship between transphonologization and structured varibility in VOT and f0 weights across speakers (individuals, and speakers of different genders). Our findings suggest that such correlations exist in all languages with a short-lag long-lag stop contrast, and may serve as a precursor to sound change. Transphonologization would then proceed by 'exaggerating' in some sense the cross-speaker correlations which already exist. The question then arises of how this sound change is triggered. Our finding that both synchronic reduction of total cue informativity and more advanced (diachronic) transphonologization occur in the same linguistic conditions suggest a link between the two. We propose that transphonologization is triggered by the enhancement of the secondary cue in the lingusitic conditions where the primary cue to signal the stop voicing contrast is weakened (Kirby, 2013). This weakened primary cue may cause misparsing of the speech signal, leading the listener to infer a different category from what the speaker intended (Ohala, 1981, 1993a). If the listener-turned-speaker selects a variant from the pronunciation variants that exist across speakers, reflects it in their own production (Lindblom et al., 1995), and this selected form is imitated by other speakers (Baker et al., 2011; Harrington, 2012), the new variant could propogate in the community, and the sound change would be said to have 'actuated'. The 'actuation problem' in this case would be: why has this process happened in Korean, but not German or English? In the case of Korean, the language-specific high informativity of f0 associated with the role of f0 marking intonational boundaries (Jun, 1996, 1998) may have been the language-specific motivation that actuated the change, as suggested in our previous study (Bang et al., in press).

Taken together, our results from three languages provide evidence that trade-offs between the weights of VOT and f0 in contrasting stop voicing categories across speakers

may serve as a precursor for transphonologization. In contrast, the trade-offs in VOT and f0 weights as a function of linguistic factors observed in Korean may be a *consequence* of the language undergoing transphonologization—an adaptive response to avoid merger, especially in conditions prone to contrast reduction.

### 3.5.3 Further remarks

The findings from the current study are a first step towards understanding how cues are structured in the signal in a way that facilitates speech perception and may at the same time serve as a precursor to sound change. This study has several limitations which suggest directions for future work.

First, our findings on cross-linguistic cue structures are based on only three languages (two of which are related). More languages must be investigated to test our idea that total cue informativity may remain more or less constant across speakers in most languages, and that this pattern is a precursor to transphonologization. Further, our findings on cue structures are limited to languages whose stop systems (at least partly) contrast short- and long-lag stops. Languages with different stop systems, including 'true voicing' languages, may have distinct structures. A similar account was suggested by Kirby (2016) who examined correlations between the weights of onset f0 and those of VOT across speakers in six languages. In the study, strong negative correlations at the speaker level were observed only in the languages whose stop system contains long-lag stops but no robust correlations were noted in the languages that contrast prevoiced and short-lag categories. This finding is consistent with the emerging re-interpretation of VOT that speakers do not regulate how early voicing begins prior to the release in the same way that they regulate how late it begins after the release. In other words, VOT only applies to

lags and not leads (see Mikuteit & Reetz (2007)).[6] Further studies should address similar research questions to the ones addressed in the current study across a broader range of languages. This line of work could shed light on whether VOT/f0 covariation across speakers is present in all languages as a universal property or is limited to languages that, for instance, do not use prevoicing as a robust cue to stop categorization.

A second caveat is that only two cues were analyzed to explain how multiple cues combine together to signal phonetic contrasts. In fact, listeners integrate all available available acoustic cues to make categorization judgements (McMurray & Jongman, 2011; Nearey, 1997). For example, in a study where cue-integration models are compared using a classification method, (McMurray & Jongman, 2011) found that the models based on maximal cues (24 cues) performed more similarly to listeners' than a model using fewer invariant cues (13 cues), indicating that listeners make use of most of the cues available to them in speech categorization. Therefore, even though the cues examined here (VOT, f0) are known to be among the most important in signaling stop contrasts, to widen our uderstanding of how cues are weighted in speech production, further studies are required that would expand the current method and findings to multiple cues.

Finally, the less controlled prosodic positions where the target stop tokens were extracted in our data may account for some of the cross-lingusitic differences reported in the current study. A future study controlling for this complication with similar findings will strengthen the arguments of this study.

In summary, we found structured variability in multiple cues to phonetic categorization, for the case of VOT and f0 for stop voicing contrasts. The structure may be explained by speech efficiency associated with the listeners' need of cue informativity, speakers' social identity, and the priciple of total cue informativity acoross speakers in

[6]These two sentences are quoted from John Kingston's comments.

multiple dimensions. Furthermore, the speakers' goal of maintaining total cue informativity with others while expressing socio-phonetically conditioned stylistic variations appears to serve as a precursor to sound change. Transphonologization may then initiate in the linguistic conditions where phonetic bias is present and progress by strengthening the existing cue correlations.

# Appendices

**Table B1:** English: 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.

| Laryngeal class | Stop | VOT (msec) | | f0 (Hz) | | $n$ |
|---|---|---|---|---|---|---|
| | | mean | SD | mean | SD | |
| voiced | b | 10.76 | 5.91 | 158.58 | 46.87 | 778 |
| | d | 13.84 | 5.69 | 155.23 | 46.87 | 492 |
| | g | 20.34 | 7.06 | 158.34 | 46.71 | 342 |
| voiceless | p | 59.31 | 17.14 | 165.34 | 48.11 | 1027 |
| | t | 49.95 | 16.29 | 168.51 | 48.14 | 815 |
| | k | 50.86 | 15.42 | 164.36 | 45.95 | 754 |

**Table B2:** German: 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.

| Laryngeal class | Stop | VOT (msec) | | f0 (Hz) | | $n$ |
|---|---|---|---|---|---|---|
| | | mean | SD | mean | SD | |
| voiced | b | 9.36 | 4.77 | 175.01 | 52.45 | 522 |
| | d | 12.56 | 5.26 | 172.33 | 49.53 | 554 |
| | g | 18.03 | 8.30 | 182.94 | 55.09 | 646 |
| voiceless | p | 44.90 | 21.54 | 186.32 | 52.65 | 232 |
| | t | 55.28 | 20.00 | 190.25 | 53.37 | 312 |
| | k | 57.24 | 18.50 | 184.70 | 54.77 | 394 |

# Chapter 4

# General discussion and conclusion

## 4.1 Discussion

The central goal of the dissertation was twofold: 1) to understand the structures of cue variability in multiple dimensions across linguistic factors, social factors, and individuals and 2) to identify the relationship between these structures and a specific type of sound change that involves change in cue primacy (i.e. transphonologization). These questions were addressed in two studies using speech corpus data with acoustic analyses and statistical modelling.

In the first study, I examined how cue primacy to stop voicing contrasts shifts from VOT to f0 within a speech community over generations. This question was addressed by observing an ongoing sound change in Seoul Korean using a large data set. The analysis focused on how the change in the way Korean aspirated and lax stop categories are contrasted is initiated and how it progresses across speakers (of different ages and gender) and the language (e.g. words and phonetic environments), and how it impacts other systems of the language over the course of the change.

As for the path of change across speakers, both VOT contrast reduction and f0 contrast enhancement were most advanced in younger female speakers' speech and least advanced in older male speakers' speech, which is consistent with previous findings (Kang, 2014). To examine the effects of linguistic factors on the diachronic VOT contrast reduction and f0 contrast enhancement, word frequency and vowel height were considered. One novel finding was that the change is most advanced in high frequency words and in nonhigh vowel contexts; contexts where the VOT contrast is (presumably) synchronically weakest.

Another novel finding in this study was that the enhancement (i.e. increased informativity) of f0 in contrasting the stop categories also impacts the intrinsic f0 (IF0) related to the tongue gestures for vowel height contrasts. Furthermore, the attenuation of IF0 was most noticeable for aspirated stops even though the diachronic change in the role of f0 in contrasting Seoul Korean stops progresses similarly for both aspirated and tense stops. It seems that this interesting phenomenon is due to a combined effect of the reorganized laryngeal settings for contrastive f0 production for stop categorization and the level of importance of f0 to signal each category; VOT is informative enough to distinguish tense stops from the other categories even after the sounds change, while VOT is not as informative in distinguishing aspirated stops, necessitating enhancement of f0 across vowel environments.

These findings from Seoul Korean shed light on the mechanisms of transphonologization that involve cue trade-offs in detail. However, it is still challenging to demarcate between synchronic cue trade-offs that may exist as preconditions to sound change and diachronic trade-offs that may be unique cue patterns that occur during transphonologization. This was the part of the question addressed in the following study.

In the second study, I first sought to better describe the relationship between VOT

and f0 as cues to stop voicing categories across the same factors considered in Study 1. Further, I compared synchronic and diachronic variation by comparing the findings from two languages (German and English) that are not undergoing sound change with the findings in Seoul Korean where change is underway. From this approach, I aimed to distinguish the cue correlations that are cross-linguistic phenomena and the ones that are properties unique to ongoing sound change.

Three possible outcomes were hypothesized. First, use of VOT and f0 could be negatively correlated such that words or speakers who use one more use the other less (cue trade-offs), preserving overall contrast informativity, as was observed in the first experiment on Seoul Korean. These cue correlations observed in Korean indicate that cue trade-offs are at least in part due to the change in progress that involves transphonologization possibly to avoid contrast merger. In Study 2, I assumed that observing this pattern in German and English, even if it is weaker than in Seoul Korean, could indicate that it provides a synchronic precursor to transphonologization. On the other hand, if VOT and f0 weights are positively correlated across for example, speakers or words, it could indicate that speakers or words may differ along a hypo- and hyper-articulation continuum. In other words, some speakers tend to articulate more clearly than others and some words tend to be articulated more clearly than other words. The occurrence of this pattern would indicate that overall informativity differs across some sources of variability; e.g. some speakers may produce all cues more clearly than others. Finally, I may only observe variation in cue use across speakers or words in a single cue (e.g. the primary cue ) without any correlations in how the two cues are used. The cue pattern would mean that total cue informativity would vary due to the noticeable variability in a single dimension and at the same time the absence of strong variability in the other

dimensions.

With these predictions in mind, the corpus data from German and English were analyzed in a similar way as was done for Seoul Korean in Study 1. The Seoul Korean data was also analyzed further in parallel with the analyses performed on the German and English data. In German and English, the results did not provide strong evidence that the size of the VOT and f0 cue contrasts are correlated across words of different frequencies or vowels of different heights, in contrast to the findings in Seoul Korean. In these languages, cue variation in these linguistic conditions was noticeable only in one cue dimension and which cue is most affected seems to be language-specific; VOT for German and f0 for English. The lack of cue correlations seems to indicate that the total cue informativity that signals stop voicing categories is greater in certain words and contexts than other words and contexts; in both languages, it was the low frequency words and high vowel contexts that showed greater total cue informativity, which is expected to be a feature of hyperspeech. Crucially, the conditions where total cue informativity are weakest are the same conditions where the sound change in Seoul Korean is more advanced, which supports the argument made in Study 1 that the change may have been triggered by phonetic bias related to contrast reduction or reduced cue informativity.

Our results on gender and individual speakers, on the other hand, showed consistent negative correlations in all three languages while the correlation was strongest for Seoul Korean. Diachronically, the stronger cue correlations in Seoul Korean than other languages may imply that transphonologization is progressing by strengthening the existing synchronic variability among speakers. Synchronically, the negative correlations across gender further suggest that each speaker's goal is to signal the contrast with a combined cue informativity comparable to other members in the speech community while at the

same time (possibly) showing gender-specific stylistic variation. That is, when multiple cues are considered together, the summed informativity is more or less similar across speakers. However, deciding which cue is weighted more than other cues in contrasting a sound category may be a way to express their social identity. Negative correlations were observed across speakers after gender differences were accounted for. One possibility is that other social factors that were not considered in the current study may contribute to the individual variability in the use of cues. Whatever factors contribute to the observed cue correlations across speakers, the findings imply a crucial mechanism in understanding transphonologization at least for the one that is taking place in Seoul Korean; the change is initiated by production bias, spreading through adaptivity to avoid merger and by strenthening existing synchronic variability which is structured across speakers.

There was a noticeable cross-linguistic difference between Korean and the other languages in the way speakers weigh one cue relative to the other. VOT is the primary cue to signal stop categories for all speakers in German and English, while f0 is the primary cue to most of the speakers in Seoul Korean. Interestingly, in English and German, female speakers place more weight on VOT and less on f0 than male speakers. In Korean, however, it is the male speakers who place more weight on VOT and less on f0 than female speakers. This cross-linguistic difference seems to suggest that there may be a tendency for female speakers to weigh the primary cue more heavily than their male counterparts. The argument is also consistent with many previous findings from acoustic studies that reported women's tendency towards clear speech compared to men - studies which mostly considered the primary cue alone.

These findings suggest an intriguing possibility for the role of gender in sound change. The results from two studies suggested that transphonologization initiates in linguistic

conditions where phonetic bias is present with reduced cue informativity and progresses by enhancing the secondary cue to compensate for the weakened primary cue (Kirby, 2013). Interestingly, however, the change in Seoul Korean is led by female speakers, who under non-change conditions would place more weight on the primary cue. It may be the case that before change initiated, Seoul Korean women put more weight on the traditionally primary cue, VOT, than men. Then, in the linguistic conditions where informativity was reduced, women may enhance the secondary cue more than men. Once the change is actuated for some language-specific reason, this tendency may make women the pioneers of sound change. This is only a speculation, which, however, could be tested when an interaction between gender and the linguistic factors is considered, which I leave to future research.

After discussing the cross-linguistic similarities in the structure of variability, one question arises; why it is only Seoul Korean that is undergoing change while the pre-conditions to transphonologization are present in all three languages (and possibly in many more languages). This answer might be found in the level of informativity of f0 in this language (Kirby, 2013). The language-specific implementation of f0 for prosodic boundary marking in Seoul Korean that makes an active use of consonant-perturbed f0 attributes (Jun, 1993, 1996) may inherently increase f0 informativity in this language compared to other languages. The language-specific level of f0 informativity may make Seoul Korean listeners easily attuned to this cue when VOT informativity is weak (e.g. high frequency words and non-high vowel contexts), actuating transphonologization.

### 4.1.1 Summary

To sum up, the cross-linguistic data showed that certain linguistic conditions reduce total cue informativity for a contrast. Furthermore, speakers' tendency to balance total cue informativity relative to other speakers seems to be present across languages. Crucially, the linguistic conditions where total cue informativity is reduced in languages not undergoing transphonologization are the same environments (high frequency words, non-high vowels) where transphonologization in Seoul Korean is more advanced. These findings indicate that trasphonologization may progress to maintain the number of sound contrasts in the language by strengthening the existing pronunciation variations across speakers.

### 4.1.2 Further remark

An additional note is to be made with respect to speech perception. This structured variability and the concept of constant total cue informativity across speakers may have important implications for speech perception. The structure may facilitate speech categorization in the face of talker variability because it implies that all listeners need to do is to identify how a subset of speech cues operate for a given speaker within the intersection of their socio-phonetic groups. In other words, variability across speakers may be detected simply from one dimensional space. The principle of total cue informativity would then make perception easier because identifying one value in one dimension for the speaker may accompany other information necessary for speech perception for free, which may include values for other cues along with socio-indexical information for the given speaker.

### 4.1.3 Future directions

The interpretation of our results on the structure of cue variability in multiple dimensions and its role on sound change is well-supported by the current data. However, it can be further strengthened when this analysis is expanded to other languages. Right now, the results are based on languages with a similar laryngeal system that (at least partly) contrast between long- and short-lag stops without having a true voicing contrast or a lexical tonal system. Even though the results are robust in the languages concerned in this dissertation, it is less clear how the availability of additional cues or other phonological structures (e.g. the existence of true voicing, the structure of tonal systems) would further constrain the relationship between cues. The specific questions that future research should ask with respect to this is "would the magnitude of VOT/f0 covariation across speakers be different in languages depending on their laryngeal structures?" This question could be further developed into more specific questions, which include; "would the amount of prevoicing affect the VOT/f0 cue correlations given that prevoicing will be another cue salient enough to signal voicing contrasts?"; "would the complexity of the tonal structure of a tonal language influence VOT/f0 cue correlations for stop voicing contrasts as increasing tonal complexity may mean more fine control of f0 is expected to signal tonal contrasts?" This line of research should broaden our understanding on the relationship between language structures and the structures of cue variability in multidimensional phonetic space.

Another line of research that is worth conducting in the near future concerns the crucial finding in this dissertation that even languages that are not undergoing a tonogenetic sound change already show inter-speaker variability similar to the variation found in Seoul Korean. Furthermore, the gender effects observed in all three languages suggested that

the inter-speaker variability in the use of multiple cues may be structured across various social factors. Identifying how social factors structure speech variability in multiple acoustic cue dimensions has not been considered in any past work. Any knowledge from future research that addresses this aspect of speech variability should shed light on the relationship between pronunciation variation and socio-phonetic factors in the use of multiple cues as well as its effect on a long-term change in pronunciation norms.

Futhermore, future studies should investigate how sensitive listeners are to social and indexical information in speech perception, and whether this sensitivity is consistent with the distributional properties in speech production in their native language. Given that there is systematic pronunciation variation associated with the difference in the use of VOT and f0 between old and young Seoul Korean speakers due to the ongoing change, one prediction is that perceptual categorization will be affected by the socio-indexical information about speakers. For example, if speakers' 'perceived' age is manipulated, for a speaker whose perceived age is 'old' listeners are expected to weigh VOT more than f0 while for a speaker who's perceived age is 'young', listeners' judgment should primarily depend on f0. This line of research could broaden our knowledge of the sensitivity of listeners to socio-indexical information present in the speech signal, especially for listeners who's language is undergoing a change in norms.

## 4.2 Conclusion

To conclude, the results from two studies suggest that cue trade-offs across individuals are observed cross-linguistically. This may reflect each speaker's goal to signal the contrast with a combined cue informativity comparable to other members in the speech community while at the same time showing socio-phonetically constrained stylistic variation in how

the cues are weighted relative to each other. This variation across talkers may serve as a catalyst to sound change. Secondly, transphonologization seems to initiate in linguistic conditions where phonetic bias is present, weakening the primary cue and triggering enhancement of the secondary cue. Then, change seems to spread by strengthening the existing cue correlations across speakers. Actuation may be decided by language-specific constraints such as the level of the informativity of the redundant cue. The results further suggest that this structured variability also may account for how listeners cope with high variability in the signal in real-time speech processing.

# Bibliography


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.

Abramson, A. S. & Lisker, L. (1985). Relative power of cues: F0 shift versus voice timing. In V. Fromkin (Ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged* (pp. 25–33). San Diego: Academic Press.

Allen, J. S., Miller, J. L., & DeSteno, D. (2003). Individual talker differences in voice–onset–time. *The Journal of the Acoustical Society of America*, *113*(1), 544–552.

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.

Bailey, P. J. & Summerfield, Q. (1980). Information in speech: observations on the perception of [s]-stop clusters. *Journal of Experimental Psychology: Human Perception and Performance*, *6*, 536–563.

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. (2017). The acoustic counterpart to coarticulatory resistance and aggressiveness in locus equation metrics and vowel dispersion. *Journal of the Acoustical Society of America*, *141*, EL345–EL350.

Bang, H.-Y. & Clayards, M. (2016). Structured Variation across Sound Contrasts, Talkers, and Speech Styles. In *The 15th Conference on Laboratory Phonology*. Cornell University, Ithaca, USA.

Bang, H.-Y., Clayards, M., & Goad, H. (2017). Compensatory strategies in the developmental patterns of english /s/: Gender and vowel context effects. *Journal of Speech, Language, and Hearing Research*, *60*, 571–591.

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 *the 18th International Congress of Phonetic Science*, Glasgow, Scotland, UK.

Bang, H. Y., Sonderegger, M., Kang, Y., Clayards, M., & Yoon, T.-J. (2018). The emergence, progress, and impact of sound change in progress in Seoul Korean: implications for mechanisms of tonogenesis. *Journal of Phonetics*, *66*, 120–144.

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.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Beckman, M. E., Li, F., & Kong, E. 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.

Bell-Berti, F. & Harris, K. S. (1979). Anticipatory coarticulation: Some implication from a study of lip rounding. *Journal of the Acoustical Society of America*, *65*(5), 1268–1270.

Benguerel, A. P. & Cowan, H. A. (1974). Coarticulation of upper lip protrusion in French. *Phonetica*, *30*, 41–55.

Benkí, J. (2001). Place of articulation and first formant transition pattern both affect perception of voicing in english. *Journal of Phonetics*, *29*(1), 1–22.

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.

Blevins, J. (2004). *Evolutionary phonology: The emergence of sound patterns.* Cambridge University Press.

Bradlow, A. R., Kraus, N., & Hayes, E. (2003). Speaking clearly for learning-impaired children: sentence perception in noise. *Journal of Speech and Hearing Research*, *46*, 80–97.

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

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in german. *Experimental Psychology*, *58*, 412–424.

Brysbaert, M. & New, B. (2009). Moving beyond kucera and francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for american english. *Behavior Research Methods*, *41*, 977–990.

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.

Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication*, *15*(1-2), 39–54.

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.

Chodroff, E. & Wilson, C. (2014). Burst spectrum as a cue for the stop voicing contrast in American English. *The Journal of the Acoustical Society of America*, *136*(5), 2762–2772.

Chodroff, E. & Wilson, C. (2017). Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *Journal of Phonetics*, *61*, 30–47.

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 variability in multiple cues to stop voicing. *Phonetica*, *75*, 1–23.

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, *108*(3), 804–809.

Coetzee, A. W., Beddor, P. S., & Wissing, D. P. (2014). Emergent tonogenesis in Afrikaans. *Journal of the Acoustical Society of America*, *135*, 2421.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Routledge Academic.

Cole, J., Linebaugh, G., Munson, C., & McMurray, B. (2010). Unmasking the acoustic effects of vowel-to-vowel coarticulation: A statistical modeling approach. *Journal of Phonetics*, *38*(2), 167–184.

Connell, B. (2002). Tone languages and the universality of intrinsic F0: evidence from Africa. *Journal of Phonetics*, *30*(1), 101–129.

Croux, C. & Joossens, K. (2005). Influence of observations on the misclassification probability in quadratic discriminant analysis. *Journal of Multivariate Analysis*, *96*(2), 384–403.

DiCanio, C. T. (2012). Coarticulation between tone and glottal consonants in Itunyoso Trique. *Journal of Phonetics*, *40*(1), 162–176.

Diehl, R. L., Lindblom, B., Hoemeke, K. A., & Fahey, R. P. (1996). On explaining certain male-female differences in the phonetic realization of vowel categories. *Journal of Phonetics*, *24*(2), 187–208.

Dmitrieva, O., Llanos, F., Shultz, A. A., & Francis, A. L. (2015). Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics*, *49*, 77–95.

Docherty, G. J. & Foulkes, P. (1999). Derby and Newcastle: instrumental phonetics and variationist studies. In P. Foulkes & G. J. Docherty (Eds.), *In Urban Voices: Accent Studies in the British Isles* (pp. 47–71). London: Arnold.

Dorman, M. F., Raphael, L. J., & Isenberg, D. (1980). Acoustic cues for a fricative-affricate contrast in word-final position. *Journal of Phonetics*, *8*, 397–405.

Dorman, M. F., Raphael, L. J., & Liberman, A. M. (1979). Some experiments on the sound of silence in phonetic perception. *Journal of the Acoustical Society of America*, *65*(6), 1518–1532.

Dorman, M. F., Studdert-Kennedy, M., & Raphael, L. J. (1977). Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. *Perception and Psychophysics*, *22*(2), 109–122.

Drager, K. (2011a). Speaker age and vowel perception. *Language and Speech*, *54*, 99–121.

Drager, K. K. (2011b). Sociophonetic variation and the lemma. *Journal of Phonetics*, *39*, 694–707.

Draxler, C. (1995). Introduction to the Verbmobil-PhonDat Database of Spoken German. In *Prolog Applications Conference*, Paris.

Eckert, P. (1989). The whole woman: Sex and gender differences in variation. *Language Variation and Change*, *1*(3), 245–267.

Eckert, P. (2000). *Language variation as social practice*. Wiley-Blackwell.

Esposito, A. (2002). On vowel height and consonantal voicing effects: Data from Italian. *Phonetica*, *59*, 197—-231.

Fahey, R. P. & Diehl, R. L. (1996). he missing fundamental in vowel height perception. *Perception and Psychophysics*, *58*, 725–733.

Fant, G. (1966). A note on vocal tract size factors and non-uniform f-pattern scalings. *Speech Transmission Laboratory Quarterly Progress and Status Report*, *1*, 22–30.

Fant, G. (1989). Non-uniform vowel normalization. *Speech Technology Laboratory: Quarterly Progress and Status Report*, *2*, 1–19.

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.

Ferguson, S. & Kewley-Port, D. (2007). Talker differences in clear and conversational speech: Acoustic characteristics of vowels. *Journal of Speech, Language, and Hearing Research*, *50*, 1241–1255.

Ferguson, S. H. (2004). Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners. *Journal of the Acoustical Society of America*, *116*, 2365–2373.

Fischer, R. M. & Ohde, R. N. (1990). Spectral and duration properties of front vowels as cues to final stop-consonant voicing. *The Journal of the Acoustical Society of America*, *88*, 1250–1259.

Fischer-Jørgensen, E. (1990). Intrinsic F0 in tense and lax vowels with special reference to German. *Phonetica*, *47*(3-4), 99–140.

Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *Journal of the Acoustical Society of America*, *84*(1), 115–123.

Fridland, V., Kendall, T., & Farrington, C. (2014). Durational and spectral differences in American English vowels: Dialect variation within and across regions. *Journal of the Acoustical Society of America*, *136*(1), 341–349.

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.

Fujisaki, H. & O Kunisaki, O. (1976). Analysis, recognition and perception of voiceless fricative consonants in Japanese. *Annual Bulletin Research Institute of Logopedics and Phoniatrics*, *10*, 145–156.

Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, *84*, 474–496.

Garrett, A. & Johnson, K. (2013). Phonetic bias in sound change . In A. C. L. Yu (Ed.), *Origins of Sound Change: Approaches to Phonologization* (pp. 51–97). Oxford: Oxford University Press.

Gay, T. (1978). Effect of speaking rate on vowel formant movements. *Journal of the Acoustical Society of America*, *63*(1), 223–230.

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.

Gordon, P. C., Eberhardt, J. L., & Rueckl, J. G. (1993). Attentional modulation of the phonetic significance of acoustic cues. *Cognitive Psychology*, *25*, 1–42.

Hagège, C. & Haudricourt, A.-G. (1978). *La phonologie panchronique*. Paris: Presses Universitaires de France.

Haggard, M., Ambler, S., & Callow, M. (1970). Pitch as a Voicing Cue. *Journal of the Acoustical Society of America*, *47*(2), 613–617.

Hajek, J. (1997). *Universals of sound change in nasalization*. Boston: Blackwell.

Hajek, J. & Maeda, S. (2000). *Vowel height and duration on the development of distinctive nasalization*. Cambridge: Cambridge University Press.

Halle, M. & Stevens, K. N. (1971). A note on laryngeal features. *Quarterly progress report*, *101*, 198–213.

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.

Hanson, H. M. (2009). Effects of obstruent consonants on fundamental frequency at vowel onset in english. *Journal of the Acoustical Society of America*, *125*, 425–441.

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. (2008). Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study. *Journal of the Acoustical Society of America*, *123*(5), 2825–2835.

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. Solé-Casals, & F. Morabito (Eds.), *In Recent Advances in Nonlinear Speech Processing* (pp. 15–34). Switzerland: Springer International Publishing.

Harris, K. S., Hoffman, H. S., Liberman, A. M., Delattre, P. C., & Cooper, F. S. (1958). Effect of Third-Formant Transitions on the Perception of the Voiced Stop Consonants. *The Journal of the Acoustical Society of America*, *30*(2), 122–126.

Haudricourt, A.-G. (1954). De l'origine des tons du vietnamien. *Journal asiatique*, *242*, 69–82.

Hay, J. & Drager, K. (2010). Stuffed toys and speech perception. *Linguistics*, *48*, 865–892.

Hay, J. & Foulkes, P. (2016). The evolution of medial /t/ over real and remembered time. *Language*. in press.

Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*, *34*, 458–484.

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.

Heinz, J. M. & Stevens, K. N. (1961). On the properties of voiceless fricative consonants. *Journal of the Acoustical Society of America*, *33*(5), 589–596.

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.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, *97*(5), 3099–3111.

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. (1976). The effect of aspiration on the fundamental frequency of the following vowelt. In *Proceedings of the 2nd Annual Meeting of the Berkeley Linguistics Society*, (pp. 212–219).

Hombert, J.-M. (1977). Consonant types, vowel height and tone in Yoruba. *Studies in African Linguistics*, *8*(2), 1–18.

Hombert, J.-M. (1978). Consonant types, vowel quality, and tone. In V. A. Fromkin (Ed.), *Tone: a Linguistic Survey* (pp. 77–111). New York: Academic Press.

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 control 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.

Honorof, D. N. & Whalen, D. H. (2005). Perception of pitch location within a speaker's F0 range. *Journal of the Acoustical Society of America*, *117*, 2193—-2200.

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.

Horvath, B. (1985). *Variation in Australian English: The Sociolects of Sydney*. Cambridge Univ. Press.

House, A. S. & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America*, *25*, 105–113.

Hughes, G. W. & Halle, M. (1956). Spectral properties of fricative consonants. *Journal of Acoustical Society of Ametica*, *28*, 303–310.

Hyman, L. M. (1976). Phonologization. In A. Juilland (Ed.), *Linguistic Studies Offered to Joseph Greenberg*, volume 2. Saratoge, 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.

Idemaru, K. & Holt, L. L. (2011). Word recognition reflects dimension-based statistical learning. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(6), 1939–1956.

Idemaru, K. & Holt, L. L. (2014). Specificity of dimension-based statistical learning in word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(3), 1009–1021.

Idemaru, K., Holt, L. L., & Seltman, H. (2012). Individual differences in cue weights are stable across time: The case of Japanese stop lengths. *Journal of the Acoustical Society of America*, *132*(6), 3950–3964.

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.

Johnson, k., Ladefoged, P., & Lindau, M. (1993). Individual differences in vowel production. *The Journal of the Acoustical Society of America*, *94*(2), 701–714.

Johnson, K., Strand, E. A., & D'Imperio, M. (1999). Auditory–visual integration of talker gender in vowel perception. *Journal of Phonetics*, *27*(4), 359–384.

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America*, *108*(3), 1252–1263.

Jun, S. A. (1993). *The Phonetics and Phonology of Korean prosody.* PhD thesis, 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 Emegernce 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. (2016a). VOT merger in Heritage Korean in Toronto. *Language Variation and Change*, *28*, 249–272.

Kang, Y. & Nagy, N. (2016b). 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.

Kawasaki, H. (1986). Phonetic explanation for phonological universals: The case for distinctive vowel

nasalization. In J. J. Ohala & J. J. Jaeger (Eds.), *Experimental Phonology* (pp. 81–103). Cambridge, MA: MIT Press.

Keating, P., Cho, T., & Fougeron, 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 tensity 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. (2004). Correlation between VOT and F0 in the Perception of Korean Stops and Affricates. In *INTERSPEEECH*, (pp. 1–6).

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. (2005a). The phonetics of Athabaskan tonogenesis. In H. Sharon & R. Keren (Eds.), *Athabaskan Prosody* (pp. 137–184). Amsterdam: Benjamins.

Kingston, J. (2005b). The phonetics of Athabaskan tonogenesis. In S. Hargus & K. Rice (Eds.), *Athabaskan Prosody* (pp. 137–184). Amsterdam: Benjamins.

Kingston, J. (2007). The phonetics-phonology interface. In P. de Lacy (Ed.), *The Cambridge Handbook of Phonology* (pp. 401–434). Cambridge Universisy.

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.

Kingston, J. & Diehl, R. L. (1994). Phonetic knowledge. *Language*, *70*, 419–454.

Kingston, J., Rich, S., Shen, A., & Sered, S. (2015). Is perception personal? In for ICPhS 2015, T. S. C. (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, UK. The University of Glasgow.

Kirby, J. (2010). *Cue selection and category restructuring in sound change.* PhD thesis, University of Chicago.

Kirby, J. (2014). Incipient tonogenesis in Phnom Penh Khmer: Acoustic and perceptual studies. *Journal of Phonetics*, *43*, 69–85.

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. (2016). Cross-linguistic variability in cue weighting of consonant voicing. In *The 15th Conference on Laboratory Phonology.* Cornell University, Ithaca, USA.

Kirby, J. P. & Ladd, D. R. (2016). Effects of obstruent voicing on vowel F0: Evidence from "true voicing" languagesa). *Journal of the Acoustical Society of America*, *140*(4), 2400–2411.

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.

Kleinschmidt, D. F. & Jaeger, T. F. (2015). Supplemental Material for Robust Speech Perception: Recognize the Familiar, Generalize to the Similar, and Adapt to the Novel. *Psychological review*, *203*.

Kluender, K. R. (1991). Effects of first formant onset properties on voicing judgments result from processes not specific to humans. *The Journal of the Acoustical Society of America*, *90*(1), 83–96.

Koenig, L. (2000). Laryngeal factors in voiceless consonant production in men, women, and 5-year-olds. *Journal of Speech, Language and Hearing Research*, *43*, 1211–1228.

Kohler, K. J. (1985). F0 in the perception of lenis and fortis plosives. *Journal of Acoustical Society of America*, *78*, 21–32.

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. (1996). *Intonational Phonology.* Cambridge: Cambridge University Press.

Ladd, D. R. & Silverman, K. E. A. (1984). Vowel intrinsic pitch in connected speech. *Phonetica*, *41*, 31–40.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and anovas. *Frontiers in psychology*, *4*, 863.

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.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1974). Perception of the speech code. *Psychological Review*, *27*, 431–461.

Liberman, A. M., Delattre, P. C., & Cooper, F. S. (1958). Some cues for the distinction between voiced and voiceless stops in initial position. *Language and Speech*, *1*(3), 153–167.

Lieberman, E., Michel, J.-B., Jackson, J., Tang, T., & Nowak, M. A. (2007). Quantifying the evolutionary dynamics of language. *Nature*, *449*(7163), 713–716.

Lin, S., Beddor, P. S., & Coetzee, A. W. (2014). Gestural reduction, lexical frequency, and sound change: A study of post-vocalic /l/. *Laboratory Phonology*, *5*(1), 9–36.

Lindblom, B. (1990). Explaining Phonetic Variation: A Sketch of the H&H Theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modeling* (pp. 403–439). Kluwer, Dordrecht: Springer Netherlands.

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. (1975). Is it vot or a first-formant transition detector? *Journal of the Acoustical Society of America*, *57*(6), 1547–1551.

Lisker, L. & Abramson, A. S. (1964). A Cross-language study of voicing in initial stops: acoustical measurements. *Word*, *20*, 384–422.

Lisker, L. & Abramson, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics. In *Proceedings of the 6th international congress of phonetic sciences*, (pp. 563–567). Academia Prague.

Lisker, L., Liberman, A. M., Erickson, D. M., Dechovitz, D., & Mandler, R. (1977). On pushing the voice-onset-time (vot) boundary about. *Language and Speech*, *20*(3), 209–216.

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.

Maddieson, I. (1984). *Patterns of Sounds*. Cambridge: Cambridge University Press.

Magnuson, J. S. & Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human perception and performance*, *33*, 391–409.

Manaster Ramer, A. (1986). Genesis of Hopi tones. *International Journal of American Linguistics*, *52*, 154–160.

Maniwa, K., Jongman, A., & Wade, T. (2009). Acoustic characteristics of clearly spoken English fricatives. *The Journal of the Acoustical Society of America*, *125*(6), 3962–3973.

Mann, V. A. (1980). Influence of preceding liquid on stop-consonant perception . *Perception and Psychophysics*, *28*(5), 407–412.

Mann, V. A. (1981). Influence of preceding fricative on stop consonant perception. *The Journal of the Acoustical Society of America*, *69*(2), 548–558.

Mann, V. A. & Repp, B. H. (1980). Influence of vocalic context on perception of the /sh/-/s/ distinction. *Perception and Psychophysics*, *28*(3), 213–228.

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.

Martin, A., Utsugi, A., & Mazukaac, R. (2014). The multidimensional nature of hyperspeech: Evidence from Japanese vowel devoicing. *Cognition*, *132*(2), 216–228.

Massaro, D. W. & Cohen, M. M. (1976). The contribution of fundamental frequency and voice onset time to the/zi/-/si/distinction. *Journal of the Acoustical Society of America*, *60*, 704–717.

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.

Mayo, C. & Turk, A. (2004). Adult–child differences in acoustic cue weighting are influenced by segmental context: Children are not always perceptually biased toward transitions. *Journal of the Acoustical Society of America*, *115*(6), 3184–3194.

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.

McMurray, B., Clayards, M. A., Tanenhaus, M. K., & Aslin, R. N. (2008). Tracking the time course of phonetic cue integration during spoken word recognition. *Psychonomic Bulletin & Review*, *15*(6), 1064–1071.

McMurray, B. & Jongman, A. (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological review*, *118*(2), 219–246.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F. (2017). *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien*. R package version 1.6-8.

Mikuteit, S. & Reetz, H. (2007). Caught in the act: The timing of aspiration and voicing in east bengali. *Language and Speech*, *50*, 247–277.

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.

Miller, J. L. & Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception and Psychophysics*, *25*(6), 457–465.

Miller, J. L. & Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception and Psychophysics*, *46*(6), 505–512.

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.

Mitterer, H. (2006). On the causes of compensation for coarticulation: Evidence for phonological mediation. *Perception and Psychophysics*, *68*, 1227—-1240.

Morgan, J., LaRocca, S., Bellinger, S., & Ruscelli, C. (2005). West Point Company G3 American English Speech LDC2005S30. Philadelphia. Linguistic Data Consortium.

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.

Morrongiello, B. A., Robson, R. C., Best, C. T., & Clifton, R. K. (1984). Trading relations in the perception of speech by 5-year-old children. *Journal of Experimental Child Psychology*, *37*(2), 231–250.

Munson, B. (2007). The acoustic correlates of perceived masculinity, per- ceived femininity, and perceived sexual orientation. *Language and Speech*, *50*, 125–142.

Munson, B. & Solomon, N. P. (2004). The effect of phonological neighborhood density on vowel articulation. *Journal of speech, language, and hearing research*, *47*(5), 1048–1058.

Nearey, T. M. (1978). *Phonetic feature systems for vowels.* PhD thesis, Indiana University.

Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America*, *85*, 2088–2113.

Nearey, T. M. (1997). Speech perception as pattern recognition. *Journal of the Acoustical Society of America*, *101*, 3241–3254.

Newman, R. S., Clouse, S. A., & Burnham, J. L. (2001). The perceptual consequences of within-talker variability in fricative production. *Journal of the Acoustical Society of America*, *109*(3), 1181–1196.

Nolan, F. (2003). Intonational equivalence: an experimental evaluation of pitch scales. In Solé, M.-J., Recasens, D., & Romero, J. (Eds.), *the XVth 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.

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 Otolaryngolical*, *33*, 87—-108.

Peterson, G. E. & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical of America*, *24*(2), 175–184.

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.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing II: acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research*, *29*, 434–446.

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.

Pisoni, D. & Luce, P. (1987). Acoustic-phonetic representation in word recognition. In U. Frauenfelder & L. Tyler (Eds.), *Spoken word recognition* (pp. 21–52). Cambridge, MA: MIT Press.

Port, R. F. & Dalby, J. (1982). Consonant/vowel ratio as a cue for voicing in English. *Perception and Psychophysics*, *32*(2), 141–152.

Purcell, E., Villegas, G., & Young, S. (1978). A before and after for tonogenesis. *Phonetica*, *35*(5), 284–293.

Raphael, L. J. (2004). Acoustic Cues to the Perception of Segmental Phonemes. In B. D. Pisoni & E. R. Remez (Eds.), *The Handbook of Speech Perception* (pp. 182–206). Malden, MA: Blackwell Publishing Ltd.

Repp, B. H. (1979). Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. *Language and Speech*, *22*(2), 173–189.

Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin*, *92*(1), 81–110.

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 Bensamin Publishing Company.

Rhodes, R. A. (1992). Flapping in american english. In *Proceedings of the 7th International Phonology Meeting*, (pp. 217–232)., Turnin. Rosenberg and Sellier.

Rhodes, R. A. (1996). English reduced vowels and the nature of natural processes. In B. Hurch & R. A. Rhodes (Eds.), *Spoken word recognition* (pp. 239–259). Berlin: Mouton du Gruyter.

Robb, M., Gilbert, H., & Lerman, J. (2005). Influence of gender and environmental setting on voice onset time. *Folia Phoniatrica et Logopaedica*, *57*(3), 125–133.

Roubeau, B., Chevrie-Muller, C., & Saint Guily, J. L. (1972). Electromyographic activity of strap and cricothyroid muscles in pitch change. *Acta Otolaryngolical*, *117*, 459–464.

Ryalls, J. H., Zipprer, A., & Baldauff, P. (1997). A preliminary investigation of the effects of gender and race on voice onset time. *Journal of Speech, Language and Hearing Research*, *40*, 642–645.

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.

Sapir, S. (1989). The intrinsic pitch of vowels: theoretical, physiological, and clinical considerations. *Journal of Voice*, *3*(1), 44–51.

Schertz, J. (2013). Exaggeration of featural contrasts in clarifications of misheard speech in English. *Journal of Phonetics*, *41*, 249–263.

Schertz, J., Cho, T., Lotto, A., & Warner, N. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics*, *52*, 183–204.

Scobbie, J. M. (2006). Flexibility in the face of incompatible English VOT systems. *Laboratory Phonology*, *8*, 367–392.

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 *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.

Simpson, A. P. (2009). Phonetic differences between male and female speech. *Language and Linguistics Compass*, *3*(2), 621–640.

Smiljanic, R. & Bradlow, A. R. (2009). Speaking and Hearing Clearly: Talker and Listener Factors in Speaking Style Changes. *Language and Linguistics Compass*, *3*(1), 236–264.

Smith, B. L. (1978). Effects of place of articulation and vowel environment on. *Glossa*, *12*(2), 163–75.

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.

Solé, M.-J. & Ohala, J. (2010). What is and what is not under the control of the speaker: Intrinsic vowel duration. In C. Fougeron, B. Kühnert, M. D'Imperio, & N. Vallee (Eds.), *Papers in Laboratory Phonology*, volume 10 (pp. 607–655). Berlin: de Gruyter.

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.

Sonderegger, M., Bane, M., & Graff, P. (2017). The medium-term dynamics of accents on reality television. *Language*, *93*(3), 598–640.

Sonderegger, M. & Keshet, J. (2012). Automatic measurement of voice onset time using discriminative structured prediction. *Journal of the Acoustical Society of America*, *132*, 3965–3979.

Sóskuthy, M. (2015). Understanding change through stability: A computational study of sound change actuation. *Lingua*, *163*, 40–60.

Staum-Casasanto, L. (2009). *Experimental investigations of sociolinguistic knowledge.* PhD thesis, Stanford University.

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, K. N. & House, A. S. (1955). Development of a quantitative description of vowel articulation. *Journal of the Acoustical Society of America*, *27*, 484–493.

Stevens, K. N. & Klatt, D. H. (1974). Role of formant transitions in the voiced-voiceless distinction for stops. *Journal of the Acoustical Society of America*, *55*, 653–659.

Stevens, M. & Harrington, J. (2014). The individual and the actuation of sound change. *Loquens*, *1*, e003.

Strand, E. & Johnson, K. (1996). Gradient and visual speaker normalization in the perception of fricatives. In D. Gibbon (Ed.), *Natural Language Processing and Speech Technology* (pp. 14–26). Germany: Mouton de Gruyter.

Strand, E. A. (1999). Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Psychology*, *18*, 86–99.

Strange, W. (1989). Evolving theories of vowel perception. *Journal of the Acoustical Society of America*, *85*, 2081—-2087.

Stuart-Smith, J. (2007). Empirical evidence for gendered speech production:/s/in glaswegian. *Laboratory phonology*, *9*, 65–86.

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.

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of experimental psychology. Human perception and performance*, *7*(5), 1074–1095.

Summerfield, Q. & Haggard, M. (1977). On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *Journal of the Acoustical Society of America*, *62*(2), 435–448.

Sumner, M. (2014). The socially weighted encoding of spoken words: a dual-route approach to speech perception. *Frontiers in Psychology*, *4*, 1–13.

Sussman, H. M. & Shore, J. (1996). Locus equations as phonetic descriptors of consonantal place of articulation. *Perception and Psychophysics*, *58*(6), 936–946.

Svantesson, J.-O. & House, D. (2006). Tone production, tone perception and Kammu tonogenesis. *Phonology*, *23*(02), 309–333.

Swartz, B. L. (1992). Gender difference in voice onset time. *Perceptual and Motor Skills*, *75*(5), 983–992.

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

Theodore, R. M. & Miller, J. L. (2010). Characteristics of listener sensitivity to talker-specific phonetic detail. *Journal of the Acoustical Society of America*, *128*(4), 2090–2099.

Theodore, R. M., Myers, E. B., & Lomibao, J. A. (2015). Talker-specific influences on phonetic category structurea). *The Journal of the Acoustical Society of America*, *138*(2), 1068–1078.

Titze, I. R. (1989). Physiologic and acoustic differences between male and female voices. *Journal of the Acoustical Society of America*, *85*, 1699—-1707.

Titze, I. R. (1994). Toward standards in acoustic analysis of voice. *Journal of Voice*, *8*, 1–7.

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.

Venables, W. N. & Ripley, B. D. (2002). *Modern Applied Statistics with S* (Fourth ed.). New York: Springer. ISBN 0-387-95457-0.

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.

Weismer, G. (1979). Sensitivity of voice-onset time (vot) measures to certain segmental features in speech production. *Journal of Phonetics*, *7*, 197–204.

Whalen, D. H., Abramson, A. S., Lisker, L., & Mody, M. (1990). Gradient effects of fundamental frequency on stop consonant voicing judgments. *Phonetica*, *47*, 36–49.

Whalen, D. H., Abramson, A. S., Lisker, L., & Mody, M. (1993). F0 gives voicing information even with unambiguous voice onset times. *Journal of the Acoustical Society of America*, *93*(4), 2152–2159.

Whalen, D. H., Gick, B., Kumada, M., & Honda, K. (1999). Cricothyroid activity in high and low vowels: exploring the automaticity of intrinsic F0. *Journal of Phonetics*, *27*(2), 125–142.

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.

Whiteside, S. P. & Irving, C. J. (1998). Speakers' sex differences in voice onset time: a study of isolated word production. *Perceptual and motor skills*, *86*(2), 651–654.

Winn, M. B., Chatterjee, M., & Idsardi, W. J. (2013). Roles of Voice Onset Time and F0 in Stop Consonant Voicing Perception: Effects of Masking Noise and Low-Pass Filtering. *Journal of Speech and Hearing Research*, *56*(4), 1097–1107.

Wolff, E. (1987). Consonant-tone interference in chadic and its implications for a theory of tonogenesis in afroasiatic. In D. Barreteau (Ed.), *Langues et cultures dans le bassin du Lac Tchad* (pp. 193–216). Paris: ORSTOM.

Wright, J. D. (2007). *Laryngeal contrast in Seoul Korean*. PhD thesis, University of Pennsylvania.

Wurm, L. H. & Fisicaro, S. A. (2014). What residualizing predictors in regression analyses does (and what it does not do). *Journal of Memory and Language*, *72*, 37–48.

Yao, Y. (2009). Understanding VOT variation in spontaneous speech. In M. Park (Ed.), *Current numbers in unity and diversity of languages* (pp. 1122–1137). Seoul: Linguistic Society of Korea.

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. (2015). A corpus-based study on the layered duration in Standard Korean. In M. Kenstowicz, T. Levina, & R. Masuda (Eds.), *Japanese/Korean Linguistics 23*. Chicago: The 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 Scie*, *6*, 139–145.

Yu, A. C. L. (2013). Individual differences in socio-cognitive processing and the actuation of sound change. In A. C. L. Yu (Ed.), *Origins of sound change: Approaches to phonologization* (pp. 201–227). Oxford: Oxford University Press.

Yu, A. C. L., Abrego-Collier, C., & Sonderegger, M. (2013). Phonetic imitation from an individual–difference perspective: Subjective attitude, personality and "autistic" traits. *PLoS ONE*, *8*(9), e74746–274746.

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 *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.