CONTRASTIVE HYPERARTICULATION OF WORD-INITIAL STOPS IN A CONVERSATIONAL CORPUS

by

Noah Richard Nelson

Copyright © Noah Richard Nelson 2019

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF LINGUISTICS

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2019

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Noah Nelson, titled Contrastive Hyperarticulation of Word-Initial Stops in a Conversational Corpus and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Andrew Wedel — Date: Nov 22 '19

Adam Ussishkin — Date: 11/22/19

Michael Hammond — Date: 11/22/19

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Date: Nov 22 '19

Andrew Wedel
Dissertation Committee Chair
Linguistics Department

Acknowledgements

There are a great many people who deserve thanks and acknowledgment for helping to make this dissertation a reality. First and foremost, I wish to thank my wife, Elly Zimmer, whose patience, love, and support was instrumental in maintaining my mental and emotional health. I also wish to extend my gratitude to my advisor, Andy Wedel, who has been not only a collaborator and a mentor, whose guidance has always driven me to be a better linguist, but who has also been a dear friend. And of course, I also wish to thank Adam Ussishkin, who has always been both a friend and a supportive professional mentor, and Michael Hammond, who has always been sure to tend first to my goals and my needs, and then to my research. Finally, I wish to thank my family, one and all, for encouraging me to follow my path and helping me to reach the finish line in one, happy piece.

I also wish to acknowledge the help and contributions of many others who have helped me through discussion, critical feedback, and emotional support. First, I want to extend my appreciation to all of my fellow graduate students in the Department of Linguistics at the University of Arizona, who have always been supportive of one another. I especially want to acknowledge the other members of my cohort who participated in our "prelim" course, who patiently helped me to formulate the paper that would ultimately lead to this dissertation. Second, I want to thank the many faculty of the Department of Linguistics at the University of Arizona, who helped to shape me into the researcher I am today. I especially want to acknowledge, in addition to my committee members, Amy Fountain and Natasha Warner, for feedback and insights directly related to this dissertation; Diane Ohala, for always being available and supportive; Masha Fedzechkina, for teaching me skills I might otherwise lack and for her professional guidance; and Heidi Harley, for introducing me to the wonderful Hiaki language and her wonderful Hiaki language consultants, and for her patience and guidance as my minor advisor throughout my years of study.

Finally, I wish to extend a special thank-you to the fantastic folks at the assorted coffee shops where I've spent so much of my time, employees and regulars alike - I think I wrote more of this dissertation in their midst than anywhere else!

TABLE OF CONTENTS

TABLE OF CONTENTS—*Continued*

TABLE OF CONTENTS—*Continued*

List of Figures

List of Tables

LIST OF FIGURES—*Continued*

Abstract

In this dissertation, I seek to clarify the nature of *contrastive hyperarticulation*. Unlike other kinds of hyperarticulation found in the literature, such as clear-speech hyperarticulation or the hyperarticulation of Lindblom's (1990) Hyper-articulation and Hypo-articulation (H & H) Theory, contrastive hyperarticulation is defined not in terms of articulatory durations or gestural extents, but rather in terms of phonetic realizations that enhance the perceptual distance between the target and competing lexical forms.

Despite a growing number of studies over the past decade that have investigated what might be described as contrastive hyperarticulation, the phenomenon is still not very well defined or understood. For example, some researchers have suggested that the total number of lexical-phonological neighbors to a given target word is correlated with hyperarticulation of individual phonetic cues in that target word (e.g., Fox et al., 2015). Other researchers, however, have suggested that only the existence of the specific lexical-phonological neighbor that contrasts with the target word in the measured phonetic cue will correlate with hyperarticulation of that phonetic cue (e.g., Wedel et al., 2018). In addition to uncertainty regarding the kinds of lexical competition that trigger contrastive hyperarticulation, researchers have also suggested different phonetic outcomes of this competition-driven hyperarticulation. Some researchers have suggested that, as a form of hyperarticulation, contrastive hyperarticulation can only be realized in the form of phonetic enhancement (e.g., Goldrick et al., 2013; Fricke et al., 2016). Other researchers, however, have suggested that the phonetic outcomes of contrastive hyperarticulation can include instances of phonetic reduction (such as decreased phonetic durations) in addition to more traditionally defined hyperarticulation (such as increased phonetic durations; Seyfarth et al., 2016; Wedel et al., 2018).

In this dissertation I clarify both the triggers and outcomes of contrastive hyperarticulation in two studies. The first study was designed to sample a large hypothesis space regarding the kinds of lexical competitors that might trigger contrastive hyperarticulation of word-initial stop voicing contrasts in spontaneous English speech. Contrastive hyperarticulation was identified in the form of enhanced voice onset time contrasts between voiced and voiceless stops (longer voice onset times in voiceless stops, and shorter voice onset times in voiced stops). A number of possible metrics of lexical competition were evaluated as predictors of these voice onset time realizations based on the literature, including overall neighborhood density, the existence of the initial stop voicing minimal pair competitor, and a number of metrics varying in the position or type of competition between target and competitor. It was found that the existence of the initial stop voicing minimal pair competitor was a better predictor of voice onset time measurements than any other metric of lexical competition for nearly all statistical models. The outcomes of contrastive hyperarticulation were also partially addressed in this study in that voiced stops in words with voiceless stop minimal pairs were realized with shorter voice onset times than voiced stops in words without such minimal pairs. This is indicative of shorter phonetic durations that are used to

enhance a lexical-phonological contrast (see also Seyfarth et al., 2016; Wedel et al., 2018).

The second study was designed to further investigate the phonetic outcomes of contrastive hyperarticulation by measuring a secondary cue to the lexical-phonological contrast in question. Specifically, word-initial stop voicing in English is primarily signaled by voice onset time, but is also signaled by the fundamental frequency that follows the stop at the onset of voicing. In this study, I investigated fundamental frequency measurements following voiced and voiceless stops to see whether they are also contrastively hyperarticulated as a function of stop voicing minimal pair competitor existence. It was found that fundamental frequency correlated positively with voice onset time in voiceless stops of words with initial voiced stop minimal pairs, but this correlation did not hold for words without stop voicing minimal pairs. In other words, fundamental frequency appears to be selectively hyperarticulated alongside the primary voice onset time cue only for voiceless stops in words with voiced stop minimal pairs. This suggests that the effect is not intrinsic, but rather is selectively applied in cases where the phonetic contrast will serve a lexical-phonological contrast.

The results of these studies were taken to support the notion that contrastive hyperarticulation is driven by lexical-phonological competition that is rooted in phonetic contrast. As voice onset time is a major cue to initial stop voicing in English, it is recruited contrastively in words with initial stop voicing minimal pairs. As fundamental frequency, however, is an ancillary cue, it is recruited more selectively in service of the lexical-phonological contrast. These results are considered alongside other results related to contrastive hyperarticulation, including results showing that multiple cues to a given lexical-phonological contrast can be hyperarticulated contrastively, with both cues moving in opposing directions on either side of the lexical-phonological contrast (Seyfarth et al., 2016).

CONTRIBUTIONS OF AUTHORS

This dissertation consists of two primary studies. Study 1 (Chapter 2) has been previously published in the *Journal of Phonetics* as a co-authored article entitled ‘The phonetic specificity of competition: Contrastive hyperarticulation of voice onset time in conversational English’ (Nelson and Wedel, 2017; this published version of the article can be found in Appendix C). This article was written with Prof. Andrew Wedel, and I am the primary author of the manuscript in both its published form (Appendix C) and as Chapter 2 of this dissertation. The voice onset time data used in Study 1 were originally acquired for another published article of which I am the second author (Wedel et al., 2018); those data involved hand-annotation of voice onset time by Prof. Wedel and undergraduate assistants, and I was not involved in those annotations. Apart from those already-available data, I was primarily responsible for conceptualization of the research questions, other data collection and analysis, and the bulk of the writing for Study 1, all of which was done in consultation with Prof. Wedel. Additional feedback on written drafts at various stages of preparation came from Profs. Adam Ussishkin, Michael Hammond, and Amy Fountain.

Study 2 (Chapter 3) is being prepared for submission for publication. I am either primarily or solely responsible for the conceptualization of the research questions, acquisition of all fundamental frequency data, formulation of the lexicon used to determine minimal pair relationships, visual and statistical analyses, and writing of the manuscript. This was all done in consultation with Prof. Andrew Wedel, with additional insight on acquiring and manipulating the data from Prof. Natasha Warner. Profs. Michael Hammond and Adam Ussishkin provided feedback on written drafts.

The general introduction (Chapter 1) and general discussion (Chapter 4) were written by me, with feedback from Profs. Wedel, Hammond, and Ussishkin.

CHAPTER 1

GENERAL INTRODUCTION

Spoken language, like any human behavior, is inherently variable. Even within a given language, there is variability in the form of different regional dialects, preferred sentence structures, word and phrase choices, and even pronunciation of particular words. At an even more fundamental level, there is variability in the acoustic speech signals that speakers produce and listeners must interpret. Not only do individual speakers vary due to different articulatory anatomy, speaking style, or any number of social factors, but a given speaker is unlikely to ever produce two tokens of the same word in precisely the same way. Low-level variability of this sort is expected in any behavior, especially behaviors like speech production, which requires coordination of a complex system of articulators. It is therefore no surprise that, for many linguists, fine-grained phonetic variability has been treated as separate and independent from the grammar of a language.

But to what extent are portions of the phonetic variability found in speech attributable to systematic properties of language, words, or the speech production system? It has been suggested that attributes of "higher order" levels of the linguistic system, such as the lexicon or syntactic structure, can indeed impact the "lower order" phonetic realizations of speech targets. For example, it has been shown that some lexical-level properties of words can systematically influence the phonetic realizations of those words. Sometimes referred to as lexical-phonetic interactions (e.g., Baese-Berk and Goldrick, 2009), these phenomena include a wide variety of lexical properties otherwise assumed to be independent of the phonetics or phonology of the word itself. Some of the lexical properties that have been shown to affect phonetic realizations of words range from lexical frequency (e.g., Bell et al., 2009; Sóskuthy and Hay, 2017) to lexical-phonological relationships such as neighborhood density (e.g., Luce and Pisoni, 1998; Gahl et al., 2012; Fox et al., 2015; but see Gahl, 2015; Wedel et al., 2018) and minimal pair existence (e.g., Baese-Berk and Goldrick, 2009; Schertz, 2013; Buz et al., 2016). These lexical properties can affect phonetic realizations in a number of ways, including production onset latencies (e.g., Oldfield and Wingfield, 1965; Buz et al., 2014), reduction of phonetic content (e.g., Bell et al., 2009; Ernestus and Warner, 2011), and hyperarticulation (e.g., Baese-Berk and Goldrick, 2009; Fox et al., 2015; Wedel et al., 2018).

Hyperarticulation has proven to be of particular interest in the field, with a large number of publications investigating hyperarticulation emerging over the past few decades (Goldinger and Summers, 1989; Wright, 1997; Munson, 2007; Baese-Berk and Goldrick, 2009; Peramunage et al., 2011; Schertz, 2013; Fox et al., 2015; Wedel et al., 2018, to name but a few). However, "hyperarticulation" in this literature has not always been clearly defined, and studies investigating hyperarticulation phenomena have varied considerably in what form of hyperarticulation they investigate. This broad use of the term may be due, in part, to an attempt to achieve continuity between disparate research programs. For example, one line of research has studied hyperarticulation of individual words or segments as a function of competition from within the lexicon (e.g., Munson and Solomon, 2004; Munson, 2007; Baese-Berk and Goldrick, 2009; Fox et al., 2015; Buz and Jaeger, 2016; Seyfarth et al., 2016; Wedel et al., 2018). The general findings of these studies have indicated that the existence of phonologically similar alternatives to the target word affect the phonetic realization of that word. By contrast, another prolific research program has studied a form of hyperarticulation in a number of languages referred to as *clear-speech hyperarticulation*. Clear speech is often produced by speakers under circumstances of perceived decreased intelligibility, such as when addressing someone who is hard of hearing, or under conditions of substantial environmental noise (e.g., Smiljanić and Bradlow, 2008). Such clear speech has been reported to involve overall slower speech rates, more dynamic use of pitch, and expanded vowel spaces, among other phenomena (Smiljanić and Bradlow, 2008). Though undoubtedly a form of hyperarticulation, clear speech is not believed to be conditioned by lexical properties or competition relationships, making it potentially distinct from the competition-driven hyperarticulation described above.

These differences between hyperarticulation associated with lexical-phonological competition and hyperarticulation associated with clear speech have been muddied by the fact that researchers often refer to both simply as "hyperarticulation". What is more, early researchers studying lexically conditioned variation made deliberate attempts to link their research with research of clear speech phenomena (e.g., Wright, 2004, pg. 79). In addition, a growing line of research has studied hyperarticulation of words or segments under conditions of misperception, where speakers may be presumed to actively hyperarticulate the target for clarity (e.g., Ohala, 1994; Schertz, 2013; Buz and Jaeger, 2016). In these studies, speakers are led to believe that certain productions are misperceived, and are therefore prompted to hyperarticulate, either explicitly (Ohala, 1994; Schertz, 2013) or implicitly (Buz and Jaeger, 2016). The specific misperceptions, however, are limited to similar phonological alternatives, usually (but not exclusively) lexical minimal pairs. Consequently, such research bridges a gap between hyperarticulation that is lexically conditioned and that which is the result of perceived communication difficulty (like clear speech). However, such research may also further obscure the potentially different sources of these two kinds of hyperarticulation.

It has been suggested that researchers should distinguish between different kinds of hyperarticulation phenomena (e.g., Schertz, 2013; Wedel et al., 2018; see also Warner, 2011 for discussion of the many ways speech styles can lead to variation that is orthogonal to hyperarticulation). Critical to this agenda is clearly outlining the different kinds of hyperarticulation found in human speech, and clearly defining the nature of each type of hyperarticulation in terms of both the sources that trigger it and the outcomes that result from it. The existing literature on clear-speech hyperarticulation has gone a long way toward achieving this goal for that particular kind of hyperarticulation. For example, clear-speech hyperarticulation has been suggested as a response to perceived communication difficulties (the context that triggers clear-speech hyperarticulation) whereby higher order perceptual properties of speech are targeted, such as overall speaking rate and amplitude (the outcomes of clear-speech hyperarticulation).

Other kinds of hyperarticulation, however, are much less clearly defined. As noted earlier, these other kinds of hyperarticulation include hyperarticulation presumed to be in response to lexical competition. Though “lexical competition”, broadly defined, is the suggested trigger of this kind of hyperarticulation, research into this phenomenon has varied widely in terms of the kinds of lexical competition purported to trigger hyperarticulation. For example, a number of researchers have suggested that lexical-phonological neighborhood density, i.e., the number of words in the lexicon that differ from the target word by the addition, subtraction, or substitution of a single phoneme, is correlated with hyperarticulation of particular phonetic cues (e.g., Munson and Solomon, 2004; Munson, 2007; Scarborough, 2013; Fox et al., 2015). Other researchers have suggested that the number of lexical-phonological neighbors for a given segmental position in a word correlates with hyperarticulation of phonetic cues within that segmental position (e.g., Vitevitch and Chu, 2004; Goldrick et al., 2010; Fricke, 2013; Fricke et al., 2016). Yet other researchers have suggested that hyperarticulation of a particular phonetic cue is correlated with the existence of the specific lexical-phonological minimal pair competitor defined by that phonetic cue (e.g., Baese-Berk and Goldrick, 2009; Schertz, 2013; Buz et al., 2016; Seyfarth et al., 2016; Wedel et al., 2018). While these three possibilities are not mutually exclusive, many of these researchers have argued in favor of one approach over others (e.g., in favor of neighborhood density over minimal pair relationships: Fox et al., 2015; in favor of segment-specific neighborhood density over overall neighborhood density or minimal pair relationships: Fricke et al., 2016; in favor of minimal pair relationships over neighborhood density: Wedel et al., 2018).

Clearly, the precise kinds of lexical competition that lead to hyperarticulation are not yet well understood. But the results of this hyperarticulation are not well understood either. For many researchers, hyperarticulation of any kind, including that which is driven by lexical competition, can only lead to increased phonetic durations or articulatory extents (e.g., Wright, 2004; Fricke, 2013). Other researchers, however, have suggested that lexical competition can lead to *contrastive* enhancement of the phonetic cues that distinguish the target from its competitors (e.g., Seyfarth et al., 2016; Wedel et al., 2018). Furthermore, researchers looking for any kind of competition-driven hyperarticulation have focused on *primary* phonetic cues to the lexical-phonological contrasts in question (though see Goldrick et al., 2013; Schertz, 2013; Seyfarth et al., 2016). That is, researchers have focused on the cues considered most salient or critical to the phonemic contrast under investigation, and have not considered other phonetic cues that may contribute to the contrast. For example, many researchers have investigated effects of various kinds of lexical competition on the realization of word-initial voice onset time (VOT) in English (e.g., Baese-Berk and Goldrick, 2009; Schertz, 2013; Fricke, 2013; Fox et al., 2015; Wedel et al., 2018, to name a few). VOT is considered the primary cue to the word-initial stop consonant voicing contrast in English in that it is the cue that speakers use most reliably in production and that listeners rely on most consistently in perception (Abramson and Lisker, 1985; Lisker, 1986). However, VOT is not the only cue, as stop voicing seems to affect a variety of acoustic properties of the signal, such as fundamental frequency (F0) in, and overall duration of, surrounding vowels (e.g., Lisker, 1986; Clayards, 2018). Though these other cues have been included in investigations of word-initial stop voicing, to my knowledge they have not been included in investigations of hyperarticulation resulting from lexical competition.

1.1 The Goals and Structure of the Dissertation

In this dissertation, I set out to clarify the triggers and outcomes of a particular kind of lexically conditioned hyperarticulation which I refer to as "contrastive hyperarticulation" (see also Schertz, 2013; Wedel et al., 2018, for the same or similar terminology). Contrastive hyperarticulation stands apart from other kinds of hyperarticulation in that it is specifically aimed at enhancing phonetic cues involved in signaling lexical-phonological contrasts (for a related formalization, see Hall, 2011. For discussion of the importance of meaning, and thereby the lexical nature of these lexical-phonological contrasts, see Hall et al., 2016). That is, contrastive hyperarticulation affects phonetic realizations of sub-lexical units (such as phonemes) in ways that increase the perceptual distance between competing lexical-phonological neighbors[1].

This description of contrastive hyperarticulation is distinct from more conventional descriptions of hyperarticulation, such as that described in Lindblom (1990) under the Hyper- and Hypo-articulation (H & H) Theory. Under such accounts, hyperarticulation generally takes the form of increased phonetic durations or articulatory extents (e.g., Picheny et al., 1986; De Jong, 1995; Bradlow, 2002; Ferguson and Kewley-Port, 2007; Smiljanić and Bradlow, 2008; Cho et al., 2011; Granlund et al., 2012). Contrastive hyperarticulation, however, is defined in terms of phonetic contrast rather than durations or gestural extents. As such, it is a priori possible for reduced phonetic forms to serve to enhance a lexical-phonological contrast. For example, Seyfarth et al. (2016) found that speakers contrastively hyperarticulated word-final voicing contrasts through reduced vowel durations (among other things). The use of shorter durations, something generally characterized as a form of reduction, to increase a lexical-phonological contrast is indicative of the contrastive nature of this hyperarticulation. I propose that such contrastive hyperarticulation is distinct from other kinds of hyperarticulation described in the literature, such as clear-speech hyperarticulation. Furthermore, I propose that the kind of contrastive hyperarticulation under investigation here may also be distinct from more online, intentional, communicatively-driven hyperarticulation of the type reported in many laboratory studies (see especially Schertz, 2013; Buz et al., 2016; Seyfarth et al., 2016; Fricke et al., 2016), though it need not be so (for more related discussion, see Hall et al., 2016; Wedel et al., 2018).

[1] It should be noted that the concept of 'perceptual distance' here is theoretical. In most research investigating contrastive hyperarticulation, what is actually measured is some form of phonetic-acoustic distance.

1.1.1 Goals

The primary goals of this dissertation are twofold:

1. To clarify the *triggers* of contrastive hyperarticulation by investigating the degree of phonetic similarity between lexical-phonological neighbors that leads to contrastive hyperarticulation of a particular phonetic cue

2. To clarify the *outcomes* of contrastive hyperarticulation by investigating whether and how a non-primary phonetic cue to a lexical-phonological contrast is contrastively hyperarticulated

Of particular interest in the present work is how contrastive hyperarticulation is realized in spontaneous, conversational speech. There are two primary reasons for this. First, it has been argued that certain kinds of phonetic variation are more readily identifiable under more natural conditions where speech is more likely to be reduced, as opposed to laboratory elicitation settings that may lead to more generalized hyperarticulation (Wedel et al., 2018). Second, contrastive hyperarticulation of the type under study here has been implicated in processes of long-term sound change affecting the phoneme inventories of languages (Wedel et al., 2013b). In order for contrastive hyperarticulation to affect real world phoneme inventories, it is hypothesized (at least under some theoretical models) that it must be present under "normal" day-to-day speaking conditions (Wedel et al., 2018).

1.1.2 Structure

To these ends, this dissertation consists of two studies of spontaneous speech from the Buckeye Corpus of Conversational English (Pitt et al., 2005, 2007). Both studies have been written so as to be able to stand alone as intelligible and comprehensive works. As such, there is some redundancy in their background, methodological descriptions, and discussion, but this redundancy serves to remind the reader of the relevant details as they emerge.

Study 1, presented in Chapter 2, addresses the "triggers" of contrastive hyperarticulation[2]. In this study, I investigate the kinds of lexical competition that lead to contrastive hyperarticulation of voice onset time (VOT), the primary cue to word-initial stop voicing contrasts in English (Lisker and Abramson, 1964; Lisker, 1986).

Study 2, presented in Chapter 3, addresses the outcomes of contrastive hyperarticulation. In this study, I look for evidence of contrastive hyperarticulation of fundamental frequency (F0), a secondary cue to the word-initial stop voicing contrast in English (Lisker and Abramson, 1964; Lisker, 1986).

The results of these two studies are interpreted together in Chapter 4 as preliminary evidence of the nature of contrastive hyperarticulation in relatively conversational speech contexts. Though the current introductory chapter (Chapter 1) and the general discussion (Chapter 4) cover the highlights of the relevant background and implications, more detailed background and discussions can be found within each study's respective chapter.

[2]This study has been previously published in the *Journal of Phonetics* (Nelson and Wedel, 2017) and the published article is included as supplementary material in Appendix C.

Chapter 2

Study 1: The triggers of contrastive hyperarticulation

2.1 Introduction and Background

A number of experimental and observational studies have reported that competition at the lexical level is correlated with hyperarticulation of phonetic properties in the target word. This correlation has been reported for English in a number of studies investigating the realization of vowel formants (e.g., Wright, 1997, 2004; Munson, 2007; Scarborough, 2012), vowel durations (Schertz, 2013; Seyfarth et al., 2016; but see Goldrick et al., 2013), degree of coarticulation (Scarborough, 2012, 2013), perseveration of voicing in coda fricatives (Kharlamov, 2014; Seyfarth et al., 2016), and initial stop voice onset time (Baese-Berk and Goldrick, 2009; Peramunage et al., 2011; Kirov and Wilson, 2012; Fricke, 2013; Schertz, 2013; Fox et al., 2015; Buz et al., 2016; Fricke et al., 2016). In each of these cases, some form of lexical competition has been found to correlate with hyperarticulation of phonetic properties of individual segments (e.g., Wright, 2004; Fricke, 2013; Buz et al., 2016; but see Goldrick et al., 2013; Gahl, 2015).

This work on competition-associated hyperarticulation is complicated, however, by the fact that 'competition' can be operationalized in a variety of ways. The two most common approaches to operationalizing competition are in terms of lexical-phonological *neighbor-*

hood density, defined as the number of words that can be formed by adding, deleting, or substituting any single segment of the target word (often weighted for frequency; e.g., Luce and Pisoni, 1998), or in terms of *minimal pair relationships* defined for a specific phonetic cue (e.g., Baese-Berk and Goldrick, 2009). These two general approaches differ widely in terms of the specificity of the competition measure. For neighborhood density measures, competition anywhere in the word contributes to the measure, regardless of the phonetic relationship between the target and the competitor (though see Strand and Sommers, 2011, and Gahl and Strand, 2016, for a version of neighborhood density weighted according to perceptual similarity). On the other hand, the cue-specific minimal pair measure identifies the neighbor that differs from the target solely in the cue of interest. Consequently, we can think of the cue-specific minimal pair measure as being more phonetically specific than the neighborhood density measure.

Given the differences in the way these two competition metrics are defined, there is a great deal of intermediate ground between them. One way to think about this intermediate ground is in terms of a continuum of specificity in lexical competition (but see section 2.5.1 for discussion of alternatives). For simplicity, we conceptualize this continuum relative to a given phonetic cue in terms of two parameters that define lexical neighborhoods of varying specificity: the relative *type* of competition between target and competitor, and the relative *position* of that competition (Figure 2.1). The most specific type and position on this continuum is defined by a cue-specific minimal pair competitor, which differs from the target word only in the measured cue, in the same segmental position that the cue is realized. For example, voice onset time is the primary cue distinguishing word-initial stop voicing in English (Lisker and Abramson, 1964; Lisker, 1986). Given the reference word *bill*, *pill* is a neighbor differing in the same cue (voice onset time) in the same segmental position (the word-initial segment). We can now instead define a more inclusive neighborhood for neighbors that share, for example, a manner of articulation in the same segment as the measured cue (e.g., *pill, kill, dill*...). Similarly, we can define the neighborhood to include any onset-competitor (e.g., *pill, will, mill*...). These two examples vary the phonetic *type* of competition, but hold the *position* of competition steady. We can also vary the relative position of competition. As an example, we can consider only those neighbors that differ in the second position of the word; given the target word *bill*, neighbors differing only in the second segment include such words as *bell*, *bowl*, and *ball*.

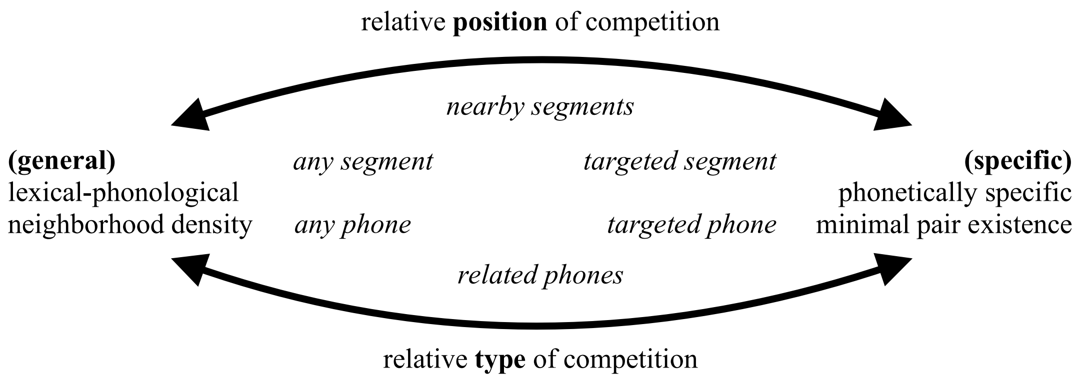

Figure 2.1: Schematization of a continuum of specificity in competition. Single-edit competition metrics can lie at various positions along this continuum according to the relative type or position of competition with the target, in terms of neither (neighborhood density), or in terms of both (phonetically specific minimal pair competitor existence).

The present study compares a number of lexical competition metrics based on this continuum of specificity. Based on a systematic sample of alternatives from this continuum, we used mixed effects regression to predict the realization of voice onset time in a corpus of conversational English speech. We tested this sample of alternative metrics separately for both voiced and voiceless word-initial stops. While we compared a substantial set of competing hypotheses, we note that this range of metrics does not exhaust the set of plausible ways of operationalizing lexical competition. Indeed, a number of alternative metrics can be found in the literature, including neighborhood measures weighted for (i) the proportion of segmental overlap (Goldrick et al., 2010), (ii) variable Levenshtein edit distance between target and competitors (Yarkoni et al., 2008), or (iii) perceptual similarity (Strand and Sommers, 2011; Gahl and Strand, 2016). Our results are considered with regard to these alternatives in the discussion.

This paper takes the following course. First, we briefly review cognitive mechanisms that have been proposed to underlie competition-associated hyperarticulation. Next we review prior research on competition-associated hyperarticulation, and the implications for each of these cognitive mechanisms. We then present results of analyses comparing a range of competition metrics for their ability to predict voice onset time in conversational speech. Finally, we discuss the implications of these results for the proposed cognitive mechanisms behind competition-associated hyperarticulation and for diachronic sound-change.

2.1.1 Accounts of competition-associated hyperarticulation

Why does competition among lexical items lead to hyperarticulation? One proposal appeals to processes of lexical selection and/or speech planning during production. In these production-internal approaches, lexical items compete for activation (e.g., Dell, 1986; see Fricke, 2013, ch. 2, for a review). This competition leads to overall higher activation levels for the target word, which in turn leads to enhancement of articulatory gestures in production (see Baese-Berk and Goldrick, 2009, and Fricke, 2013, chs. 2 & 6, for discussion). A common implication of this hypothesis is that competition should only result in *increases* in phonetic durations such as voice onset time. Further, these accounts generally assume that competition-associated hyperarticulation should be relatively insensitive to the precise phonetic relationship between target and competitor (e.g., Baese-Berk and Goldrick, 2009; Fricke, 2013; Watson et al., 2015; see Jaeger et al., 2016; Buz and Jaeger, 2016, for reviews).

Two alternative proposals predict that competition can result in targeted hyperarticulation of cues that distinguish a word from a specific competitor. According to communicative accounts (reviewed in Jaeger et al., 2016), speakers are implicitly aware that some words are more confusable than others, and are able to hyperarticulate the cues that maximally differentiate a given word from plausible alternatives. According to this approach, hyperarticulation is a tool that can be employed online by speakers in order to promote communicative efficiency. Over time, this online hyperarticulation can result in shifts in long-term lexical representations (reviewed in Hall et al., 2016). In a listener-internal approach, however, *hyper*-articulated tokens of potentially ambiguous words are more likely to be recognized, and therefore stored in memory, than *hypo*-articulated tokens. As a result, more ambiguous tokens contribute less to the exemplar cloud representing a given word or category, shifting it toward the hyperarticulated variant (see, e.g., Wedel, 2006). Because this same exemplar cloud serves as the source for future productions, this process can, over time, lead to notable differences in the phonetic forms of words. Such an account does not predict that hyperarticulation happens online, but that, through filtering by listeners, words should come to be articulated in a way that enhances the specific phonetic contrasts that distinguish them. It is worth noting that all of these proposed mechanisms operate at different levels, and are therefore mutually compatible. We consider it a priori possible that all of these mechanisms contribute to the range of hyperarticulation effects found in human language.

2.1.2 The phonetic specificity of competition

The three cognitive mechanisms outlined above are rooted in very different assumptions about the nature of lexical competition. One critical question is how sensitive competition is to the phonetic or phonological relationship between target and competitor. According to most production-internal accounts, competition is relatively abstract, occurring among phonological representations prior to phonetic encoding (e.g., Baese-Berk and Goldrick, 2009; see also Goldrick and Rapp, 2007). For example, Fricke's (2013) "Articulate As Soon As Possible Principle" (AASAPP) posits that articulation unfolds on a segment-by-segment basis, and begins as soon as competition in a given segmental position is resolved. According to this model, the relative position of difference between the target and its competitors is therefore of paramount importance, but the phonetic relationships among those competitors is not relevant—a greater number of competitors defined for a particular segmental position leads to increased activation of the target segment regardless of phonetic relationship (Fricke, 2013; see also Vitevitch and Chu, 2004; Fricke et al., 2016). In contrast, both communicative and listener-internal accounts are based on perceptual confusability, such that hyperarticulation is targeted to those cues that maximize perceptual distinctiveness between competitors.

This question of how sensitive competition is to phonetic relationships is still open. In the case of word-initial stops, the focus of this chapter, a number of researchers have argued that voice onset time is hyperarticulated in response to phonetically specific minimal pair competitors. This correlation has been reported for voiceless stops (Baese-Berk and Goldrick, 2009; Peramunage et al., 2011; Kirov and Wilson, 2012; Schertz, 2013; Buz et al., 2016), but also for voiced stops, for which voice onset times have been found to *decrease* (Schertz, 2013; Wedel et al., 2018; but see Ohala, 1994 and Goldrick et al., 2013 for different findings). Similar results have also been found for contrastive cues to coda voicing (vowel length and perseveration of voicing in English fricatives: Seyfarth et al., 2016; duration of fricatives and glottal pulsing in both stops and fricatives in Russian: Kharlamov, 2014). These results suggest that competition-driven hyperarticulation is *contrastive*, increasing the phonetic distance between target and competitor, and may be quite phonetically specific (e.g., Schertz, 2013; Seyfarth et al., 2016). Furthermore, Baese-Berk and Goldrick (2009) found a smaller, but significant effect of the mere lexical existence of the phonetically specific competitor, despite not being present in the experimental context (see also Peramunage et al., 2011). This additional result suggests that these effects may be partially mediated by processes that are not dependent on the presence of the competitor in the immediate context.

However, other researchers have reported evidence that this kind of minimal pair competition does not correlate with contrastive hyperarticulation. Goldrick et al. (2013) used a word list reading task to elicit productions of voiced stop-initial words as well as both voiced and voiceless stop-final words. For some of these words, the stop voicing minimal pair was a word of English (e.g., *bun* ~ *pun*, *coat* ~ *code*) and for others it was not (e.g., *bum* ~ **pum*, *thud* ~ **thut*). They found no significant effect of initial stop voicing minimal pair existence on the realization of voice onset time in voiced stops, nor did they find a significant effect of final stop voicing minimal pair existence on the realization of any cues to final voicing for voiceless stops (vowel duration, stop release, closure duration, or perseveration of voicing). However, they did find that final stop voicing minimal pair existence correlated significantly with vowel duration for final voiced stops, but the effect was opposite what would be expected under contrastive hyperarticulation. Rather than getting longer, which is a common cue to coda stop voicing cross-linguistically and in English, vowels before voiced stops in words with coda stop voicing minimal pairs were *shorter* than those in words without such minimal pairs, indicating a reduction of the phonetic contrast with the minimal pair competitor for the vowel duration cue.

Finally, a number of researchers have argued that hyperarticulation results from more general competition. Fox et al. (2015) reported that phonetically specific minimal pair existence did not significantly predict voice onset time realizations in word list and sentence reading tasks, while neighborhood density did. Fricke (2013) and Fricke et al. (2016) reported that hyperarticulation of word-initial voice onset time is most robustly predicted by the number of competitors in the onset position. Fricke and colleagues argued that both cue-specific minimal pair competition and more general neighborhood density correlate with hyperarticulation of word-initial voice onset time because they *also* correlate with this position-based measure (see also Vitevitch and Chu, 2004; Caselli et al., 2015). Similarly, Kirov and Wilson (2012) reported that hyperarticulation of voice onset time in word-initial stops correlated with the presence of a minimal pair competitor for either stop voicing (e.g., *cap* ~ *gap*) or place of articulation (e.g., *cap* ~ *tap*), both in the initial segment. They further found that positions other than the initial segment, i.e., the vowel or coda of CVC monosyllables, did not affect initial stop voice onset time realizations (e.g., *cat* ~ *kit*, *cat* ~ *cap*). In a related study, Schertz (2013) reported contrastive voice onset time hyperarticulation in both voiced and voiceless word-initial stops associated with the presence of a voicing competitor. However, she found no effect of either place or manner of articulation competitors.

In summary, lexical competition, broadly construed, has been found to be consistently associated with contrastive hyperarticulation of voice onset time, but the nature of the associated competition has varied substantially across studies. One possible explanation for these differences is methodological. The majority of studies exploring contrastive hyperarticulation have used laboratory elicitation of defined materials in order to control for the many factors correlated with phonetic realization. Depending on experimental conditions, elicited speech may be prone to particularly clear or slow articulations that have been associated with increased phonetic durations and exaggerated articulatory gestures (Picheny et al., 1986; De Jong et al., 1993; Smiljanic and Bradlow, 2008). This "clear speech" may represent a conceptually distinct form of hyperarticulation from competition-driven contrastive hyperarticulation of the type reviewed above (for similar terminology, see Ohala, 1994; Seyfarth et al., 2016; for other discussions of different kinds of hyperarticulation, see Cho et al., 2011; Schertz, 2013). We refer to this kind of generalized hyperarticulation as *clear-speech hyperarticulation*, which has been associated with increased voice onset times, particularly in voiceless stops (Smiljanic and Bradlow, 2008). As most studies examining the effects of competition on the realization of voice onset time have focused on voiceless stops (with the exception of Ohala, 1994; Schertz, 2013; Goldrick et al., 2013), the predictions of both clear-speech and contrastive hyperarticulation have coincided. In these studies, clear-speech hyperarticulation may raise voice onset times toward a ceiling that makes contrastive hyperarticulation effects more difficult to detect (e.g., Kirov and Wilson, 2012; see also Wedel et al., 2018 for discussion). Indeed, most of the studies that have found evidence of contrastive hyperarticulation have used task environments designed to amplify these potential effects, for example by including the minimal pair competitor in the immediate context (Baese-Berk and Goldrick, 2009, study 2; Kirov and Wilson, 2012; Seyfarth et al., 2016; Buz et al., 2016; but see Baese-Berk and Goldrick, 2009, study 1; Peramunage et al., 2011), or by explicitly indicating that the speaker's production was misperceived as the minimal pair competitor (Schertz, 2013; Buz et al., 2016).

One approach to minimize the influence of clear-speech hyperarticulation is to study speech in contexts promoting greater reduction, such as conversation (see, e.g., Gahl et al., 2012). A complementary strategy is to study phonetic cues for which the predictions of clear-speech and contrastive hyperarticulation diverge, such as in word-initial voiced stops (see Ohala, 1994; Schertz, 2013; Goldrick et al., 2013). Contrastive hyperarticulation of voiced stops should lead to *shorter* voice onset times relative to non-hyperarticulated words (Schertz, 2013), while clear-speech conditions are not associated with shorter voice onset times (e.g., Miller et al., 1986; Kessinger and Blumstein, 1997). Here, we study both voiced and voiceless stops in conversational speech because contrastive hyperarticulation should move voice onset times in opposite directions for these two stop types (e.g., Schertz, 2013).

2.1.3 The present study

The present study was designed to clarify the specificity of the relationship among targets and competitors that leads to hyperarticulation. To this end, we compared a sample of competition measures from the continuum of specificity (Figure 2.1) for their ability to predict the realizations of word-initial stop voice onset times in conversational English. In particular, we explored (i) the role of segmental position and (ii) the role of the phonetic relationship between segments within a given position. Though these two dimensions do not represent all possible hypotheses about how lexical competition can be operationalized, both of these dimensions have important theoretical consequences for accounts of competition effects. For example, Fricke's (2013) AASAPP, among other production-internal accounts, predicts that hyperarticulation is targeted only to specific segmental positions, and can only lead to increased durations. Both trade-off and perception-based approaches, on the other hand, predict that hyperarticulation is targeted to specific cues and is contrastive, i.e., can be realized as either increased or decreased durations. These different expectations regarding the possible triggers and outcomes of lexically conditioned hyperarticulation have not been clearly reconciled in the literature (compare, e.g., Fox et al., 2015; Seyfarth et al., 2016; Fricke et al., 2016).

In this study we look for contrastive hyperarticulation in the form of both longer voice onset times in voiceless stops, and shorter voice onset times in voiced stops. We investigate a sample of competing hypotheses regarding the relationship between target and competitor that gives rise to hyperarticulation, looking for these effects in conversational speech, where reduction is more likely to occur. Our study must therefore be considered alongside additional research testing alternative hypotheses, as well as investigating these phenomena in elicited speech paradigms.

2.2 Materials and Methods

We examined the effects of competition on the realization of voice onset time in conversational English based on a sampling of different competition metrics. These metrics took the form of modified neighborhood densities designed to sample two dimensions of relatedness in competition characterizing the differences between overall neighborhood density and cue-specific minimal pair competition. These modified neighborhood density measures targeted either the relative *position* of competition between target and competitor (i.e., *where*, in terms of word-internal segmental position, the two words differ) or the relative *type* of competition between target and competitor (i.e., *how* the two words differ within a given segmental position). We then used these metrics to predict voice onset time realizations of voiced and voiceless stops in the Buckeye Corpus of Conversational Speech (Pitt et al., 2005, 2007). The use of conversational speech allows us to study the effects of competition in a generally more reduced context, limiting the potential for clear-speech hyperarticulation to mask the effects of contrastive hyperarticulation. The hypothesis that competition drives contrastive hyperarticulation predicts opposite effects for voiced versus voiceless stops: for voiced stops, voice onset times should decrease, while for voiceless stops, voice onset times should increase (Schertz, 2013). Including both stop types provides two distinct tests of this hypothesis.

We used linear mixed effects models to analyze the predictive relationship between competition and the realization of voice onset time. Due to the large number of hypotheses regarding the operationalization of competition, we evaluated these models using corrected Akaike's Information Criterion (AIC_c) comparisons and evidence ratios (Burnham and Anderson, 2004; Lukacs et al., 2007; Richards et al., 2011). This approach does not test for significance as in null hypothesis testing approaches to model comparison such as log-likelihood ratio tests, making it better suited to multiple hypothesis testing, in which multiple comparisons increase the risk of family-wise error (see Chamberlin, 1890, and Shadish, 1993, for discussion of the merits of testing multiple working hypotheses). In addition, evidence ratios allow for quantified statements about the relative support in favor of one model over another, ideal for comparing competing hypotheses (Burnham and Anderson, 2004; Richards et al., 2011). After comparing the models using AIC_c, we tested the top-performing model for both voiced and voiceless stops to ensure that the competition measure was contributing significantly to model fit using log-likelihood ratio tests of nested models. If the competition measure in the top-performing model does not contribute significantly to model fit, the relative ranking of models is not clearly interpretable.

2.2.1 The data source

We used natural speech data from the Buckeye Corpus of Conversational Speech (Pitt et al., 2005, 2007). The Buckeye Corpus includes 40 hours of conversation, spread over one-hour long interviews with 40 individuals. Half of the interviewees are male, and half are female; half are over the age of 40 and half are under the age of 30. Interviews were conducted in Columbus, Ohio, and all speakers are from Columbus or surrounding regions of Ohio. The Buckeye Corpus is annotated for utterances as well as words and their syntactic category, and includes phonetic transcriptions with segment-level durations, but no sub-segmental measurements such as voice onset time.

2.2.2 Voice onset time measurements

We used measurements of the voice onset times of voiced and voiceless stops from a set of 24 out of the 40 speakers in the corpus (14 younger, 10 older; 13 female, 11 male). Measurements were made by hand for stop-initial content words (labeled in the corpus as noun, verb, adjective, or adverb) of one or two syllables produced by these speakers. Our study thus differs from much previous work on competition effects in that our data includes two-syllable words. In addition, unlike in many other studies where only CV-initial words were included (e.g., Baese-Berk and Goldrick, 2009; Kirov and Wilson, 2012; Schertz, 2013; Fricke, 2013; Fox et al., 2015), we expanded our word types to include complex onsets. Because of the phonotactics of English, this meant the inclusion of words with an initial stop followed by a liquid or a glide.

Each word was annotated by hand in Praat (Boersma and Weenink, 2010) for the beginning of the stop closure, the beginning of the burst, and the beginning of the following sonorant. Based on those measurements, the total stop duration, the closure duration, and the voice onset time were calculated. We excluded tokens with pre-voicing, tokens with no identifiable burst, and tokens with closure durations or voice onset times that were more than 3 standard deviations from the specific speaker's mean for that stop consonant. For the remaining tokens, we calculated the proportion of the total stop duration that was taken up by the voice onset time (VOT) according to the following formula:

$$\frac{burst + aspiration}{closure + burst + aspiration} \tag{2.1}$$

Where closure, burst, and aspiration are all measures of duration in milliseconds for stop closure, stop release burst, and post-release aspiration, respectively. This results in values between 0 and 1 representing the relative proportion of total stop duration that is taken up by the VOT (i.e., burst + aspiration).

Henceforth, we refer to this measurement as the "VOT/Stop-length ratio". We elected to use the VOT/Stop-length ratio rather than raw voice onset times because it provides a very local control for speech rate. Speech rate is highly correlated with voice onset time in voiceless stops of English (Kessinger and Blumstein, 1998; Yao, 2007), but this correlation is largely attributable to an effect on the entire stop duration. However, an increase or decrease in the VOT/Stop-length ratio necessarily reflects a change in the voice onset time independently of any change to the total stop duration (see Smiljanic and Bradlow, 2008, for use of a related proportional measure of voice onset time).

We excluded high frequency discourse markers and any content words homophonous with function words (Bell et al., 2009; Gahl et al., 2012; Seyfarth, 2014). Because we based our neighborhood measures on the lemma forms of words (see section 2.2.3 below), we further excluded verbs with stem-vowel changes in their morphology (e.g., *buy* ~ *bought*, *come* ~ *came*), as the specific minimal pair competitor lemma is not consistent across the paradigm (e.g., *buy* ~ *pie* versus *bought* ~ *pot*). Of the remaining words included in our dataset, tokens were excluded if there was no identifiable burst, if the token was immediately preceded by another stop consonant (due to unreliability of assigning the beginning of the stop closure), if the token followed an annotated pause or disfluency, or if either the aspiration or closure length of the token was more than three standard deviations from the speaker's mean.

2.2.3 Sources for competition metrics

All of the competition metrics used in both studies were calculated based on lemma forms. We used the Corpus Of Contemporary American English (COCA: Davies, 2009) to filter the Carnegie-Melon University pronouncing dictionary (CMU: Carnegie-Mellon University, 2015) such that only distinct lemma forms remained. The CMU pronouncing dictionary, however, includes multiple pronunciations of many words, and so in order to prevent the same lemma from counting in any neighborhood measure more than once, we included only the first pronunciation of each distinct lemma. Only unique phonemic forms were retained, with the result that homophonic lemmas (e.g., *bear*, *bare*) correspond to the same entry. Because the CMU pronouncing dictionary is organized by orthographic form, a set of English homographs corresponding to distinct lemmas were manually retained (e.g., *tear* can be either the verb /tɛɹ/ or the noun /tɪɹ/). The full list of these homographs can be found in Table 2.1. We retained the /ɑ ~ ɔ/ contrast (Durian, 2012), but collapsed /w ~ ʍ/ (Labov et al., 2006), consistent with pronunciation norms in Central Ohio.

Finally, we filtered this lemmatized CMU pronouncing dictionary based on contextual diversity, excluding any forms appearing in less than 0.5% of films in the SUBTLEX-US database (Brysbaert and New, 2009). This was done to reduce the contribution of jargon or uncommon words to the resulting lexicon. Contextual diversity has been shown to be

Homographs retained in final source lexicon

abuse	advocate	allied	alternate	appropriate	articulate	bass
bow	buffet	close	combine	compact	complex	compound
concert	conduct	conflict	console	content	contract	convict
decrease	desert	dove	intimate	invalid	laminate	lead
learned	live	minute	moped	object	polish	present
produce	progress	read	rebel	record	refuse	resign
resume	separate	subject	tear	use	wind	wound

Table 2.1: List of English homographs manually retained in our source lexicon. Note that this was done without discriminating based on the number of syllables because two-syllable words can have poly-syllabic neighbors. Additional homographs not present on this list either involve inflectional variants (e.g., *confines* (n.) ~ *confines* (v., inflected)), have pronunciations varying only in stress, which in our phonemic representations do not differ (e.g., *exploit* (n.) ~ *exploit* (v.)), or involve a form not represented in CMU at all (e.g., *blessed* (v.) ~ *blessed* (adj., not represented)).

a better predictor of lexical decision accuracy and reaction time than frequency in both the visual (Brysbaert and New, 2009) and auditory (Geary, 2019) domains. The 0.5% cutoff results in a lexicon of 11,692 lemmas; for comparison, a frequency cut-off of one per million produces a lexicon of similar size (12,811 lemmas). This lexicon served as the source file when calculating all of our competition metrics.

2.2.4 Competition metrics

We created a systematic sample of competition metrics from the continuum of specificity as conceptualized in Figure 2.1. This continuum is defined by the two metrics that have dominated the literature on competition-induced hyperarticulation: lexical-phonological neighborhood density and cue-specific minimal pair competitor existence. For comparability with prior work, we included both of these competition measures in our study. Lexical-phonological neighborhood density was defined as the tally of all words that can be derived from the target by adding, deleting, or replacing any single phoneme of the target word (Luce and Pisoni, 1998). Henceforth, we will refer to this measure as the 'overall neighborhood density'. Phonetically specific minimal pair competitor existence was coded as a binary neighborhood density. If switching the voicing value of the initial stop consonant of the target lemma resulted in a unique lemma in our lexicon (e.g., *bat* ~ *pat*), then this measure was coded as 1; otherwise, it was coded as 0. We refer to this metric as 'minimal pair competitor existence'.

To explore the space between these measures on the continuum, we used a sample of intermediate neighborhood density measures targeting either the relative position or type of competition. For the relative position of competition, we calculated three modified neighborhood densities, each targeting a different segmental position, or set of segmental positions, within the word (Figure 2.2). These three neighborhood densities targeted the first segmental position, the second segmental position, and the rest of the word (from the third segment to the last segment). We chose to include a metric for competition in the onset position because competition in this position has been found to be predictive in a number of studies (Vitevitch and Chu, 2004; Goldrick et al., 2010; Fricke, 2013; Caselli et al., 2015; Fricke et al., 2016). We included a metric for competition in the second segment because, as stop bursts often contain cues to the formants of following sonorants (Suchato and Punyabukkana, 2005), hyperarticulation of bursts might arise when there are many competitors in the following segment. Finally, we included a metric for competition elsewhere in the word to evaluate whether phonetic variation is insensitive to segmental position. Note that all of these neighborhoods include only neighbors formed by substituting a single segment in the relevant position or set of positions; none of the modified neighborhood densities used in this study included additions or deletions of segments.

For the relative type of competition, we calculated modified neighborhood densities targeting three types of competition, each within the first segmental position: the place of articulation, manner of articulation, and voicing value of the competitor relative to the target. For each of these, we calculated two neighborhood densities, one for competitors that share the feature with the target and one for competitors that have a different value for that feature relative to the target. This resulted in a total of six modified neighborhood densities targeting the relative type of competition (Figure 2.3). This set of position- and type-based competition measures was not intended to provide an exhaustive survey of all

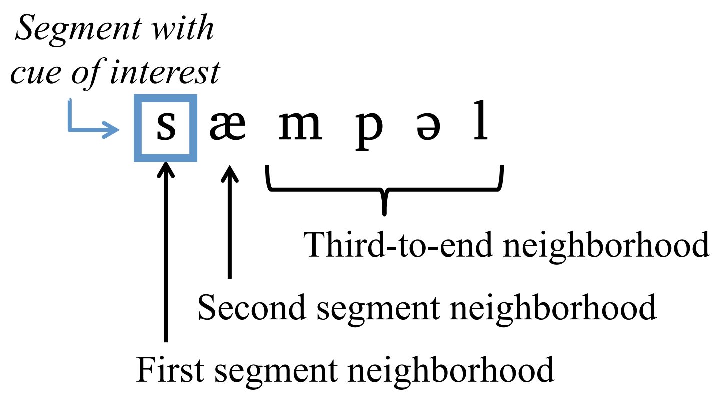

Figure 2.2: Neighborhoods for the relative position of competition in the word "sample". Each neighborhood density references a segmental position or set of segmental positions. The tally of the density of each of these neighborhoods is increased by one if replacing the segment in one of that neighborhood's segmental positions results in a unique lemma.

possible competition relationships, but rather was chosen to provide a reasonable sample of possible phonological relationships between target and competitor (for similar approaches, see Kirov and Wilson, 2012; Schertz, 2013).

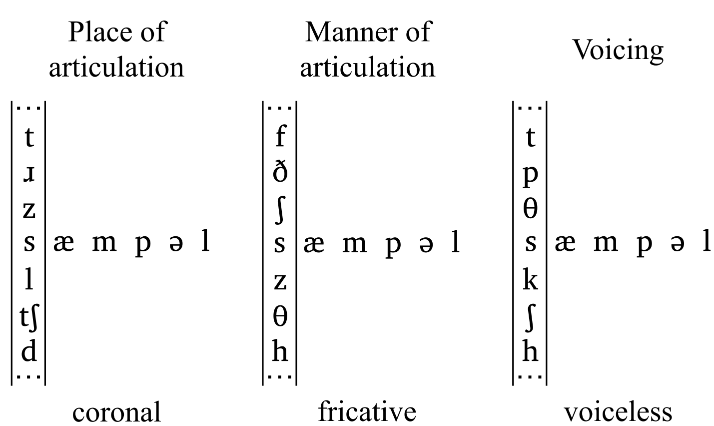

Figure 2.3: The three types of same-feature competition included in our study, for the word sample. Each neighborhood represented here targets competitors with the same featural value as the target. Three other neighborhoods targeting competitors with different values for each of these three features were also calculated. The density of each of these neighborhoods is increased by one if replacing the first segment with another segment fitting that neighborhood's targeted featural type results in a unique lemma.

In order to calculate these competition metrics, we sorted the various phonemes of English into a simplified set of possible feature values. As our dependent measure was voice onset time, we sorted all consonants into one of three places of articulation: labial, coronal, or dorsal[1]. If a consonant could not be fully described by one of these feature values, that consonant was coded as the closest approximate place of articulation of these three (e.g., /f/ was coded as labial, as it is at least partly labial in articulation; /h/ was coded as dorsal, as dorsal is the closest place of articulation in our list to glottal). All consonants were additionally coded as either plosive or non-plosive, where "plosive" was defined as all oral stops and affricates (/p|b|t|d|k|g|t͡ʃ|d͡ʒ/), and "non-plosive" contained every other consonant, including nasal stops, approximates, and fricatives. Finally, all consonants were coded as either voiced or voiceless according to their standard phonemic transcription.

For a given target word, all neighborhood density measures were calculated by extracting the lemma form of the word and identifying the phonemic form of that lemma in our lexicon (section 2.2.3). We then identified each distinct lemma in that lexicon that fit the criteria of each of our competition metrics using regular expressions, and added it to the appropriate metric's neighborhood. Thus, each neighborhood density measure received a count of +1 for each unique lemma entry in our lexicon that could be formed by replacing the appropriate phoneme of the target word with another single phoneme, provided that the substitution fit the criteria of the neighborhood in question. In an effort to further test the range of possibilities in lexical competition, we also created versions of each of these neighborhood density measures that were weighted for frequency. We will turn to the description of these frequency-weighted measures in section 2.4.

2.2.5 Mixed effects models

The competition metrics were included in linear mixed effects models conducted in R (R Core Team, 2018) using the lme4 package (Bates et al., 2008). Models were evaluated separately for voiced and voiceless stops due to different expectations for the effects of

[1]The single exception to this is /w/, which was coded as both labial and dorsal.

Competition metric	Targeted segments/features	Examples for /pæt/
Overall ND	Any segment Any feature	spat (/spæt/) at (/æt/) pit (/pɪt/)
First segment ND	First segment only Any feature	cat (/kæt/) fat (/fæt/)
Second segment ND	Second segment only Any feature	pit (/pɪt/) pot (/pɔt/)
Third-to-end ND	Any segment after second segment Any feature	pan (/pæn/) patch (/pæt͡ʃ/) pack (/pæk/)
Same place ND	First segment only Same place of articulation	fat (/fæt/) mat (/mæt/)
Different place ND	First segment only Different place of articulation	cat (/kæt/) sat (/sæt/)
Same manner ND	First segment only Same manner of articulation	cat (/kæt/) bat (/bæt/)
Different manner ND	First segment only Different manner of articulation	fat (/fæt/) that (/θæt/)
Same voicing ND	First segment only Same voicing value	cat (/kæt/) sat (/sæt/)
Different voicing ND	First segment only Different voicing value	that (/θæt/) rat (/ɹæt/)
Minimal pair existence	First segment only Same place of articulation Same manner of articulation Different voicing value	bat (/bæt/)

Table 2.2: The competition metrics with descriptions and examples. Examples are non-exhaustive, provided for the target word *pat*.

competition on the realization of voice onset times for each stop type (e.g., Schertz, 2013). For each of our stop types, each competition metric was included in a separate linear mixed effects regression model along with a consistent set of control predictors that have been shown to influence the realization of either voice onset time specifically, or word or segment durations more generally. For each subset of the data (voiced and voiceless stops), an additional model with only the control predictors and no competition metric was included as a baseline to which the other models could be compared (the “base model”). Thus, for both voiced and voiceless stops, the models in each set differed only in terms of the particular competition metric included (when one was included at all).

2.2.6 Factors of interest

VOT/Stop-length ratio

The dependent measure used in this study was the proportion of total stop duration taken up by the voice onset time (VOT/Stop-length ratio; equation 2.1). As previously noted in section 2.2.2, this was calculated as the total voice onset time divided by the total stop duration. As a proportion, the VOT/Stop-length ratios all lay on a scale between 0 and 1.

Competition metrics

Each model either included one of the competition metrics described above, or did not include any competition metric (base models). If the model included a competition metric, that value was centered, and then linearly transformed to a scale between -1 and 1 (note that this does not apply to the logical minimal pair existence factor).

2.2.7 Control predictors (fixed effects)

We included the following control predictors in all models. All continuous variables were centered, and then linearly transformed to a scale between -1 and 1 to facilitate model convergence.

Stop phoneme

The identity of the initial stop phoneme (/p|t|k/, /b|d|g/).

Following liquid

Whether or not the segment following the initial stop is one of /l, ɹ, ɝ/ (`true` or `false`). Visual inspection of the data indicated that words followed by a liquid had considerably longer voice onset times.

Number of syllables

The number of syllables in the phonemic transcription of the word (`1` or `2`). This factor was included under the assumption that bisyllabic words would have systematically shorter voice onset times on average.

Syntactic category

The syntactic category of the word-token (`noun`, `verb`, `adjective`, or `adverb`) as reported in the Buckeye Corpus.

Previous mention

Whether or not the lemma (not the word) had already appeared in the speaker's prior discourse at the time this token was produced (`true` or `false`). Approximately 39% of the tokens in this dataset occurred after a prior use of the same lemma in a speaker's own speech (2345/5957 tokens). Repeated use in discourse is correlated with reduction (Bell et al., 2009).

Speech rate

The number of syllables per second in the phonemic transcription of the stretch of continuous speech surrounding the token. This was calculated as the number of vowels in the phonemic transcript for the region of continuous speech within which the token is found, divided by the number of seconds in that region of continuous speech (Bell et al., 2003). The boundaries for regions of continuous speech were based on the Buckeye transcripts, and included such things as silences, pauses or disfluencies, interviewer speech intervals, and the beginnings and ends of audio files. Speech rate is highly correlated with durational cues, including voiceless stop voice onset time (e.g., Yao, 2007).

Word familiarity

The contextual diversity of the orthographic word form of each token, represented as the percent of films in which the word appears in the SUBTLEX-US database. Word familiarity measures correlate with reduction (e.g., word frequency: Bell et al., 2009), and contextual diversity has been found to be a better predictor of behavioral data than other familiarity measures such as frequency in both the visual (Brysbaert and New, 2009) and auditory domains (Geary, 2019).

Forward and backward contextual probability

Forward and backward conditional bigram probabilities, log-transformed. Conditional bigram probabilities were calculated per two-word pair using a combined corpus consisting of both the Buckeye Corpus (Pitt et al., 2007) and the Fisher Part 2 transcript corpus (Cieri et al., 2005), a set of 5849 English telephone conversations containing over 12 million words. A combined corpus was used in order to (i) ensure more realistic bigram probability estimates based on a larger sample of American English conversational speech than the Buckeye Corpus alone provides (note that the Buckeye Corpus contains only about 300,000 words. For use of the Fisher corpus in this way, see, e.g., Arnon and Snider, 2010; Seyfarth, 2014; Wedel et al., 2018), and (ii) ensure that every bigram in the Buckeye Corpus appears at least once in the combined corpus, thus avoiding conditional bigram probabilities of zero (note, however, that this approach likely over-estimates conditional probabilities for infrequent bigrams from the Buckeye Corpus). Such contextual probability measures are predictive of durations in the Buckeye corpus (Seyfarth, 2014).

Phonotactic probability

The average biphoneme sequence probability of the phonological form of the lemma measured without reference to stress, available through the IPhOD2 database, log-transformed.

This measure calculates the average probability of each two-phoneme sequence in the lemma, specific to its position as measured from the start of the lemma (Vaden et al., 2009). This measure is correlated with general phonetic reduction in speech production (Vitevitch and Chu, 2004).

2.2.8 Random effects

All models included random intercepts for both the speaker and the lemma. In addition, for each model with a competition metric, we included a correlated random slope for that metric on the speaker intercept (Barr et al., 2013). We did not include random slopes for our control predictors as this could lead to problems with model convergence, and we had no principled reason to include any particular slopes over others.

2.2.9 Model evaluation procedure

Models were first evaluated according to their corrected Akaike's Information Criterion (AIC_c), an information-theoretic approach to model comparison based on entropy (see Burnham and Anderson, 2004), using the AICcmodavg package (Mazerolle, 2016). The AIC_c value can be thought of in terms of information loss, where a lower AIC_c value corresponds to less information loss and therefore a more accurate model. However, AIC_c values by themselves are essentially meaningless, and must be interpreted relative to alternative AIC_c values. To do this, models are included in a candidate set and ranked according to their AIC_c value. These models are given weights based on their normalized log-likelihood, and compared based on these weights (Burnham and Anderson, 2004). This relative evaluation is operationalized in terms of the change in AIC_c value between the current model and the top-performing model (ΔAIC_c). In general, models with a $\Delta AIC_c \leq 2$ are considered to have substantial support relative to the top-performing model (i.e., they are not deemed considerably inferior to that model), and models with a $\Delta AIC_c \geq 10$ are taken to have very little support (Burnham and Anderson, 2004). In addition, AIC_c comparison allows evidence ratios to be calculated. Evidence ratios compare two models directly, and are calculated as the ratio of the AIC_c weights of the two models (Richards et al., 2011). These evidence ratios allow us to describe the models in terms of the amount of evidence in favor of the better model with respect to the other.

In addition to the AIC_c model comparison, we further tested the top-performing model of each set (voiced and voiceless stop models) for the significance of the competition metric. To test for statistical significance, we subjected these models to nested model comparison using the log-likelihood ratio test. In this case, the models in question were compared to the corresponding model with the fixed effect of the competition metric removed. Thus, for both voiced and voiceless stops separately, the model with the lowest AIC_c value was compared to a restricted version of itself without the fixed effect of the relevant competition metric.

2.3 Results

Results are presented for voiced (section 2.3.1) and voiceless (section 2.3.2) stops separately. Summaries of the top-performing models can be found in Appendices B (voiced) and C (voiceless).

2.3.1 Voiced stops

The voiced stop data included 2,267 observations meeting our criteria from 24 speakers, distributed over 293 lemmas. Variance inflation factors (VIFs) for all of the predictors of interest were less than 1.9, indicating low multi-collinearity between the factors of interest and control predictors. The overall neighborhood density metric had the highest VIF of all factors of interest (1.80), while minimal pair existence had the smallest VIF (1.21). VIFs less than 2 are generally not cause for concern (see O'brien, 2007 and Belsley et al., 2005, ch. 3, for discussion).

Corrected Akaike's Information Criterion comparison

The AIC_c comparison table for all voiced stop models is presented in Table 2.3. Models are ranked according to their AIC_c value, with a lower AIC_c value corresponding to a better model fit. The model including a factor for minimal pair existence had the lowest AIC_c

value ($\Delta AIC_c = 0$), and therefore the best overall fit to the data. The model including a factor for neighbors with the same place of articulation as the target was the second-best model (same place ND: $\Delta AIC_c = 4.49$), but evidence ratios indicate that there was about 9 times more evidence in favor of the minimal pair model. The only other model with a $\Delta AIC_c < 10$ was the neighborhood density for second segment competitors (second segment ND: $\Delta AIC_c = 8.13$), with about 58 times more evidence in favor of the minimal pair model. AIC_c comparisons suggest virtually no evidence in favor of any of the other competition metrics in the voiced stop dataset (AIC_c weight < 0.01; $\Delta AIC_c > 10$). These results suggest that the existence of the phonetically specific minimal pair competitor in the lexicon was the best predictor of voice onset time values for voiced stops in this dataset.

Rank	Model	K	AIC_c	ΔAIC_c	AIC_cWt	Cum.Wt	LL
1	Min. pair exist	20	-4131.63	0.00	0.88	0.88	2086.00
2	Same place ND	20	-4127.14	4.49	0.09	0.97	2083.76
3	Second seg. ND	20	-4123.49	8.13	0.02	0.99	2081.93
4	Overall ND	20	-4121.26	10.36	0.00	0.99	2080.82
5	First seg. ND	20	-4118.50	13.13	0.00	1.00	2079.43
6	Base model	17	-4118.22	13.41	0.00	1.00	2076.24
7	Diff. manner ND	20	-4117.89	13.74	0.00	1.00	2079.13
8	Third-to-end ND	20	-4117.28	14.35	0.00	1.00	2078.83
9	Same voice ND	20	-4116.93	14.70	0.00	1.00	2078.65
10	Diff. voice ND	20	-4116.49	15.14	0.00	1.00	2078.43
11	Same manner ND	20	-4115.87	15.76	0.00	1.00	2078.12
12	Diff. place ND	20	-4115.83	15.80	0.00	1.00	2078.10

Table 2.3: AIC_c comparison table for all models predicting word-initial voiced stop VOT/Stop-length ratios. Models are ranked in order of AIC_c value. K = the number of estimable parameters in the model. AIC_c = the corrected AIC value. ΔAIC_c = the change in corrected AIC value between each model and the top performing model. AIC_cWt = the relative likelihood that the present model is the best model, presented as a proportional weight. Cum.Wt = the cumulative AIC_c weight of the present model and all higher-ranked models. LL = the log-likelihood of the model.

The top model: minimal pair existence

The minimal pair existence model is summarized in Table 2.4. As predicted, minimal pair existence correlates with a decrease in the VOT/Stop-length ratio, suggesting that these words are contrastively hyperarticulated away from their voice onset time competitor. As the top-performing model, we evaluated whether the minimal pair existence factor significantly contributed to model fit using nested model comparison. Removal of the fixed effect of minimal pair existence significantly affected model fit ($\chi^2(1) = 16.319$, $p < 0.001$), indicating that minimal pair existence is a significant predictor of voiced stop VOT/Stop-length ratios.

Variable	β	SE	t
(Intercept)	0.241	0.013	18.654
Min Pair Exist (true-false)	**-0.036**	**0.009**	**-4.146**
Stop Phoneme (b-d)	-0.107	0.008	-13.511
Stop Phoneme (g-d)	0.034	0.010	3.417
Speech Rate	0.026	0.017	1.499
Following Liquid (true-false)	0.086	0.008	10.671
Word Familiarity	-0.010	0.015	-0.658
Phonotactic Probability	-0.007	0.039	-0.174
Forward Probability	0.039	0.010	3.864
Backward Probability	-0.012	0.010	-1.232
Number of Syllables (1-2)	-0.007	0.007	-0.986
Previous Mention (true-false)	0.006	0.004	1.335
Syntactic Category (Noun-Adjective)	-0.003	0.009	-0.354
Syntactic Category (Adverb-Adjective)	0.017	0.013	1.261
Syntactic Category (Verb-Adjective)	0.024	0.011	2.114

(a) Fixed effects summary

Group	Variable	Variance	SD	Corr
Lemma	(Intercept)	0.0009	0.030	
Speaker	(Intercept)	0.0012	0.035	
	Min Pair Exist (true-false)	**0.0001**	**0.011**	**-0.71**
Residual		0.0086	0.093	

(b) Random effects summary.

Table 2.4: Summaries of fixed and random effects for the model including minimal pair existence for voiced stops.

2.3.2 Voiceless stops

The voiceless stop data included 3,690 observations meeting our criteria from 24 speakers, distributed over 417 lemmas. VIFs for all of the predictors of interest were less than 1.8, indicating low multi-collinearity between the factors of interest and control predictors. The overall neighborhood density metric had the highest VIF of all factors of interest (1.70), while minimal pair existence had the smallest VIF (1.10).

Corrected Akaike's Information Criterion comparison

The AICc comparison table for all voiceless stop models is presented in Table 2.5. The model including a factor for minimal pair existence had the lowest AIC_c value ($\Delta AIC_c = 0$), and therefore the best overall fit to the data. The model including a factor for the second segment neighborhood density was the second-best model ($\Delta AIC_c = 2.56$), and evidence ratios indicate that there was only about four times more evidence in favor of the minimal pair model. Relative to all other models, the minimal pair model had limited support ($\Delta AIC_c > 2$). However, the minimal pair model had more substantial support relative to the base model ($\Delta AIC_c = 5.82$), with about 18 times more evidence in favor of the minimal pair existence model over the base model. These results suggest that the existence of the phonetically specific minimal pair competitor in the lexicon was the best predictor of voiceless stop voice onset times, but that the support for this model over many of the other models was limited.

The top model: minimal pair existence

The minimal pair existence model is summarized in Table 2.6. As predicted, minimal pair existence correlates with an increase in the VOT/Stop-length ratio, suggesting that these words are contrastively hyperarticulated away from their voice onset time competitor. As the top-performing model, we evaluated whether the minimal pair existence factor significantly contributed to model fit using nested model comparison. Removal of the fixed effect of minimal pair existence significantly affected model fit ($\chi^2(1) = 5.25, p < 0.05$), indicat-

Rank	Model	K	AIC_c	ΔAIC_c	AIC_cWt	Cum.Wt	LL
1	Min. pair exist	20	-6124.87	0.00	0.55	0.55	3082.55
2	Second seg. ND	20	-6122.31	2.56	0.15	0.70	3081.27
3	Same place ND	20	-6121.32	3.55	0.09	0.79	3080.77
4	Same manner ND	20	-6120.24	4.63	0.05	0.85	3080.24
5	Diff. voice ND	20	-6120.24	4.63	0.05	0.90	3080.24
6	Overall ND	20	-6119.39	5.48	0.04	0.94	3079.81
7	Base model	17	-6119.05	5.82	0.03	0.97	3076.61
8	First seg. ND	20	-6117.66	7.21	0.01	0.98	3078.95
9	Diff. manner ND	20	-6116.30	8.57	0.01	0.99	3078.27
10	Diff. place ND	20	-6115.79	9.08	0.01	0.99	3078.01
11	Third-to-end ND	20	-6114.44	10.43	0.00	1.00	3077.34
12	Same voice ND	20	-6113.72	11.15	0.00	1.00	3076.98

Table 2.5: AIC_c comparison table for all models predicting word-initial voiceless stop VOT/Stop-length ratios. Models are ranked in order of AIC_c value. K = the number of estimable parameters in the model. AIC_c = the corrected AIC value. ΔAIC_c = the change in corrected AIC value between each model and the top performing model. AIC_cWt = the relative likelihood that the present model is the best model, presented as a proportional weight. Cum.Wt = the cumulative AIC_c weight of the present model and all higher-ranked models. LL = the log-likelihood of the model.

ing that minimal pair existence is a significant predictor of voiceless stop VOT/Stop-length ratios.

2.3.3 Summary and discussion

These results support the claim that minimal pairs are hyperarticulated away from each other in a manner that enhances the phonetic contrast between them. We found evidence of this contrastive hyperarticulation in the form of both longer voice onset times in voiceless stops, and shorter voice onset times in voiced stops. For both voiced and voiceless stops, the best model in terms of AIC_c included a factor for cue-specific minimal pair existence that contributed significantly to model fit. This suggests that competition defined for the cue of interest is the best predictor of contrastive hyperarticulation in word-initial stops of conversational English. We also found limited evidence that competition among neighbors that share a place of articulation in their initial segment (same place ND), as well as com-

Variable	β	SE	t
(Intercept)	0.557	0.015	37.98
Min Pair Exist (true-false)	**0.023**	**0.010**	**2.34**
Stop Phoneme (p-t)	-0.137	0.008	-17.23
Stop Phoneme (k-t)	-0.018	0.008	-2.26
Speech Rate	-0.033	0.015	-2.21
Following Liquid (true-false)	0.039	0.007	5.88
Word Familiarity	-0.019	0.013	-1.46
Phonotactic Probability	-0.001	0.036	-0.03
Forward Probability	0.007	0.009	0.79
Backward Probability	-0.012	0.009	-1.41
Number of Syllables (1-2)	-0.0004	0.006	-0.08
Previous Mention (true-false)	-0.0004	0.004	-0.12
Syntactic Category (Noun-Adjective)	0.004	0.010	0.42
Syntactic Category (Adverb-Adjective)	0.015	0.023	0.65
Syntactic Category (Verb-Adjective)	-0.006	0.011	-0.50

(a) Fixed effects summary.

Group	Variable	Variance	SD	Corr
Lemma	(Intercept)	0.0010	0.032	
Speaker	(Intercept)	0.0015	0.039	
	Min Pair Exist (true-false)	**0.0004**	**0.020**	**0.08**
Residual		0.0103	0.101	

(b) Random effects summary.

Table 2.6: Summaries of fixed and random effects for the model including minimal pair existence for voiceless stops.

petition in the following segment (second segment ND), can influence voice onset time in word-initial stops (these were the second- and third-best models in terms of AIC_c for both voiced and voiceless stops); however, the evidence in support of these models was relatively weak (all $\Delta AIC_c > 2$).

In this dataset, we find evidence of contrastive hyperarticulation in both voiced and voiceless stops. To our knowledge, only three prior investigations of contrastive hyperarticulation have examined voiced stops (Ohala, 1994; Schertz, 2013; Goldrick et al., 2013). Schertz (2013) found that voiced stops were realized with shorter voice onset times in clarifications following misperception as their voiceless counterpart, paralleling our results here. Ohala (1994) found the same trend, but it was not significant. Goldrick et al. (2013), however, found that word-initial voiced stops were not articulated differently depending on whether they had an initial stop voicing minimal pair in a word list reading task. These differences in results may be due, in part, to the use of different tasks. Studies employing word or sentence reading tasks to look for contrastive hyperarticulation of minimal pairs have sometimes found small effects (Baese-Berk and Goldrick, 2009, study 1; Peramunage et al., 2011; but see Fricke et al., 2016), but have sometimes failed to find any effect (e.g., Goldrick et al., 2013; Fox et al., 2015). On the other hand, studies in which cue-specific minimal pairs compete directly in communicative tasks have reported contrastive hyperarticulation effects more consistently or more robustly (Baese-Berk and Goldrick, 2009, study 2; Kirov and Wilson, 2012; Schertz, 2013; Seyfarth et al., 2016; Buz et al., 2016; but see Ohala, 1994). This suggests that contrastive hyperarticulation effects of the type reported here may be more easily detected under conditions promoting communication (Buz et al., 2016). That we find evidence of contrastive hyperarticulation of voiced stops may partly reflect the spontaneous and communicative nature of the speech in this dataset, as opposed to speech elicited via less communicative tasks, such as word or sentence list reading tasks.

2.4 Alternative Analyses Using Frequency-Weighted Neighborhoods

We repeated our analyses using frequency-weighted versions of the competition metrics described above. Following Luce and Pisoni (1998), these frequency-weighted measures were constructed as the ratio of the log-transformed target word frequency over the total neighborhood frequency, defined as the sum of the log-transformed frequencies of every member of the neighborhood (including the target). This method of frequency-weighting neighborhood density is commonly used in studies of competition-induced hyperarticulation (e.g., Munson, 2007; Scarborough, 2013; Buz and Jaeger, 2016). This measure is typically used in the choice of experimental materials to bin test items into ‘hard’ and ‘easy’ categories. Hard words have low frequency relative to their (relatively numerous) neighbors, while easy words have high frequency relative to their (relatively less numerous) neighbors (e.g., Wright, 2004). Log-transformed frequencies were extracted directly from SUBTLEX-US (Brysbaert and New, 2009). Henceforth, we will refer to these frequency-weighted ratios as 'neighborhood frequency' measures.

As an illustration of the simplest case, the minimal pair neighborhood frequency consists of the log-transformed word frequency of the target divided by the sum of the log-transformed frequencies of the target and its voicing minimal pair competitor. Thus, if the target does not have a voicing minimal pair competitor (e.g., *bright*), the minimal pair neighborhood frequency is 1 (log-transformed target frequency divided by log-transformed target frequency + 0). If the target does have a minimal pair competitor, however, the denominator will include the frequency of that competitor, will therefore be larger than the numerator, and the resulting value of the ratio will be less than 1. Consequently, these neighborhood frequency measures range in value from 0 to 1, where 1 indicates that the target has no neighbors as defined for that neighborhood[2].

We repeated our mixed effects analyses as reported above for these neighborhood frequency measures. As before, the competition metrics were centered, and each appeared in its own model as both a fixed effect and as a random slope on speaker. Results are presented for voiced (section 2.4.1) and voiceless (section 2.4.2) stops separately.

2.4.1 Voiced stops

As before, the voiced stop data included 2,267 observations meeting our criteria from 24 speakers, distributed over 293 lemmas. VIFs for all of the predictors of interest were less than 1.8, indicating low multi-collinearity between the factors of interest and control predictors. The frequency-weighted metric corresponding to all neighbors with the same voicing value as the target had the highest VIF of all factors of interest (1.73), while the frequency-weighted minimal pair factor had the smallest VIF (1.17).

[2]Mathematically, a value of 0 is impossible, but the value can come arbitrarily close to 0 in theory.

Corrected Akaike's Information Criterion comparison

The AIC_c comparison table for all voiced stop models is presented in Table 2.7. The model including the frequency-weighted minimal pair factor had the lowest AIC_c value ($\Delta AIC_c = 0$), and therefore the best overall fit to the data. The model including a factor for the second segment neighborhood frequency was the second-best model (second segment ND: $\Delta AIC_c = 4.44$), and evidence ratios indicate that there was about 9 times more evidence in favor of the minimal pair model. The only other model with limited support was the model including a factor for neighbors that share their place of articulation with the target (same place ND: $\Delta AIC_c = 4.52$). All other models had little or no support (all $\Delta AIC_c > 12$). This result suggests that a frequency-weighted measure of cue-specific minimal pair competition is a better predictor of voiced stop voice onset times than other frequency-weighted neighborhood measures, mirroring the results for unweighted neighborhood densities.

Rank	Model	K	AIC_c	ΔAIC_c	AIC_cWt	Cum.Wt	LL
1	Min. pair NF	20	-4132.31	0.00	0.82	0.82	2086.34
2	Second seg. NF	20	-4127.86	4.44	0.09	0.91	2084.12
3	Same place NF	20	-4127.78	4.52	0.09	0.99	2084.08
4	Same voice NF	20	-4120.05	12.26	0.00	1.00	2080.21
5	Diff. manner NF	20	-4119.53	12.78	0.00	1.00	2079.95
6	Base model	17	-4118.22	14.09	0.00	1.00	2076.24
7	First seg. NF	20	-4118.02	14.28	0.00	1.00	2079.20
8	Third-to-end NF	20	-4117.78	14.52	0.00	1.00	2079.08
9	Overall NF	20	-4116.44	15.86	0.00	1.00	2078.41
10	Same manner NF	20	-4115.58	16.73	0.00	1.00	2077.98
11	Diff. voice NF	20	-4115.37	16.94	0.00	1.00	2077.87
12	Diff. place NF	20	-4114.57	17.74	0.00	1.00	2077.47

Table 2.7: AIC_c comparison table for all models predicting word-initial voiced stop VOT/Stop-length ratios. Models are ranked in order of AIC_c value. K = the number of estimable parameters in the model. AIC_c = the corrected AIC value. ΔAIC_c = the change in corrected AIC value between each model and the top performing model. AIC_cWt = the relative likelihood that the present model is the best model, presented as a proportional weight. Cum.Wt = the cumulative AIC_c weight of the present model and all higher-ranked models. LL = the log-likelihood of the model.

The top model: frequency-weighted minimal pair competition

The frequency-weighted minimal pair model is summarized in Table 2.8. As predicted, the correlation between this factor and the VOT/Stop-length ratio is positive, indicating that voice onset time decreases as competition increases (i.e., as the frequency ratio decreases). This suggests that these words are contrastively hyperarticulated away from their voice onset time competitor. We evaluated whether the minimal pair existence factor significantly contributed to model fit using nested model comparison, and found that removal of the fixed effect of minimal pair existence significantly affected model fit ($\chi^2(1) = 17.13, p < 0.001$). As with the unweighted neighborhood densities, this indicates that cue-specific minimal pair competition is a significant predictor of initial voiced stop voice onset times.

2.4.2 Voiceless stops

As before, the voiceless stop data included 3,690 observations meeting our criteria from 24 speakers, distributed over 417 lemmas. VIFs for all of the predictors of interest were less than 1.6, indicating low multi-collinearity between the factors of interest and control predictors. The second segment neighborhood frequency metric had the highest VIF of all factors of interest (1.53), while the frequency-weighted minimal pair factor had the smallest VIF (1.10).

Corrected Akaike's Information Criterion comparison

The AIC_c comparison table for all voiceless stop models is presented in Table 2.9. The model including the overall neighborhood frequency had the lowest AIC_c value ($\Delta AIC_c = 0$), and therefore the best overall fit to the data. No other model had substantial support (all $\Delta AIC_c > 12$), and evidence ratios indicate that there was about 552 times more evidence in favor of the overall neighborhood frequency model over the next-best model (first segment NF: $\Delta AIC_c = 12.63$). This result stands in stark contrast to the results of our other three analyses, in which cue-specific minimal pair competitor existence or neighborhood frequency provided the most predictive models of voice onset times.

Variable	β	SE	t
(Intercept)	0.228	0.012	18.883
Min. Pair Neighb. Freq.	**0.084**	**0.020**	**4.296**
Stop Phoneme (b-d)	-0.108	0.008	-13.771
Stop Phoneme (g-d)	0.033	0.010	3.383
Speech Rate	0.026	0.017	1.493
Following Liquid (true-false)	0.085	0.008	10.603
Word Familiarity	-0.013	0.015	-0.906
Phonotactic Probability	-0.008	0.039	-0.202
Forward Probability	0.039	0.010	3.863
Backward Probability	-0.012	0.010	-1.252
Number of Syllables (1-2)	-0.007	0.007	-1.007
Previous Mention (true-false)	0.006	0.004	1.342
Syntactic Category (Noun-Adjective)	-0.003	0.009	-0.362
Syntactic Category (Adverb-Adjective)	0.018	0.013	1.318
Syntactic Category (Verb-Adjective)	0.025	0.011	2.218

(a) Fixed effects summary

Group	Variable	Variance	SD	Corr
Lemma	(Intercept)	0.0009	0.030	
Speaker	(Intercept)	0.0010	0.032	
	Min. Pair Neighb. Freq.	**0.0005**	**0.023**	**0.64**
Residual		0.0086	0.093	

(b) Random effects summary.

Table 2.8: Summaries of fixed and random effects for the model including minimal pair neighborhood frequency for voiced stops.

The top model: overall neighborhood frequency

The overall neighborhood frequency model is summarized in Table 2.10. As predicted, the correlation between this factor and the VOT/Stop-length ratio is negative, indicating that voice onset time increases as competition increases (i.e., as the frequency ratio decreases). We evaluated whether the overall neighborhood frequency factor significantly contributed to model fit using nested model comparison, and found that removal of the fixed effect of overall neighborhood frequency significantly affected model fit ($\chi^2(1) = 9.74, p < 0.01$).

Rank	Model	K	AIC_c	ΔAIC_c	AIC_cWt	Cum.Wt	LL
1	Overall NF	20	-6141.40	0.00	1	1	3090.82
2	First seg. NF	20	-6128.78	12.63	0	1	3084.50
3	Diff. voice NF	20	-6126.32	15.08	0	1	3083.27
4	Second seg. NF	20	-6123.87	17.53	0	1	3082.05
5	Diff. manner NF	20	-6122.92	18.48	0	1	3081.58
6	Same place NF	20	-6122.75	18.65	0	1	3081.49
7	Min. pair NF	20	-6122.59	18.82	0	1	3081.41
8	Same manner NF	20	-6122.18	19.22	0	1	3081.21
9	Diff. place NF	20	-6121.11	20.29	0	1	3080.67
10	Base model	17	-6119.05	22.35	0	1	3076.61
11	Same voice NF	20	-6117.03	24.38	0	1	3078.63
12	Third-to-end NF	20	-6116.15	25.25	0	1	3078.19

Table 2.9: AIC_c comparison table for all models predicting word-initial voiceless stop VOT/Stop-length ratios. Models are ranked in order of AIC_c value. K = the number of estimable parameters in the model. AIC_c = the corrected AIC value. ΔAIC_c = the change in corrected AIC value between each model and the top performing model. AIC_cWt = the relative likelihood that the present model is the best model, presented as a proportional weight. Cum.Wt = the cumulative AIC_c weight of the present model and all higher-ranked models. LL = the log-likelihood of the model.

A closer look at the neighborhood frequency metric

For the analyses of both frequency-weighted and unweighted competition metrics in the voiced stop data, as well as the analysis of unweighted measures in the voiceless stop data, cue-specific minimal pair competition provided the best predictor of voice onset time realizations. However, for frequency-weighted competition metrics in the voiceless stop data, the opposite measure as defined by our continuum of specificity was most predictive: overall neighborhood density.

Why might this result be so different from the others? One possibility lies in the distribution of data along the neighborhood frequency variable. A feature of the neighborhood frequency metric is that items with no neighbors have a neighborhood frequency of 1, because the ratio for words with no neighbors is the frequency of the word divided by itself. When items are included that have no neighbors, this can result in a distribution in

Variable	β	SE	t
(Intercept)	0.554	0.015	37.90
Overall Neighb. Freq.	**-0.048**	**0.015**	**-3.26**
Stop Phoneme (p-t)	-0.132	0.008	-16.52
Stop Phoneme (k-t)	-0.015	0.008	-1.91
Speech Rate	-0.034	0.015	-2.29
Following Liquid (true-false)	0.043	0.007	6.47
Word Familiarity	-0.021	0.013	-1.65
Phonotactic Probability	-0.003	0.035	-0.08
Forward Probability	0.008	0.009	0.92
Backward Probability	-0.014	0.008	-1.61
Number of Syllables (1-2)	0.007	0.006	1.20
Previous Mention (true-false)	-0.0008	0.004	-0.22
Syntactic Category (Noun-Adjective)	0.004	0.010	0.37
Syntactic Category (Adverb-Adjective)	0.012	0.023	0.55
Syntactic Category (Verb-Adjective)	-0.008	0.011	-0.67

(a) Fixed effects summary.

Group	Variable	Variance	SD	Corr
Lemma	(Intercept)	0.0010	0.031	
Speaker	(Intercept)	0.0016	0.040	
	Overall Neighb. Freq.	**0.0013**	**0.036**	**-0.16**
Residual		0.0102	0.101	

(b) Random effects summary.

Table 2.10: Summary of fixed and random effects for the model including overall neighborhood frequency for all voiceless stops.

which the majority of values are concentrated at the lower range, with an isolated peak at 1 corresponding to words with no neighbors. In the voiceless dataset, the bulk of the data is distributed between overall neighborhood frequency values of 0 and 0.2 (Figure 2.4). There is a substantial gap in the observed values in the upper range, followed by a single peak corresponding to a neighborhood frequency ratio of 1. Extreme values can exert undue leverage on regression models, and notably, studies using neighborhood frequency to contrast 'hard' versus 'easy' words have excluded words with no neighbors, i.e. those with a neighborhood frequency value of 1 (e.g., Wright, 2004; Munson and Solomon, 2004; Munson, 2007; Scarborough, 2013). To ask whether words with no neighbors contributed to this anomalous result, we removed from the dataset the 337 observations of the 69 lemmas with no neighbors and repeated the analysis. 67 of these lemmas were bisyllabic, and the remaining 2 monosyllabic lemmas were the phonotactically unusual *prompt* and *puke*. The resulting dataset retained 3,353 observations from 24 speakers, distributed over 348 lemmas.

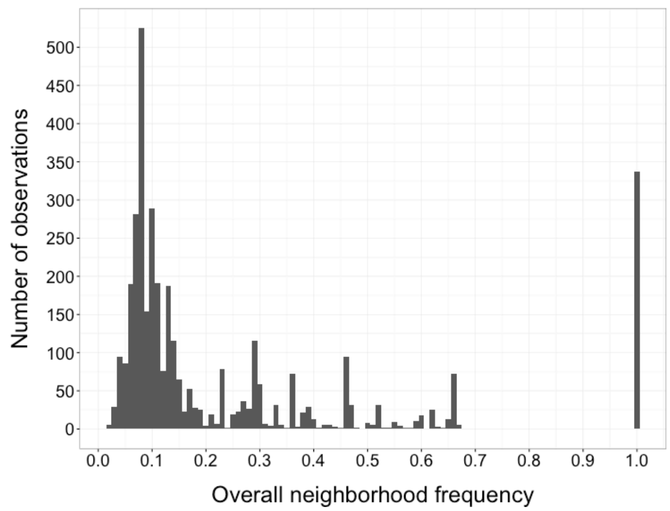

Figure 2.4: The distribution of observations in the voiceless stop dataset by overall neighborhood frequency.

The AIC_c comparison table for all voiceless stop models is presented in Table 2.11. The model including the overall neighborhood frequency still had the lowest AIC_c value ($\Delta AIC_c = 0$), and therefore the best overall fit to the data. However, the overall neighborhood frequency factor no longer contributed significantly to model fit ($\chi^2(1) = 2.74$, $p > 0.05$; see A for details). Furthermore, two other models now varied minimally from the overall neighborhood frequency model: the first segment neighborhood frequency model ($\Delta AIC_c = 2.07$) and the minimal pair neighborhood frequency model ($\Delta AIC_c = 2.27$). Evidence ratios indicate that there was about 3 times more evidence in favor of the overall neighborhood frequency model relative to both the first segment neighborhood frequency model and the minimal pair neighborhood frequency model. These results indicate that overall neighborhood frequency is still the most predictive of these frequency-weighted competition metrics in this voiceless stop dataset, but the effect of this factor was not significant and the evidence in support of this model over the two subsequent models in the AIC_c ranking is limited. We confirmed that unweighted minimal pair existence was still significantly predictive of voice onset times in this subset of the voiceless stop data ($\chi^2(1) = 5.38$, $p < 0.05$). This suggests that i the lack of significance for the overall neighborhood frequency factor is not purely due to a reduction of power, since a corresponding reduction of power does not impact the significance of the unweighted minimal pair factor, and ii the weighting of competition metrics for frequency adds unnecessary noise into the data, since even the best frequency-weighted model is less predictive of voice onset times than the unweighted minimal pair factor.

Rank	Model	K	AIC_c	ΔAIC_c	AIC_cWt	Cum.Wt	LL
1	Overall NF	20	-5555.97	0.00	0.38	0.38	2798.11
2	First seg. NF	20	-5553.90	2.07	0.13	0.51	2797.08
3	Min. pair NF	20	-5553.71	2.27	0.12	0.63	2796.98
4	Diff. voice NF	20	-5553.42	2.55	0.11	0.74	2796.84
5	Same place NF	20	-5552.42	3.55	0.06	0.80	2796.34
6	Second seg. NF	20	-5552.10	3.87	0.05	0.85	2796.18
7	Base model	17	-5551.72	4.26	0.04	0.90	2792.95
8	Diff. manner NF	20	-5551.33	4.46	0.04	0.94	2795.79
9	Same manner NF	20	-5551.05	4.92	0.03	0.97	2795.65
10	Diff. place NF	20	-5549.75	6.23	0.02	0.99	2795.00
11	Same voice NF	20	-5548.95	7.02	0.01	1.00	2794.60
12	Third-to-end NF	20	-5546.59	9.38	0.00	1.00	2793.42

Table 2.11: AIC_c comparison table for all models predicting word-initial voiceless stop VOT/Stop-length ratios in the subset of data for which overall neighborhood frequency is less than 1. Models are ranked in order of AIC_c value. K = the number of estimable parameters in the model. AIC_c = the corrected AIC value. ΔAIC_c = the change in corrected AIC value between each model and the top performing model. AIC_cWt = the relative likelihood that the present model is the best model, presented as a proportional weight. Cum.Wt = the cumulative AIC_c weight of the present model and all higher-ranked models. LL = the log-likelihood of the model.

2.4.3 Summary and discussion

The results for voiced stops were nearly identical for both frequency-weighted and unweighted measures. The competition metric corresponding to cue-specific minimal pair competition produced the best models in both cases, where minimal pair competition correlated with shorter voice onset times. The results for voiceless stops, however, varied substantially. While the analysis for unweighted competition metrics indicated that cue-specific minimal pair existence was the best predictor of voiceless stop voice onset times, the analysis for frequency-weighted competition metrics provided more evidence in favor of the overall neighborhood frequency measure. However, we noted that the neighborhood frequency metric results in a value of 1 for words without any neighbors, regardless of their frequency. This cluster of data at one extreme of the distribution creates the possibility for leverage, where a relatively small number of observations can have undue influence on the regression model. What is more, studies using this neighborhood frequency measure to categorize words as ‘easy’ or ‘hard’ have excluded words with no neighbors (e.g., Wright, 2004; Munson and Solomon, 2004; Munson, 2007; Scarborough, 2013). We repeated our analysis of the voiceless data with these words excluded and found that the overall neighborhood frequency model had the greatest support, but the effect of overall neighborhood frequency was no longer significant. These results suggest that, at least within this dataset, neighborhood frequency measures are not robustly predictive of voice onset time in voiceless stops.

We also found limited evidence in favor of both the first segment and minimal pair neighborhood frequency measures predicting voice onset times in voiceless stops (first segment: $\Delta AIC_c = 2.07$; minimal pair: $\Delta AIC_c = 2.27$). These findings parallel those of Fricke (2013), who reported that overall neighborhood density, minimal pair competitor existence, and the number of competitors differing in their onset (“rhyme neighbors” in her terminology) all correlated with hyperarticulation of voiceless stop voice onset times in the Buckeye Corpus (see also Fricke et al., 2016). However, Fricke reported that onset competition was a better predictor of voice onset time than either minimal pair competitor existence or overall neighborhood density, concluding that each of these other measures correlate with voice onset time because they also correlate with the position-dependent measure of onset competition. Instead, we found that, when frequency-weighted, the overall neighborhood density measure was a better predictor than the other two, but was not significant when we excluded words with no neighbors. Some methodological considerations may help to explain this different result. For example, Fricke only included monosyllabic words beginning with simplex onsets, while we included bisyllabic words and words with complex onsets. This difference affects our various neighborhood measures, particularly the neighborhoods corresponding to onset/first segment competition.

2.5 Discussion

We identified a set of lexical competition metrics that sampled a conceptual space between the phonetically specific measure of cue-defined minimal pair existence, and the phonetically more general measure of neighborhood density. These measures were tested for their ability to predict voice onset time in voiced and voiceless word-initial stops of conversational English. Previous studies have suggested that competition from a cue-defined minimal pair competitor induces contrastive hyperarticulation of voice onset time (e.g., Baese-Berk and Goldrick, 2009; Schertz, 2013; Buz et al., 2016), and we found support for this hypothesis in the form of shorter voice onset times in voiced stops and longer voice onset times in voiceless stops (see also Wedel et al., 2018). However, previous studies using a variety of speech elicitation paradigms have suggested that less phonetically specific measures of competition also correlate with hyperarticulation of voice onset time in voiceless stops (e.g., Kirov and Wilson, 2012), in some cases more strongly than the cue-specific minimal pair competitor measure (Fricke, 2013; Fricke et al., 2016; Fox et al., 2015). In this natural speech dataset, we found that these other metrics of competition were less predictive of the realization of voice onset time in both voiced and voiceless stops.

We also conducted analyses in which we weighted our competition metrics for the relative frequencies of neighbors. For the voiced stops, this neighborhood frequency approach did not alter the results. The model including a frequency-weighted factor for the existence of an initial stop voicing minimal pair remained the most predictive of our models. For voiceless stops, however, overall neighborhood frequency provided the most predictive model. Upon further inspection, we noted the possibility that the extreme neighborhood frequency values contributed by words with no neighbors could unduly leverage model outcomes. With these data removed, the model including overall neighborhood frequency remained the most predictive in terms of AIC_c, but the contribution of overall neighborhood frequency to the model was no longer significant. Further, the models including factors for first segment and minimal pair neighborhood frequencies were minimally different from the overall neighborhood frequency model in terms of AIC_c.

2.5.1 Alternative metrics of lexical competition

The competition measures tested in this study represent only a subset of the large and multidimensional hypothesis space surrounding lexical competition effects. One way to expand this hypothesis space is to extend the continuum beyond a single-phoneme edit distance between target and competitor, so that similar words with a multi-phoneme edit distance can still contribute to competition. An example of such a measure in the literature is to weight lexical items for their Levenshtein edit distance, so that edit distances greater than 1 contribute to the neighborhood, but less so than more closely related words (Yarkoni et al., 2008). A related, but different approach is to weight the neighbors for proportion of position-dependent segmental overlap with the target (Goldrick et al., 2010). Using either of these approaches, *trap* would not only have such neighbors as *trip*, *tap*, and *track*, but also neighbors such as *trick* and *tarp*, for which more than one segment are different. Proportion of segmental overlap has been found to be predictive of spoken and written errors in subjects with acquired language impairment, especially when additionally weighted for frequency, syntactic category, and onset overlap (i.e., the Lex-Form Composite: Goldrick et al., 2010). However, broader competition metrics such as these are less phonetically specific than the range of neighborhood measures tested in our study. We find that hyperarticulation of word-initial voice onset time in conversational English is primarily predicted by a competitor differing solely in word-initial voicing, suggesting that these broader alternatives should be less predictive of contrastive hyperarticulation in voice onset time.

Another competition metric from the literature compares the target to all words in the lexicon, weighting them for their perceptual similarity (Phi-square density: Strand and Sommers, 2011). Under this approach, *fist* and *fish* would be considered stronger competitors than *fist* and *fit*, because /s/ and /ʃ/ are perceptually similar, while the absence of /s/ is perceptually salient. This measure has been found to be predictive of word recognition accuracy, but whether it contributes predictive power above and beyond overall neighborhood density is unclear, and it was not found to be significantly predictive of word durations in the Buckeye corpus (Gahl and Strand, 2016). Although the phi-square density measure is gradiently sensitive to similarity, it considers perceptual relationships without respect to a particular cue. As such, neighbors that are perceptually similar in the segment or cue of interest, such as the cue-specific minimal pair competitor, are given no priority over neighbors differing in other positions or cues. It is unclear to what extent this measure should correlate with contrastive hyperarticulation of individual cues, but it may provide a fruitful model for studying the role of perceptual similarity in contrastive hyperarticulation (cf. Buz et al., 2016).

2.5.2 Other cue types

If these results are representative, we expect to find similar contrastive hyperarticulation of other types of cues, where this hyperarticulation specifically increases phonetic distance between competitors. In another study of conversational speech, Wedel et al. (2018) found evidence of hyperarticulation in vowels in the form of F1–F2 Euclidean distance. Vowels in words with minimal pairs defined by a nearby competitor vowel (e.g., *pit* ~ *pet*) were farther apart in F1–F2 space than the same vowels in words without such minimal pairs (e.g., *ship* ~ **shep*). This difference was not explained by more general competition in the form of overall neighborhood density. In conjunction with the results reported here, this suggests that contrastive hyperarticulation of phonetic cues arises in response to competition with nearby competitors within that cue, and that this holds across two very different cue types (voice onset time and F1–F2 Euclidean distance).

2.5.3 Cognitive mechanisms

What do our results have to say about the cognitive mechanisms proposed to underlie competition-based hyperarticulation? In production-internal accounts that appeal to interactive cascading activation, the set of competitors has often been defined broadly as the set of lexical neighbors differing by a single phoneme edit distance in any position. More recent accounts, however, have been modified to include differential neighborhood effects at different positions within words (e.g., Vitevitch and Chu, 2004; Goldrick et al., 2010; Fricke, 2013). For example, Fricke (2013) proposed that interactive cascading activation during motor planning is processed on a segment-by-segment basis. Each progressive segment is articulated as soon as competition is resolved, with greater competition leading to higher activation levels and, consequently, longer phonetic durations. According to this Articulate As Soon As Possible Principle, competition is sensitive to segmental position within the word, but is insensitive to the phonetic relationships among competitors within that segmental position. Consequently, in the case of word-initial voice onset time, this model predicts that the overall competition within the word onset should be the best predictor of competition-induced hyperarticulation.

Our findings did not support this hypothesis. In general, the existence of the minimal pair competitor defined by the initial stop voicing contrast was most predictive of voice onset times for both voiced and voiceless stops. This result is difficult to reconcile with any model of competition effects that does not consider the specific phonetic contrast between target and competitor. On the other hand, this result is consistent with communicative and listener-internal approaches, according to which the contrastive cues to phonemic and/or lexical categories can be specifically targeted for hyperarticulation (Pierrehumbert, 2002; Wedel, 2006; Buz et al., 2016; Hall et al., 2016). Notably, there is evidence that these cues may be especially targeted for hyperarticulation when they are most confusable. Buz et al. (2016) elicited productions of voiceless stop-initial monosyllabic words of English in a cooperative task. Speakers produced longer voice onset times on average when the minimal pair competitor was present in the experimental context, replicating past findings (e.g., Baese-Berk and Goldrick, 2009). Crucially, in a post-hoc analysis they found evidence that speakers were achieving this shift in average voice onset time not by increasing the target voice onset time of their productions, but by reducing the distribution of productions near the category boundary, where small changes to voice onset time have detectable effects on perception (McMurray et al., 2002). This suggests that speakers specifically avoid production variants that would result in greater perceptual similarity with a competitor.

If speakers contrastively hyperarticulate by avoiding more confusable productions, we expect that contrastive hyperarticulation may be less evident in conditions favoring clear-speech hyperarticulation, in which category contrasts may be inherently clearer. In elicited speech paradigms such as those reported above, voice onset times might be distinct enough a priori that the voiced and voiceless categories are completely non-overlapping[3]. In such a case, speakers may not attempt to hyperarticulate further, as the added effort would offer little in communicative benefit. On the other hand, if less formal registers induce greater overall reduction, we would expect phonetically contrastive cues to be less distinctive overall. In such a scenario, increasing articulatory effort to enhance this contrast would be worthwhile in words with minimal pair competitors defined for the relevant cue, but not for words without such a competitor.

What is more, the communicative aspect of this story is likely playing a further role. The studies that have found contrastive effects most consistently have involved (i) communication with a (sometimes simulated) human partner or (ii) clarifications of previous utterances (e.g., Baese-Berk and Goldrick, 2009, study 2; Kirov and Wilson, 2012; Schertz, 2013; Seyfarth et al., 2016; Buz et al., 2016). These conditions highlight the possibility that hyperarticulation is part of a communicative goal (Buz et al., 2016; Hall et al., 2016). Our conversational data similarly involve communication with an interlocutor (see also Wedel et al., 2018). In contrast, studies that have not found evidence of contrastive hyperarticulation of the kind reported here have largely used word or sentence reading tasks to elicit

[3]Unfortunately, most of the relevant studies have investigated only voiced or voiceless stops, but not both. The one exception is Schertz (2013), who reports changes to voice onset time for each condition, but not raw voice onset time values. Consequently, it is difficult to test this hypothesis based on the existing data.

productions (e.g., Goldrick et al., 2013; Fox et al., 2015; but see Ohala, 1994. For studies finding minimal pair effects using list reading paradigms, see Baese-Berk and Goldrick, 2009, study 1; Peramunage et al., 2011). It is possible that, in such tasks, cue-defined minimal pair competitor existence fails to predict contrastive hyperarticulation as reliably as in more communicative tasks because there would be little communicative benefit to this hyperarticulation[4].

2.5.4 Implications for sound change

In what follows, we assume that consistent utterance-level bias in phonetic output can shift long term mental representations, which can in turn shape the trajectory of sound change in a speech community (discussed in Wedel, 2012 and Seyfarth, 2014). In this dataset, we observed that only competitors specifically distinguished from target words by the voice onset time cue robustly predicted contrastive hyperarticulation of that cue. This suggests in turn that contrastive hyperarticulation—and associated longer-term sound change—will be best predicted by the finer grained, cue-specific minimal pair relationships in the lexicon, rather than more abstract neighborhood relationships. Conversely, it suggests that more general measures of the functional load of phonetic cues such as phoneme-level entropy (Hockett, 1966; Surendran and Niyogi, 2006) will be predictive of sound change only through their correlation with the probability that a cue defines a minimal pair contrast. This is consistent with cross-linguistic evidence that phoneme contrasts distinguishing more minimal pairs are significantly less likely to merge over time (Wedel et al., 2013b).

Further, the finding that competition results in significant shortening of voice onset time for voiced stops supports the notion that contrastive hyperarticulation creates greater perceptual distance to a competitor (Seyfarth et al., 2016; Buz et al., 2016; Wedel et al., 2018; see Hall et al., 2016, for discussion), rather than resulting in a general exaggeration of du-

[4]In the case of Fox et al. (2015), greater phonological neighborhood density was nonetheless predictive of longer voice onset time realizations (see also Fricke et al., 2016). Alongside previous findings correlating neighborhood density with longer phonetic durations (see Fricke, 2013, chs. 2 & 6 for discussion), this raises the question of whether cue-defined minimal pair competition and overall neighborhood density index different kinds of competition and, consequently, different kinds of hyperarticulation.

ration or extent of articulatory gestures. Slow or clear speech conditions, which are associated with increased phonetic durations, have not been found to correlate with shortening of voice onset time for initial voiced stops (Miller et al., 1986; Kessinger and Blumstein, 1997). This in turn suggests that phonetic distinctions that are on average more confusable should be more likely to trigger contrastive hyperarticulation. As reviewed above, this is supported by evidence that, for voiceless stops, contrastive hyperarticulation primarily suppresses shorter voice onset time productions that would be close to the voiced–voiceless category boundary (Buz et al., 2016). Taken together, these two observations predict that contrastive hyperarticulation will exert a greater influence on the trajectory of change for cue-distinctions which (i) are more perceptually confusable, and (ii) distinguish a greater number of cue-specific minimal pairs. A corollary of this prediction is that contrastive hyperarticulation is more likely to exert an effect through modulating production of reduced/rapid speech, where phonetic distinctions tend to be reduced, rather than through effects on careful/slow speech, where phonetic distinctions tend to be more robust.

2.6 Conclusion

We have argued that voice onset time is contrastively hyperarticulated in conversational speech in a way that increases the perceptual distance between lexical minimal pairs defined specifically for the voice onset time cue. The effect of minimal pair existence was more robust for voiced stops, for which voice onset times were found to decrease under competition. These results are most consistent with models of hyperarticulation effects that consider the fine-grained phonetic relationships among competitors, such as listener-oriented models that consider the communicative goals of speakers (e.g., to be understood). Furthermore, this is consistent with prior work arguing that fine-grained difference in individual-level productions of lexical minimal pairs can help to explain patterns of sound change such as phoneme merger (Wedel et al., 2013b). In addition, we noted that prior work has emphasized (i) voiceless stop hyperarticulation, for which increased voice onset times are predicted under both contrastive and clear-speech hyperarticulation, and (ii) elicited speech paradigms, some of which may not provide motivation for contrastive hyperarticulation in speakers (e.g., word list or sentence reading tasks), and which may instead promote clear-speech hyperarticulation more generally. We argued that such studies are not ideally suited to inducing or detecting contrastive hyperarticulation, potentially explaining why some of these studies find small effects, but others do not. Furthermore, we noted that the pattern of results in the literature and in the data presented here suggest the importance of considering language as a tool for communication, where speakers hyperarticulate cues that facilitate effective communication, but are less likely to hyperarticulate when there is little benefit to doing so.

CHAPTER 3

STUDY 2: THE OUTCOMES OF CONTRASTIVE HYPERARTICULATION

3.1 Introduction and Background

In the languages of the world, voiceless obstruents tend to correlate with higher fundamental frequency (F0) on the following segment than their voiced counterparts (e.g., House and Fairbanks, 1953; Löfqvist, 1975; Löfqvist et al., 1989; Kingston and Diehl, 1994; Kirby et al., 2015)[1]. For stop consonants in particular, the voicing of the stop has been shown to influence following F0 in both true "voicing languages" such as Spanish and Afrikaans, for which phonologically voiced stops are typically phonetically voiced (i.e., the vocal folds vibrate during production of the stop closure) and phonologically voiceless stops are phonetically unvoiced (i.e., the vocal folds do not vibrate during production of the stop closure; e.g., Dmitrieva et al., 2015; Coetzee et al., 2018), as well as in "aspiration languages" such as English and Korean, for which both phonologically voiced and voiceless stops are typically phonetically unvoiced (i.e., the vocal folds do not vibrate during production of the stop closure for either voiced or voiceless stops; e.g., Dmitrieva et al., 2015;

[1]The effects reported in these and other sources were all measured in syllables with simplex onsets, and thus the following segment on which F0 perturbations were recorded was always a vowel. In fact, this author knows of no study specifically examining whether these effects hold for sonorous consonants in addition to vowels. However, there is no a priori reason to believe that this effect does not apply to all sonorous segments. See sections 3.1.1 and 3.1.2 for further discussion.

Bang et al., 2018). It has been argued that this cross-linguistic correlation is due to physiological factors involved in the suppression of voicing (e.g., vocal fold tightening: Löfqvist et al., 1989; Hoole and Honda, 2011, discussed further in section 3.1.1). As such, the correlation between obstruent voicing and following F0 may have an automatic, principled, articulatory cause. Nonetheless, that this correlation holds for stop consonants in many aspiration languages (such as English), for which both phonologically voiced and voiceless stops are phonetically unvoiced, suggests the possibility that this phonetic source of covariation has become phonologized in these languages (Hombert et al., 1979; Blevins, 2004; Hyman, 2013, among others)[2]. Some researchers have therefore suggested that physiological factors provide a principled and systematic articulatory *source* for obstruent voicing-dependent F0 variation, but that speakers capitalize on the resulting correlation to enhance phonemic contrasts in their language, even when the underlying articulatory source is no longer relevant (Hoole and Honda, 2011; Dmitrieva et al., 2015; Clayards, 2018; see sections 3.1.1 and 3.1.2 for further argumentation).

In light of this possibility, this section provides a brief review of the literature on the relationship between obstruent voicing and following F0, with special attention to stop consonants (the focus of this chapter). First, we review work on the phonetic/physiological source of this relationship. Second, we review work suggesting that, in many languages, F0 is contrasted phonologically in the absence of a clear phonetic/physiological source. Third, we review work on the relationship between voice onset time (VOT) and F0 as cues to initial stop voicing. Finally, we review work on the role of lexical-phonological contrast on the realization of phonetic cues.

[2] Of course, it is also possible that disparate sources of F0 variation are at play in voicing languages and aspiration languages, such as vocal fold tightening versus glottal spreading. However, as is discussed below, vocal fold tightening appears to be responsible for at least some voicing-dependent F0 variation in aspiration languages (Hoole and Honda, 2011), and there is evidence that aspiration-related sources of potential voicing-dependent F0 variation such as glottal spreading are insufficient to explain the pitch variations found in these languages (Kingston and Diehl, 1994). Note also that in many languages such as English, aspiration is not consistent across same-voicing allophones of stop consonants, but post-stop F0 is consistently higher following all phonologically voiceless allophones (see Kingston and Diehl, 1994, for discussion, among others).

3.1.1 The phonetic relationship between voicing and F0

A number of physiological sources have been suggested to account for systematic variation in F0 following voiced versus voiceless obstruents. For example, Kingston and Diehl (1994) note three such possibilities:

1. Lowering of the larynx during the initiation of voicing, which increases the size of the oral cavity and thereby decreases F0 surrounding voiced obstruents
2. Widening of the glottis during voiceless obstruent productions, which increases transglottal airflow and thereby increases F0 surrounding voiceless obstruents
3. Tightening of the vocal folds in order to suppress voicing, which causes higher frequency vocal fold vibrations and thereby increases F0 surrounding voiceless obstruents

Perhaps most promising of these possibilities is the role of the cricothyroid (CT) muscle in the tightening of the vocal folds in order to suppress voicing. Löfqvist et al. (1989) measured activation levels of the CT muscle during productions of syllables beginning with voiced and voiceless obstruents by two speakers of American English and one speaker of Dutch. They found that CT activity was greater, and F0 higher, following voiceless obstruents relative to voiced ones. Hoole and Honda (2011) measured CT activation levels and F0 realizations in three German-speaking subjects. They found that CT activation and following F0 were higher for voiceless obstruents compared to voiced obstruents, consistent with Löfqvist et al. (1989). Furthermore, Hoole and Honda (2011) reported that CT muscle activation could continue beyond the articulation of a voiceless stop closure and into the following vowel (though the activation peak tends to be during the stop closure). They suggest that this may reflect an attempt on the part of speakers to precisely control the onset of voicing with a crisp transition from voiceless to voiced (by raising the pressure threshold required for initiation of vocal fold vibration; see Hoole and Honda, 2011, for discussion). Regardless of why they do it, that some speakers maintain CT activity into the following vowel may help to explain why heightened F0s can be detected even in vowels following long-lag voiceless stops.

The primary argument made by Hoole and Honda (2011) is that obstruent voicing and following F0 are fundamentally correlated for physiological reasons, but that (some) speakers capitalize on this correlation to enhance contrasts in the language. As we will see in the following sections, this view is supported by evidence that the phonological voicing status of obstruents may have a greater influence on F0 than the phonetic voicing status of those obstruents (Sec. 3.1.2) and that the production and perception of the phonetic outcomes of obstruent voicing may vary more between individual speakers than they do between the voicing categories themselves (Sec. 3.1.3).

3.1.2 The phonological relationship between voicing and F0

There is substantial evidence that phonological factors may play a larger role than phonetic or physiological factors in determining the realization of F0 after stop consonants. One piece of evidence already noted above is that F0 is lower for phonologically voiced stops even in aspiration languages like English, where these stops are phonetically voiceless. Indeed, Ohde (1984) examined voice onset time and F0 measurements in productions of word initial voiced stops (/b/, /d/, /g/), voiceless unaspirated stops from /s/-initial onset clusters (/sp/, /st/, /sk/), and voiceless aspirated stops (/p^h/, /t^h/, /k^h/). He found that F0 was similar across voiceless aspirated and unaspirated stops, but considerably lower in voiced stops. This suggests that F0 is conditioned more by the phonological voicing value of the stop than by other phonetic correlates such as VOT (he also found that VOT was similar across voiced and voiceless unaspirated stops, but considerably longer in voiceless aspirated stops).

In addition, the correlation between stop voicing and following F0 is weak and temporally limited in a number of tonal languages, where F0 is already used to cue other phonological contrasts (Gandour, 1974; Hombert, 1977; Francis et al., 2006). Furthermore, Kingston and Diehl (1994) present evidence from a number of languages that phonetic factors, such as degree of glottal aperture and voice onset time, explain less variability in F0 than the phonological voicing status of the obstruents surrounding the vowel.

One good test of this hypothesis comes from examining phonetically similar segments with different phonological voicing statuses across different languages. One such test comes from Dmitrieva et al. (2015), who investigated patterns of post-stop F0 in Spanish and English. The English voicing contrast is primarily one of aspiration (short-lag vs. long-lag VOT), but where voiced stops in utterance-initial position can sometimes be realized as phonetically voiced. The Spanish voicing contrast is one of true phonetic voicing. Thus, phonetically voiced stops and phonetically unvoiced short-lag stops are allophones of the voiced stop category in English, but belong to distinct phonemic categories in Spanish. Dmitrieva et al. found that, despite comparable VOTs, F0 varied significantly only when the contrast was phonemic (in Spanish) and not when it was allophonic (in English), suggesting that the phonological voicing status of a stop is more predictive of F0 patterns than the VOT of the stop.

Another useful test of the hypothesis comes from languages in which phonological generalizations about F0 are extended to cases in which the actual phonetic effects on F0 run directly counter to the phonological generalization. One such example comes from Zina Kotoko (Mielke, 2004). As described by Mielke, Zina Kotoko is a tone language in which the recent past verbal inflection leads to a lowering of mid tones to low tones following voiced obstruents in the first syllable of the inflected verb. Of particular note, however, is that this process is generalized to implosives, a category of obstruent known to cause *raising* of F0 on following vowels (Hyman and Schuh, 1974). In this case, then, speakers have generalized the F0-lowering process to a phonologically related case that should, ceteris paribus, produce the opposite phonetic effect.

These examples suggest that the correlation between obstruent voicing and F0 correlates better with phonological factors than with phonetic or physiological factors. This is indicative of phonologization, a process whereby phonetic side-effects of articulations become reanalyzed as the phonetic target of the articulation, and take on phonological importance in the language (e.g., Blevins, 2004; Hyman, 2013). In the case of consonant voicing effects on F0, the control of voicing in obstruents likely leads to systematic effects on vocal fold tension, and thereby F0. In many languages, this phonetic side-effect appears to have been phonologized as an integral part of the contrast.

3.1.3 Individual variation in the use of F0 to signal voicing contrasts

Despite consistent phonological effects of voicing on F0, individuals vary substantially in how they weigh F0 against other cues to stop voicing contrasts, such as VOT (for a convenient case in point, see Kingston, 2007). Bang et al. (2018) investigated the effect of speaker age and gender on the realization of VOT and F0 in tense, lax, and aspirated stop consonants in Seoul Korean. They found that the lax/aspirated stop contrast was more strongly cued by VOT in older speakers, but by F0 in younger speakers. Furthermore, they found that this shift in cue weighting was more advanced for female speakers than for males. Coetzee et al. (2018) investigated the effect of age on the use of pre-voicing (negative VOT) versus F0 as cues to stop voicing in labial and alveolar stops in Afrikaans (there is no phonological velar stop voicing contrast in Afrikaans, except in loan words). They found that all of their speakers produced significant F0 contrasts and relied on F0 to signal voicing contrasts in perception, but that older speakers were far more likely to also pre-voice their stops in production and rely on pre-voicing in perception. The authors of both of these studies argue that these patterns of variation reflect language change in progress (specifically, tonogenesis, a kind of phonological restructuring leading to phonological tone), though at possibly different stages for each of these languages (Bang et al., 2018; Coetzee et al., 2018).

Less systematic variation across speakers in the use of VOT and F0 as signals to a voicing contrast has also been noted in languages that do not show signs of phonological restructuring. As noted above, Hoole and Honda (2011) investigated levels of cricothyroid (CT) activation during the production of stop consonants in speakers of German (note that, like English, German stop voicing is primarily cued by short-lag versus long-lag VOT).

They found increased CT activation and higher F0 following voiceless stops relative to voiced stops. However, they found that speakers varied considerably in how long they maintained activation of the CT muscle (and subsequently maintained higher F0) during the vowel following the voiceless stop. They interpreted their findings as evidence for a physiological source of voicing-contingent F0 variation, but where certain speakers actively enhance these effects beyond what is physiologically necessary.

Individual variation in the production of F0 as a voicing cue in English

Researchers investigating the relative weights of VOT and F0 in production and perception of English have also found considerable individual-level variation. On the production side, Clayards (2018) investigated the roles of VOT, following vowel onset F0, and following vowel duration as cues to initial labial stop voicing contrasts in English using a word list reading task. In particular, she looked at talker and token variability in the production of these cues. She found that (i) variation was greater between talkers than between tokens, (ii) that correlations between cues were more consistent within talkers than within stop voicing categories, and (iii) although all speakers appeared to utilize VOT above either F0 or vowel duration to signal voicing contrasts, the relative strength of the VOT cue varied substantially across individual speakers, whereas the relative strength of the other two cues varied less. These patterns indicate that individuals vary considerably in how they utilize the multiple cues to stop voicing contrasts, but that this variability is actually greatest in VOT itself, and individual speakers themselves are relatively systematic from token to token.

Further evidence that the relationship between VOT and F0 is variable in English comes from the different correlations reported for these factors across different studies. When voiced and voiceless production data from English are viewed together, there is a positive correlation between VOT and F0 as we would expect given that VOT is longer and F0 is higher in words with initial voiceless stops relative to words with initial voiced stops (e.g., Shultz et al., 2012; Dmitrieva et al., 2015; Clayards, 2018). Within categories, however, the picture is different. For voiceless stops, Dmitrieva et al. (2015) reported a significant *negative* correlation between VOT and F0 in an analysis of data from Shultz et al. (2012), which they interpreted as a sign of a trade-off relationship between these cues whereby shorter voiceless VOT values are compensated for by higher following F0 values. Dmitrieva et al. (2015) note, however, that this correlation was quite weak. Clayards (2018), on the other hand, failed to find any significant correlation between VOT and following F0 within the voiceless stop category. The different results between these two studies may be due to methodological factors, such as task effects or materials, or due to individual variation and random sample biases. Despite these differences, however, the phonological effect of the voicing category on the realization of F0 is consistent across languages and for English specifically. Indeed, regardless of whether F0 is correlated with VOT, it would appear that the phonological voicing status of the stop is a more robust predictor of F0 realizations than VOT alone (see especially Dmitrieva et al., 2015). This lends support to the view that F0 is independently manipulated as a cue to stop voicing by at least some speakers of English (cf. Hoole and Honda, 2011, for German).

The role of F0 in the perception of voicing in English

On the perceptual side, a number of studies have found that listeners of English will use F0 as a cue to a stop voicing contrast only when VOT is ambiguous. Abramson and Lisker (1985) investigated listener voicing judgments to synthetically altered syllables in a two-alternative forced choice task. Using CV syllables consisting of a labial stop followed by the vowel /a/, they produced stimuli with a variety of VOT values and a variety of post-stop F0 values. Manipulating these two phonetic properties independently, they found that listeners responded very consistently and in line with the VOT for most syllables. However, when VOT was closest to the typical voicing category boundary for labial stops of English (20ms), adjustments to the F0 value affected listeners judgments significantly, such that higher F0 values led to a greater proportion of voiceless judgments. They interpreted this as evidence that F0 is a redundant, or secondary, cue to initial stop voicing contrasts that listeners only make use of when VOT itself is not sufficient (see also Whalen et al., 1990).

In a subsequent study, however, Whalen et al. (1993) showed that F0 affects *reaction times* to such syllables even when VOT is not ambiguous. That is, even though listeners would consistently rate VOT values of, e.g., 15ms as voiced and 35ms as voiceless regardless of the following F0 value, F0 values nonetheless affected listeners' reaction times to those stimuli. For voiced stops, the lowest F0 value (98 Hz) resulted in faster reaction times to stops with 5-20ms VOTs. For voiceless stops, the highest F0 value (130 Hz) resulted in faster reaction times to stops with 20-50ms. They interpreted this result as evidence that F0 affects listeners' judgments about preceding stop voicing even when VOT is unambiguous; the effect is simply only detectable in reaction times to stimuli.

Taken all together, the results for English suggest a picture where (i) VOT is the primary signal to word-initial stop voicing contrasts for both speakers and listeners (Lisker and Abramson, 1967; Abramson and Lisker, 1985; Whalen et al., 1993; Clayards, 2018); (ii) F0 is a secondary (or redundant) signal to this contrast, with listeners attending to F0 in all cases (Whalen et al., 1993), but *relying* on F0 only when VOT is insufficient (Abramson and Lisker, 1985; Whalen et al., 1993); (iii) speakers vary considerably in how they weigh VOT and F0 as production signals to the voicing contrast, but it is always clear that VOT is more strongly weighted than F0 (Clayards, 2018); and (iv) the relationship between VOT and F0 *within* each stop voicing category is unclear (cf. Dmitrieva et al., 2015; Clayards, 2018).

3.1.4 Lexical structure and contrastive hyperarticulation

So far, this discussion has focused on the realization of VOT and F0 as signals to voicing contrasts. The primary purpose of the present investigation, however, pertains to how these cues are realized in the context of contrastive hyperarticulation. In this section we review evidence that a variety of phonetic cues are contrastively hyprarticulated, and whether or how F0 may be contrastively hyperarticulated as a cue to word-initial stop voicing contrasts.

As was noted in Chapter 1 (see also section 2.1.1), it has been argued that the realization of many phonetic cues varies systematically as a function of lexical structure (e.g., Kirov and Wilson, 2012; Fox et al., 2015; Fricke et al., 2016; Wedel et al., 2018). These effects are typically explained in terms of lexical competition, where words with sufficiently similar phonological or phonetic structure compete against one another for activation (Dell, 1986; Luce and Pisoni, 1998; see Baese-Berk and Goldrick, 2009 and Fricke, 2013, chapters 2 & 6, for relevant discussion). Though different theories vary in the details, this lexical competition is generally assumed to affect phonetic realizations, for example due to effects on phonological planning (Fricke, 2013). Lexical competition of this sort can be represented broadly, in terms of phonological neighborhood density (e.g., Wright, 2004; Fox et al., 2015; see also Jaeger et al., 2016; Buz and Jaeger, 2016, for reviews), but often it is represented more narrowly, in terms of the existence of a particular lexical minimal pair competitor (e.g., Baese-Berk and Goldrick, 2009; Schertz, 2013; Seyfarth et al., 2016; Wedel et al., 2018).

The latter case represents an instance of what Ohala (1994) refers to as the contrastive hypothesis. In order to avoid confusion with the Contrastive Hypothesis of Clark (1980) and others, which deals with child language development, and the Contrastive Analysis Hypothesis of Wardhaugh (1970) and others, which deals with second language pedagogy, we will prefer the term "contrastive hyperarticulation hypothesis". The contrastive hyperarticulation hypothesis holds that the phonetic cues which signal a particular phonological contrast will be enhanced in words for which that phonological contrast defines a lexical minimal pair, relative to words for which it does not[3]. In the case of word-initial stops in spontaneous English, the focus of this dissertation, the contrastive hyperarticulation hypothesis has already been demonstrated for VOT (Chapter 2; see also Wedel et al., 2018; but see Fricke, 2013 for different results), the primary cue to the word-initial stop voicing

[3]It should be noted that this hypothesis is closely related to Hall's (2007) Contrastivist Hypothesis, whereby phonemic inventories are defined in terms of minimal sets of contrastive phonological features (see also Hall, 2011). Presumably, the same underlying phenomena might lead to contrastive hyperarticulation of phonetic cues to phonemic contrasts as those that might lead to phonemic inventories that are defined in terms of minimally contrastive sets of phonological features or phonetic cues. We set these questions aside for future research.

contrast (Lisker, 1986). That is, VOT is contrastively enhanced in words with word-initial stop voicing minimal pair competitors, such that voiceless stops in words with a voiced stop minimal pair (e.g., *time*, minimal pair *dime*) have longer VOTs than voiceless stops in words without a voiced stop minimal pair (e.g., *type*, no minimal pair **dipe*). Similarly, voiced stops in words with a voiceless stop minimal pair (e.g., *grime*, minimal pair *crime*) have shorter VOTs than voiced stops in words without a voiceless stop minimal pair (e.g., *gripe*, no minimal pair **cripe*). Indeed, in Chapter 2 it was argued that operationalizing lexical competition in terms of the existence of cue-specific minimal pair competitor lemmas was a better predictor of VOT realizations for both voiced and voiceless stops than a wide variety of plausible alternatives (such as neighborhood density or the total number of neighbors differing in the onset position. For alternative views and additional discussion, see Goldrick et al., 2010; Fricke, 2013; Fox et al., 2015, among others).

It remains an open question, however, to what extent F0 may also be contrastively hyperarticulated as a cue to stop voicing, if at all. As discussed above, the effect of preceding stop voicing on F0 may have phonetic/physiological origins, but there is reason to believe that this contrast has become phonologized in English, suggesting that speakers may have some control of F0 following stops independently of any physiological effect related to the suppression of voicing. As such, F0 is theoretically available as a cue for contrastive hyperarticulation of stop voicing contrasts. Indeed, speakers can and do enhance multiple cues to a given contrast, even ones which reside on a different segment than the one being contrasted. In a study of word-final voicing contrasts, Seyfarth et al. (2016) showed that speakers will contrastively enhance both vowel length and the perseveration of voicing into a coda fricative. They found that words ending in voiced fricatives with word-final voiceless fricative minimal pair competitors were realized with longer vowels and with longer perseveration of voicing into the final fricative, relative to words without such competitors. Similarly, in words with final voiceless fricatives, they found that vowels were shorter and final fricatives were less voiced in words with a final voiced fricative minimal pair competitor, relative to words without such competitors. Not only does this show that multiple cues can be contrastively enhanced, but it also shows that the cues which are enhanced can reside on segments neighboring the one that is being contrasted.

This suggests that, in addition to contrastive enhancement of VOT, speakers may also contrastively enhance F0 on the following segment. However, there are some qualifying factors that may be at play. First, speakers may only contrastively enhance “primary” cues to phonological contrasts. In the case of Seyfarth et al.’s (2016) study, both vowel duration and coda voicing may contribute equally to signaling coda voicing contrasts for speakers and listeners of English. Given the degree of variation in the relationship between VOT and F0 as cues to stop voicing across speakers, and that some studies have found that speakers and listeners of English may only rely on F0 to signal initial stop voicing when VOT alone is not a strong enough cue, it may be that F0 is not a sufficiently “primary” cue to be contrastively enhanced by speakers. Indeed, it may be precisely in the case of stop-voicing minimal pairs, where VOT contrasts are especially large (Chapter 2; see also Wedel et al., 2018), that we are least likely to find contrastive enhancement of F0.

3.1.5 The present study

The present study is designed to expand our understanding of the role of F0 in contrastive hyperarticulation of stop voicing in English. This study involves the analysis of F0 following word-initial voiced and voiceless stops in a conversational speech corpus (The Buckeye Corpus of Conversational English: Pitt et al., 2007). In particular, this study investigates the effect of lexical minimal pairs for word-initial stop voicing on the realization of F0 following these stops, and the relationship between F0 and VOT in these stops. This study has 2 primary aims:

1. To determine whether F0 is contrastively hyperarticulated after voiced and voiceless stops that have lexical minimal pairs for stop voicing
2. To determine whether variation in stop-offset F0 is correlated with variation in VOT, or whether F0 following stops varies independently of VOT

These aims are of significant theoretical interest. If F0 is contrastively hyperarticulated, it would suggest that speakers manipulate even secondary phonetic cues to lexical-phonological contrasts. Although it has already been shown that multiple cues to the same contrast can be enhanced together (Seyfarth et al., 2016), it is unclear to what extent this phenomenon is dependent on the relative importance of the cues in question. In the case of phonetic cues to initial stop voicing in English, listeners predominantly rely on F0 when VOT information is not good enough to determine the voicing value of a stop (Abramson and Lisker, 1985; Whalen et al., 1990). If speakers nonetheless enhance F0 alongside VOT to strengthen the contrast (consistent with the finding that F0 perturbations affect reaction times even when VOT is unambiguous; Whalen et al., 1993), this would present clear evidence that secondary phonetic cues are contrastively hyperarticulated. If, however, F0 is not contrastively hyperarticulated, this may suggest that speakers are sensitive to the phonetic cues that listeners need most to understand them, and enhance only those particular cues. In addition, if F0 variation is significantly predicted by variation in VOT, this could suggest a more automatic relationship between these two cues. In other words, it could suggest that speakers vary VOT with control, but F0 just "comes along for the ride". If, however, F0 variation is not predicted by variation in VOT, this may suggest that speakers have independent control of these two phonetic cues[4].

These aims are also of interest to theories of sound change. Previous work (Wedel et al., 2013a,b) has suggested that phonemic contrasts are less likely to be lost (i.e., phonemes are less likely to merge) if a relatively large number of lexical items contribute to that phonemic contrast (a measure often referred to as the "functional load" of a phonemic contrast; Martinet, 1952; King, 1967; Wedel et al., 2013a). A question of interest is whether the same functional load mechanisms can explain processes of shifting phonologization such as tonogenesis, whereby one cue to a phonological contrast takes over for another. As applied to word initial stop voicing in English, contrastive hyperarticulation of VOT in words with stop voicing minimal pairs supports the notion that functional load prevents

[4]Or, of course, that factors not considered in the analysis of VOT and F0 are driving additional variation.

merger through repeated small enhancements of phonetic cues to phonemic contrasts (as discussed in Wedel et al., 2018). But if only VOT is hyperarticulated in these contexts, and not F0, will F0 gradually become lost as a phonetic cue to stop voicing? Given previous hypotheses regarding the source of voicing-dependent F0 variation, it seems likely that the F0 contrast has emerged and become phonologized over time, rather than a pre-existing F0 contrast that is waning in favor of VOT (Kingston and Diehl, 1994; Hoole and Honda, 2011). In light of this, by the hypotheses underlying the concepts of functional load and contrastive hyperarticulation, we should expect that F0 contrasts will also be preserved through phonetic enhancement in words with minimal pairs.

Preview of findings

The previously described study aims were addressed in three sub-studies by investigating patterns of F0 realization in particular word tokens. The first sub-study (section 3.5) addresses aim (1) by examining word tokens from a large dataset looking for (i) higher F0 following voiceless stops that have voiced stop minimal pairs relative to words that do not have such minimal pairs; (ii) lower F0 following voiced stops that have voiceless stop minimal pairs relative to words that do not have such minimal pairs; and (iii) the temporal prolongation of higher F0 following voiceless stops, as some German speakers do (Hoole and Honda, 2011) and as is done in languages for which F0 is becoming a more prominent cue to initial stop voicing (Bang et al., 2018; Coetzee et al., 2018). Evidence of any one of these possibilities would suggest that F0 contrasts are enhanced as a function of minimal pair existence, in addition to the VOT contrast enhancement found in Chapter 2 and in Wedel et al. (2018). Under the assumption that speakers can and do manipulate multiple cues to lexical-phonological contrasts (Seyfarth et al., 2016), it was hypothesized that F0 would indeed be contrastively hyperarticulated in words with stop voicing minimal pairs. To preview the results of this investigation, no evidence was found to suggest that F0 is contrastively enhanced in words with stop voicing minimal pairs relative to words without these minimal pairs (section 3.5).

The second sub-study (section 3.6) addresses aim (2) by examining word tokens from a modified version of the dataset under investigation in Chapter 2 and in Wedel et al. (2018), for which hand-annotated VOT measurements were readily available. In this sub-study, the data were analyzed to look for a correlation between the VOT of a word-initial stop and the F0 following that stop, within the voiced and voiceless stop categories[5]. Results on this matter from prior studies have been equivocal, as described above (Dmitrieva et al., 2015; Clayards, 2018). However, under the assumption that contrastive enhancement of phonetic cues is more easily identified in more reduced spontaneous speech contexts as compared to laboratory elicitation contexts (Wedel et al., 2018), it was tentatively hypothesized that these cues may be positively correlated in this dataset. To preview the results of this investigation, no evidence was found of a correlation between VOT and F0 in either the voiced or voiceless stops (section 3.6).

In light of these results (covered in sections 3.5 and 3.6), and in light of the evidence in favor of contrastive hyperarticulation of VOT in a comparable dataset (Chapter 2), aims (1) and (2) were jointly addressed in a third, post-hoc sub-study (section 3.7). This post-hoc investigation was conducted to determine whether F0 does indeed correlate with VOT, but *for words with minimal pairs only*. Previewing the results, VOT was found to correlate significantly with F0 for tokens of voiceless stops with voiced stop minimal pairs. No correlation was found for voiceless stop tokens without such minimal pairs, and no correlation was found between VOT and F0 for any voiced stop tokens (with or without minimal pairs). This result is discussed in sections 3.7 and 3.8.

[5]Under the assumption that decades of research have not been erroneous, VOT and F0 should absolutely correlate *across* stop voicing categories. That is, as voiceless stops have both longer VOTs and higher following F0, there should be a positive cross-category correlation between these two cues (see, e.g., Clayards, 2018, and citations therein). The present investigation, however, is about whether these cues correlate *within* either voicing category, for which previous laboratory-elicited results vary (cf. Dmitrieva et al., 2015; Clayards, 2018).

3.2 Materials and Methods

3.2.1 Speech corpus

The materials for this study were taken from the Buckeye Corpus of Conversational Speech (Pitt et al., 2007). This corpus consists of interviewer-guided conversations with 40 participants carried out in Columbus, Ohio during 1999–2000. 20 participants were over the age of 40 at the time of recording, and 20 were under the age of 30. 20 participants were female, and 20 were male. Each conversation with a participant is approximately one-hour long, and participant speech has been segmented into utterances, words, phonemes, and phonetic segments. Transcripts have been further annotated for vocal modality (e.g., creaky voicing) and words have been tagged with syntactic category labels.

The data analyzed in this study consist of 39 of the 40 speakers from the corpus (speaker 35 was excluded based on previous reports of vowel segmentation errors; Gahl et al., 2012). For 24 of those 39 speakers (previously selected for prior studies; Nelson and Wedel, 2017; Wedel et al., 2018), the data include hand-annotated voice onset time measurements (see Wedel et al., 2018, for details regarding these measurements). The remaining data consist of 15 speakers that have not been annotated for voice onset time.

For each of the 39 speakers in the data, stop-initial nouns, verbs, adjectives, and adverbs of one or two syllables with initial stress were identified. Each word's stress pattern, total number of syllables, and initial phoneme were determined based on the Carnegie Melon University Pronouncing Dictionary entry for the orthographic form of the word as it appears in the Buckeye Corpus (CMU Dict: Carnegie-Mellon University, 2015). Lexical category information was based on part-of-speech annotations in the Buckeye Corpus itself.

Words were excluded from the data based on a number of criteria commonly used in the literature. Function words, contractions including functional elements (such as *kinda*), and content words that are homophonous with function words or which can serve as high frequency discourse markers (such as *please* and *cool*) were excluded from the data (Bell et al., 2009; Gahl et al., 2012; Seyfarth, 2014). Following Wedel et al. (2018), words with a stem vowel change in their morphological paradigm were excluded from the data set because the existence of a minimal pair competitor is not always consistent throughout the paradigm for a given lemma (e.g., *drink* ~ **trink*, but *drunk* ~ *trunk*). This allowed minimal pair competitor existence to be determined based on lemma forms (see section 3.2.4 below).

Individual tokens were also excluded from the data if they were produced at an utterance boundary, were immediately preceded or followed by a disfluency, or were not produced entirely with modal voice as annotated in the Buckeye Corpus, as any of these can affect pitch or make it undetectable. Tokens were further dropped if the phonetic transcription of the word in the Buckeye Corpus did not begin with a stop consonant or the voiceless affricate [t͡ʃ] (transcribed in the corpus as "ch")[6]. The result was a set of 9,372 tokens of 1,181 unique orthographic word forms across 886 unique lemmas.

3.2.2 Pitch data

Fundamental frequency (F0) information was extracted from a region of interest (ROI) for each word token in the dataset matching the above criteria. The targeted ROI for each token was set to begin with the offset of the word-initial stop consonant and end with the offset of the nucleus of that syllable[7] as annotated in the Buckeye Corpus' phonetic transcripts. The ROI for each target word thus included any liquids (/l, ɹ/) or glides (/j, w/) that were part of a complex onset as well as the syllable nucleus. For example, in the word *play* (/p^{h}lej/), pitch information was extracted for the interval spanning /lej/ (Fig. 3.1). Extracting F0 data for the entire pre-coda portion of each target syllable allowed for some statistical control of prosody- and intonation-based variability in F0 throughout the data, discussed in greater detail in section 3.2.3 below.

F0 was extracted for each ROI using Praat's `To Pitch...` function (Boersma and Weenink, 2010), which uses an autocorrelation method to detect acoustic periodicity (see Boersma, 1993, for full details). This pitch analysis was conducted with a 10ms analysis

[6]The affricate [t͡ʃ] was included because it is a common phonetic realization of /t/ when preceding /ɹ/ in the corpus.

[7]Syllabic nasals, laterals, and rhotic vowels were all considered syllable nuclei for these purposes, in addition to all monophthongal and diphthongal vowels.

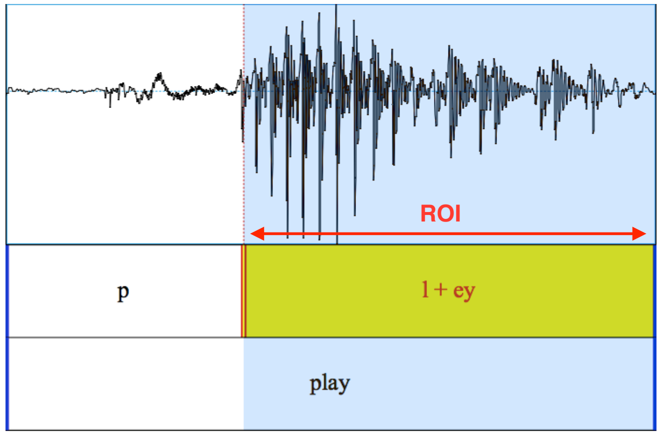

Figure 3.1: An example of the Region Of Interest (ROI) for the word *play*. F0 data was extracted from the first syllable of all target words beginning at the offset of the initial stop and ending at the onset of any coda or the end of the word. This includes any liquids or glides that make up part of the syllable onset and the full stressed vowel.

window and 5ms time steps, resulting in 10ms overlapping windows. Male speakers were analyzed using a 50-300 Hz pitch range, while female speakers were analyzed using a 100-500 Hz pitch range[8]. Due to minor errors in the placement of phoneme boundaries in the Buckeye Corpus transcripts, and the importance of capturing F0 immediately following stop offset, pitch extraction was computed beginning 25ms prior to the annotated stop offset. This ensured that the pitch extraction interval would include the pitch at stop offset for each token, but correspondingly led to attempted extraction of pitch prior to the onset of voicing for many tokens. Consequently, the ‘true’ stop offset was identified for each token as the first transition from no identifiable pitch to identifiable pitch, within +/-25ms of the annotated stop offset. If such a transition was not detectable within this 50ms window, the token was excluded from the data on the grounds that the dependent measures could not be reliably determined for such tokens. This resulted in the loss of 1,254 tokens, resulting in extracted pitch data for 8118 tokens of 1,107 word forms across 839 lemmas. The result of

[8]These relatively large pitch ranges, which can result in higher rates of pitch tracking errors, were used to accommodate significant within-gender variability across the 39 speakers in the corpus. Additional measures were taken to limit the effects of potential pitch tracking errors as outlined in this section and in section 3.2.3.

pitch extraction for each target word in the data was an arbitrary number of pitch analysis stages (depending on the length of the target syllable), each reflecting the average F0, in Hz, for 5ms of the ROI.

Although the pitch extraction algorithm used is relatively good for human speech, pitch tracking errors are not uncommon. To reduce the likelihood of such errors influencing the data, tokens for which extracted pitch values from one 5ms region to the next changed by 15% or more were excluded from the data. This led to the exclusion of a further 387 data tokens. Finally, as discussed below, the dependent measures chosen for this study required that all tokens include at least 5 pitch data points. This led to the exclusion of a further 70 tokens, resulting in 7,661 observations of 1,081 word forms across 821 lemmas.

3.2.3 Dependent measures

It has been noted that the effect of obstruent voicing on the pitch of a following vowel can sometimes be evident throughout the entire vowel, but that the effect is strongest at the onset of that vowel and tends to weaken over the course of the vowel (Hoole and Honda, 2011). As the aims of this study are principally concerned with the potential use of F0 as a means of enhancing lexical contrast, two possibilities emerge. On the one hand, speakers may directly enhance the inherent phonetic correlation between phonological voicing and following F0. As this contrast is most evident immediately following the obstruent, any contrastive enhancement will likely be most evident in these initial F0 values. Alternatively, speakers may instead enhance the contrast by maintaining the phonological effect of voicing on F0 for longer than normal, effectively reducing the rate at which F0 returns to 'normal'. We have consequently chosen two dependent measures to investigate that may contribute to the signaling of word-initial stop voicing contrasts in English:

1. Initial pitch following stop offset

2. Rate of pitch change following stop offset

We chose to represent initial pitch as the average of the first 2-3 F0 data points extracted by Praat for that token, rather than use only the first F0 data point (see section 3.2.3 below for details). This was done to make the measure more robust to possible spurious pitch measurements (i.e., due to possible pitch tracking errors). We chose to represent the rate of pitch change as change in pitch per millisecond over the range from stop offset to the midpoint of the ROI. In order to control for systematic variation of pitch that is unrelated to our hypotheses, the pitch data contributing to these dependent measures were normalized based on both speaker and token as described below.

Speaker normalization

Pitch varies substantially from individual to individual. To control for this variability, F0 values were converted from Hz to speaker-normalized semitones using the following formula (see, e.g., Dmitrieva et al., 2015):

$$\frac{12\ln(x/\bar{x})}{\ln 2} \tag{3.1}$$

Where x is the current F0 (in Hz), and $\bar{x}$ is the speaker's mean F0 (in Hz) across all target syllables in the data. This results in measures of individual pitch tokens that are normalized relative to each speaker's average F0 across all target syllables in the corpus, with positive values representing higher-than-average F0 and negative values representing lower-than-average F0.

Token normalization

Pitch also varies within a given speaker as a function of prosody and intonation. To control for some of this variation, we further normalized our dependent measures with respect to the rest of the individual token's ROI. By normalizing the pitch values of interest to the pitch in the rest of the same syllable token, we severely limit the potential impact of spurious variability in F0 on our dependent measures, such as that due to prosody or intonation (note also that all tokens come from words with initial stress, further reducing possible effects of prosody).

We achieved this token normalization for our initial pitch dependent measure in two ways. First, we calculated initial pitch as the mean pitch in semitones across the first 15ms of the ROI (i.e., across the first three pitch analysis stages); we calculated the pitch for the rest of the token as the mean pitch in semitones from *after the first 15ms* to the end of the ROI (i.e., from the fourth pitch analysis stage until the last). We chose to calculate the token's average without including the first 15ms so that pitch values corresponding to our dependent measures would not contribute to those dependent measures twice (once as the value itself, and once in the calculated token average). Note that this requires that a token's ROI span a minimum of 25ms (three 5ms pitch stages for the initial pitch, and a minimum of two more 5ms pitch stages for the rest of the token to allow for averaging). This led to the loss of 70 data tokens as noted above in section 3.2.2. We then calculated the dependent measure as the deviation in semitones from this measure of initial pitch to this measure of the token average. This is represented in the following formula:

$$\frac{\sum_{i=1}^{3}(x_i)}{3} - \frac{\sum_{j=4}^{N}(x_j)}{N} \tag{3.2}$$

Where x is pitch in semitones, N is the total number of pitch analysis stages in the ROI, and i and j each represent a given pitch analysis stage, where each pitch analysis time stage corresponds to 5ms of pitch data as described above. This results in a measure of the difference in pitch between stop offset and the rest of the ROI, such that positive values represent initial pitch values that are higher than the average pitch across the rest of the token, and negative values represent initial pitch values that are lower than the average pitch across the rest of the token.

Second, we repeated this process using the pitch at the midpoint of the ROI to represent the pitch of the rest of the token. This was done primarily to provide an alternative measure of token-normalized pitch that is more resilient to so-called "edge effects". Specifically, in addition to sentence- and phrase-level intonation, individual words often involve their own, smaller pitch contours. For example, surrounding consonants can have small but significant effects on the pitch of a vowel, particularly at the edges of the vowel (in fact, the effect of stop voicing on initial pitch is an example of just such an effect). When calculating the mean pitch for the token as we did above, more extreme pitch values at the edge may have undue influence on the mean, thus skewing the dependent measure. Including a single token-relative reference point at the middle of the syllable limits this possibility. However, using the midpoint means relying on a single pitch data point to represent the entire token, which increases the potential influence of pitch tracking errors. By comparing both of these strategies for token normalization, we improve our confidence in our dependent measures. Note that, because some of our data tokens spanned only 25ms (i.e., included only 5 pitch analysis stages), and in such cases the token midpoint is represented by the F0 data from 10-15ms (i.e., the 3rd pitch analysis stage), we calculated initial pitch for this measure as the mean pitch in semitones across the first 10ms of the ROI (2 analysis stages) rather than the first 15ms. In this way, we again prevent the same data point from contributing to the dependent measure twice (once as the value itself, and once as the midpoint representing the rest of the token). Thus, the midpoint-normalized initial pitch measure is represented by the following formula:

$$\frac{\sum_{i=1}^{2}(x_i)}{2} - x_m \tag{3.3}$$

Where x is again pitch in semitones, i is the pitch analysis time stage (each representing 5ms of data), and x_m is the pitch in semitones at the midpoint of the ROI. This results in a measure of the difference in pitch between stop offset and the midpoint of the ROI, such that positive values represent initial pitch values that are higher than the pitch in the middle of the token, and negative values represent initial pitch values that are lower than the pitch in the middle of the token.

Token normalization was less critical for our dependent measure concerning the rate of pitch change following stop offset. This measure was calculated according to the following formula:

$$\frac{x_1 - x_m}{t_m - t_1} \tag{3.4}$$

Where x_1 is the pitch, in semitones, immediately following stop offset, x_m is the pitch, in semitones, at the midpoint of the ROI, t_m is the time, in milliseconds, of the token midpoint, and t_1 is the time, in milliseconds, of the initial pitch value. Thus, the denominator represents the time that transpired between x_1 and x_m. The result is the rate of pitch change in speaker-normalized semitones per millisecond, such that positive values represent increases in speaker-normalized pitch per millisecond and negative values represent decreases in speaker-normalized pitch per millisecond.

Thus, this dependent measure inherently contains a degree of token normalization. As a representation of pitch change, it is akin to our midpoint-normalized initial pitch measure described above; as a representation of this change over time, this measure is further normalized with respect to token duration.

3.2.4 Factor of interest: Minimal pair competitor existence

Our primary factor of interest was the existence of a word-initial stop-voicing minimal pair competitor in the lexicon. Following Nelson and Wedel (2017) and Wedel et al. (2018), we built this lexicon by creating a lemmatized version of the Carnegie-Mellon University Pronouncing Dictionary (CMU Dict: Carnegie-Mellon University, 2015). Based on the lemma-to-word-form concordances included in the Corpus of Contemporary American English (COCA: Davies, 2009), we removed all inflectional variants from CMUDict, leaving only lemma forms. We further removed unfamiliar lemmas for which the orthographic form had a contextual diversity of lower than 0.5% of films in the Subtlex-US database (Brysbaert and New, 2009). The result was a lexicon of orthographic and phonemic representations for lemmas that are more likely to be familiar to the speakers in our dataset.

Using lemma forms instead of surface word forms for our lexicon prevented different surface forms of the same underlying lemma from being coded differently for minimal pair existence. For example, while the surface word form *boy* (/boj/) has no minimal pair competitor *poy* (/p^hoj/)[9], the plural word form *boys* (/bojz/) does have the minimal pair competitor *poise* (/p^hojz/). In our lemmatized lexicon, however, minimal pair competitor existence is determined for both of these surface word forms based on the lemma form *boy* (/boj/)[10].

[9] I acknowledge that, for some speakers of English, there is a word `poi`, which comes from Hawaiian and refers to fermented taro root. I consider this word sufficiently unfamiliar to the majority of English speakers, however.

[10] For evidence that using lemma versus surface word forms does not significantly affect results concerning contrastive hyperarticulation of VOT for stop voicing minimal pairs in a comparable dataset, see Wedel et al. (2018).

Minimal pair competitor existence was determined for each target word in the dataset based on this lemmatized lexicon. For each token, we replaced the initial stop of the phonemic form of that word's lemma with the corresponding voicing-distinct stop. If this new phonemic form was found in our lemmatized lexicon, the token was coded as `true` for minimal pair competitor existence. Otherwise it was coded as `false`. If a phonemic form had multiple orthographic correspondents, the minimal pair competitor factor was coded as `true` if any single correspondent was found in the lexicon.

As an example, for the inflected word form *pats*, we substituted the corresponding voicing-distinct phoneme /b/ into the phonemic representation of the lemma *pat* /pæt/ to produce /bæt/. This phonemic form corresponds to the existing word *bat* and so the value of the minimal pair competitor factor is coded `true` for all tokens of the word *pats*. The number of tokens and lemmas in the dataset beginning with each stop, as a function of minimal pair lemma existence, is given in Table 3.1.

It should be pointed out that the Buckeye Corpus files include phonemic annotations in addition to their phonetic transcriptions. We elected to use the phonemic forms of words found in CMUDict instead of those found in the Buckeye Corpus itself because (i) they are more complete than the Buckeye's phonemic annotations (which use a single phonemic form for each orthographic word, ignoring homographs), (ii) they are widely used in the literature (including metastudies, e.g., Tucker et al., 2019), and (iii) we based our lexicon on lemma forms, not the surface word forms that the phonemic annotations in the Buckeye Corpus are based on.

3.3 Preliminary Data Visualization

While the effect of obstruent voicing on the pitch of following vowels is fairly well established for English in laboratory data (e.g., Hanson, 2009), there is little direct evidence of this phenomenon for English in more spontaneous speech such as that found in the Buckeye Corpus. In addition, the relative weights of cues to stop voicing, including F0, vary across speakers (e.g., Dmitrieva et al., 2015; Clayards, 2018). For these reasons, data were

Stop phoneme	Number of tokens	Number of lemmas	Avg. tokens-per-lemma
/b/	616	126	4.889
/d/	548	88	6.227
/g/	468	52	9.000
/p/	1611	127	12.685
/t/	752	98	7.673
/k/	1600	187	8.556

(a) Number of tokens and lemmas per stop, for data without minimal pairs

Stop phoneme	Number of tokens	Number of lemmas	Avg. tokens-per-lemma
/b/	665	40	16.625
/d/	194	16	12.125
/g/	221	17	13.000
/p/	300	32	9.375
/t/	413	21	19.667
/k/	273	17	16.059

(b) Number of tokens and lemmas per stop, for data with minimal pairs

Table 3.1: Counts of tokens and lemmas per stop phoneme. Subtable (a) shows totals for observations corresponding to lemmas that do not have a word-initial stop voicing minimal pair competitor lemma, while subtable (b) shows totals for observations corresponding to lemmas that do have a word-initial stop voicing minimal pair competitor lemma.

first visualized looking for the following patterns:

1. Does stop voicing systematically affect following F0 in this dataset?
2. Are speakers consistent in the effect of phonological stop voicing on following F0?
3. Does the existence of a stop voicing minimal pair affect the realization of following F0?

3.3.1 Does stop voicing systematically affect our dependent measures?

The effect of the phonological voicing value of a stop on the realization of speaker-normalized pitch (semitones) over the course of the first 50ms of the syllable, aggregated across speakers and lemmas, is plotted in Figure 3.2. Visual inspections suggests an effect of phonological voicing on the realization of pitch, with voiceless stops followed by higher average pitch than voiced stops. This effect appears to continue (although it diminishes over time) throughout the aggregated data for the first 50ms of our target syllables[11]. Note, however, that this data is not token-normalized as described in section 3.2.3, and no controls for other variables affecting pitch have been accounted for. As such, error bars were found to overlap considerably and overcrowd the plot, and so they have not been included (for a plot that controls for much of this variability and includes standard error, see Figure 3.3).

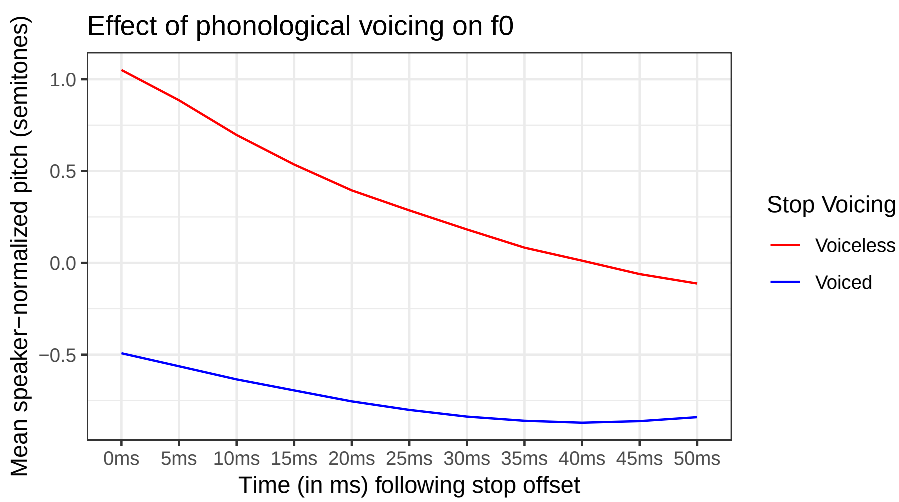

Figure 3.2: The effect of phonological stop voicing on following pitch in the dataset. The y-axis is pitch in speaker-normalized semitones. The x-axis represents the time (in ms) elapsed since the offset of the stop. The red lines represent pitch values following voiceless stops. The blue lines represent pitch values following voiced stops. Error bars are not included because they are too large.

To see whether the over-large error bars are due to systematic by-speaker and by-item variability or are instead indicative of a spurious effect, we fit simple mixed effects models to the voiced and voiceless stops separately, with random effects included for speaker and lemma. Figure 3.3 plots the model estimates together, representing the overall effect of stop voicing on following pitch across speakers and lemmas with random intercepts accounting for systematic by-speaker and by-lemma variability. Even for such simple models, where

[11]Note that not all ROIs in our dataset span a full 50ms, so the concentration of data contributing to these lines is necessarily decreasing as time increases.

a large amount of potentially systematic variation is treated as random noise, pitch appears to vary significantly as a function of stop voicing.

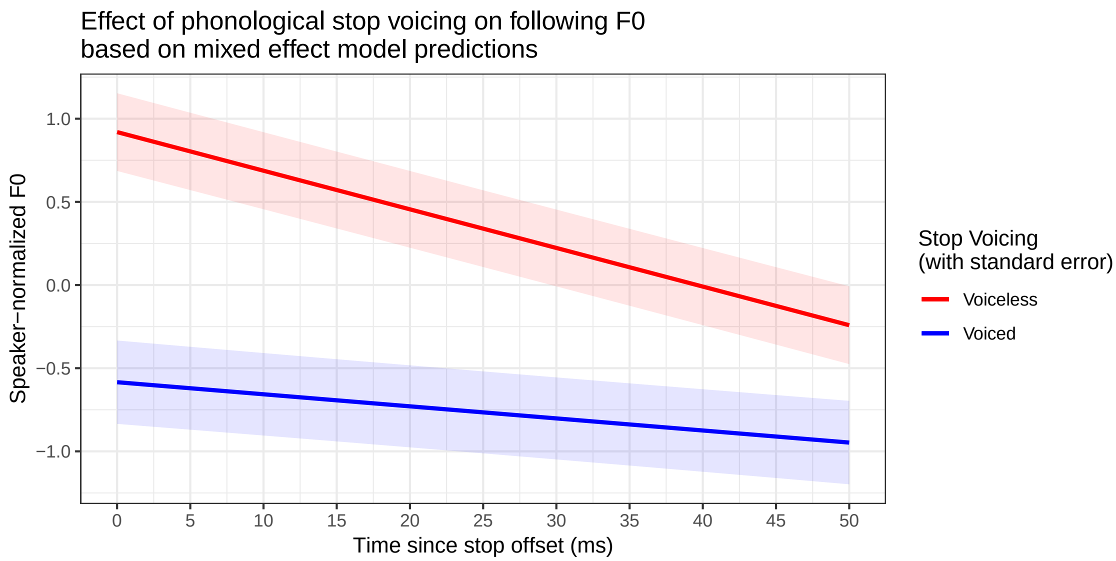

Figure 3.3: The effect of phonological stop voicing on following pitch as represented by simple mixed effects models of Pitch ~ Time for voiced and voiceless stops with random effects for speaker and lemma. The y-axis represents pitch in speaker-normalized semitones. The x-axis represents the time (in ms) elapsed since the offset of the stop. The red line represents pitch following voiceless stops. The blue line represents pitch following voiced stops. Shaded areas represent standard error values according to model estimates.

3.3.2 Are speakers consistent?

The same data as from Figure 3.2 are plotted separately for each speaker in Figure 3.4. For the majority of speakers, there is a clear voicing effect present, whereby average pitch values following voiceless stops appear to be consistently higher than those following voiced stops (albeit with a number of exceptions, most notably speakers 2 and 28). Furthermore, this difference diminishes, but is nonetheless maintained, throughout those 50ms into the target syllable for the majority of speakers (though again with a number of exceptions, e.g., speakers 31 and 36).

Figure 3.5 presents the same data tokens, but using a visually modified form of the

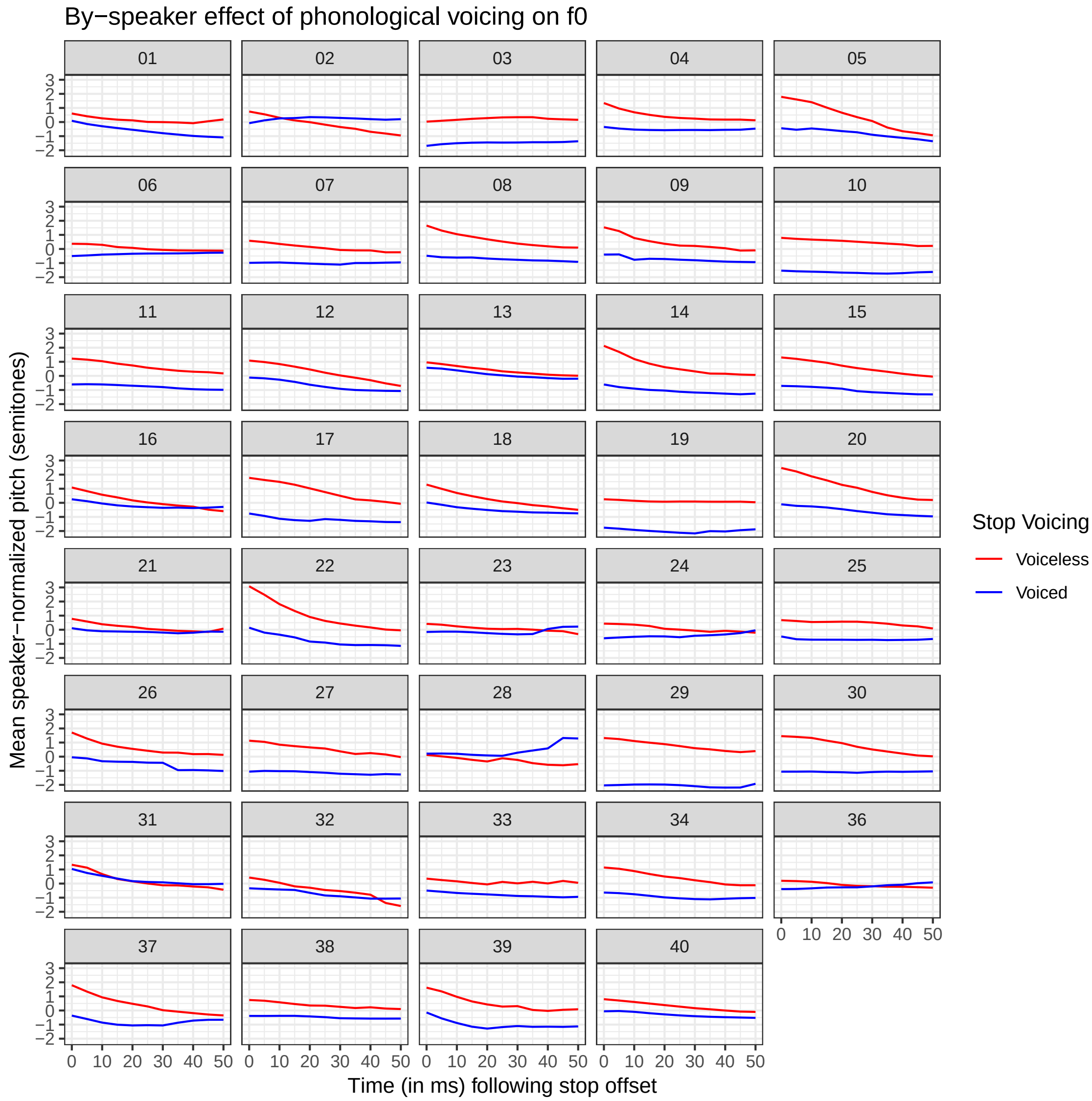

Figure 3.4: The effect of phonological stop voicing on following pitch for each speaker in the dataset. Speakers are identified according to their IDs in the Buckeye Corpus. The y-axis is pitch in speaker-normalized semitones. The x-axis represents the time (in ms) elapsed since the offset of the stop. The red lines represent pitch values following voiceless stops. The blue lines represent pitch values following voiced stops. Error bars are not included because they are too large.

average-normalized initial pitch measure described in section 3.2.3. Note that, where the average-normalized initial pitch measure averages across the first three pitch analysis stages for each token prior to subtracting the rest of the token's average, the visualization in Figure

3.5 plots the difference between each of those first three pitch analysis stages and the token average independently to improve visualization. As can be seen, a portion of the cross-speaker variability from Figure 3.4 is removed when pitch is normalized by token (note especially speakers 2, 28, 31, and 36), but considerable variability remains nonetheless (note especially speakers 1, 13, and 19).

3.3.3 Are there signs of a minimal pair effect?

Figure 3.6 shows the aggregated effect of word-initial stop voicing minimal pair existence on the realization of pitch for voiced and voiceless stops separately. Though visually the data for lemmas with minimal pairs appear to have higher speaker-normalized pitch values, the differences are likely too small to be reliable.

To help determine how consistent this pattern may be across items, Figure 3.7 represents the same data as above, plotted separately by place of articulation (labial, coronal, and velar). There appears to be an effect of place of articulation on the data, such that the apparent effect of higher pitch for lemmas with initial stop voicing minimal pairs holds for labial and coronal stops, but not for velar stops. However, the numbers of observations per category are, at this point, small enough to suggest that these apparent visual effects may not be statistically reliable (Table 3.1).

3.3.4 Summary of data visualization

Visualization of the data suggests that the phonological effect of initial stop voicing on following F0 is found in this conversational speech corpus, and is relatively consistent across speakers insofar as nearly every speaker produces voiceless stops with relatively higher offset pitch than voiced stops. However, this visualization also suggests that there likely is not an effect of initial stop minimal pair competitor existence on the realization of pitch. Furthermore, if there is such an effect, it does not appear to be contrastive.

It is important to note, however, that the realization of pitch in English is influenced by a plethora of variables unrelated to the hypotheses under study here. As such, it is important to apply rigorous statistical controls before making conclusions regarding the data. We

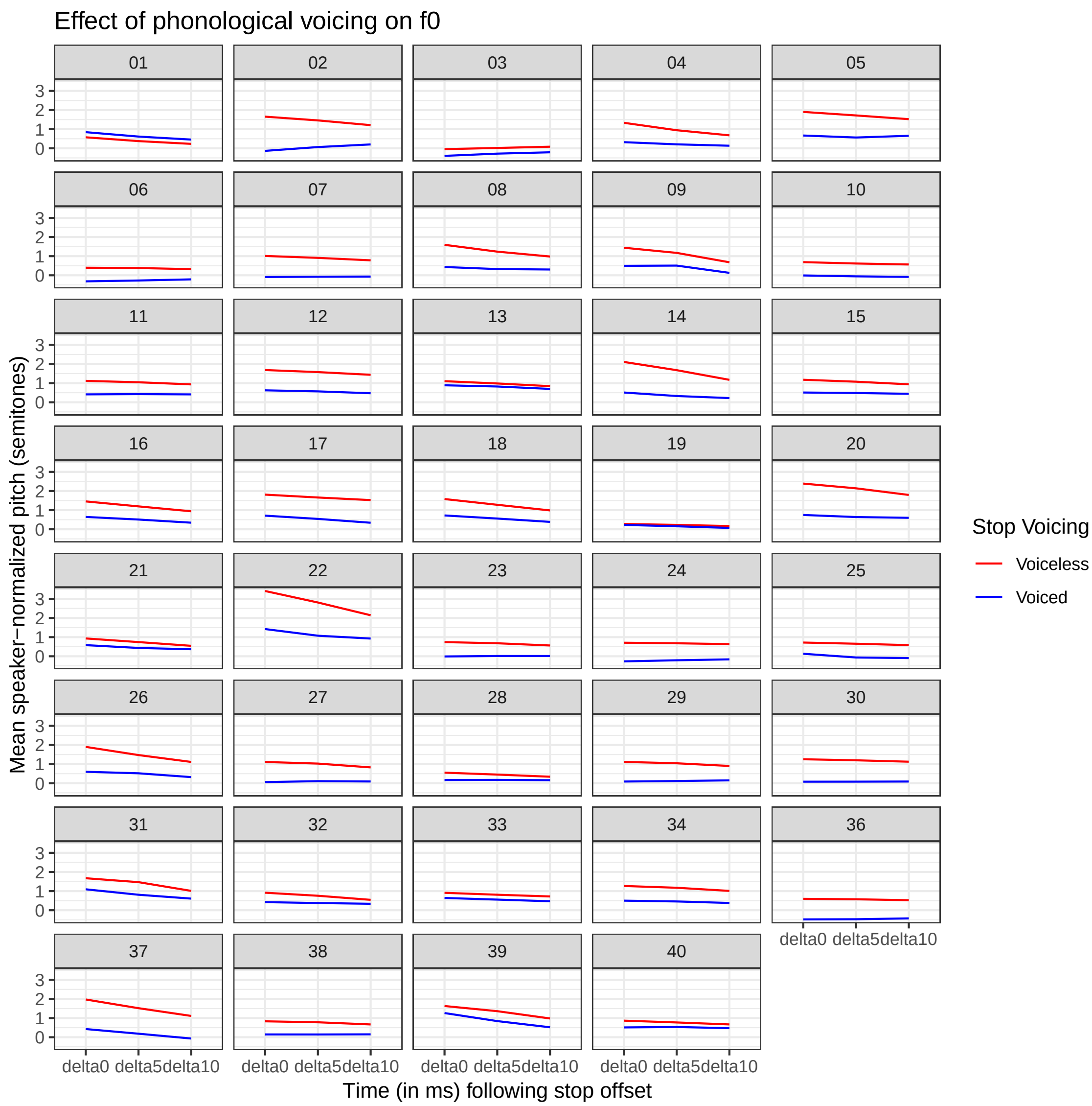

Figure 3.5: The effect of phonological stop voicing on following pitch for each speaker in the dataset, token-normalized. Speakers are identified according to their IDs in the Buckeye Corpus. The y-axis is the deviation of initial pitch from the token's average pitch, in speaker-normalized semitones. The x-axis represents the time (in ms) elapsed since the offset of the stop. The red lines represent pitch values following voiceless stops. The blue lines represent pitch values following voiced stops.

now turn to a description of the predictor variables and model fitting procedure used to statistically control the data. We will then turn to tests of each of our primary hypotheses.

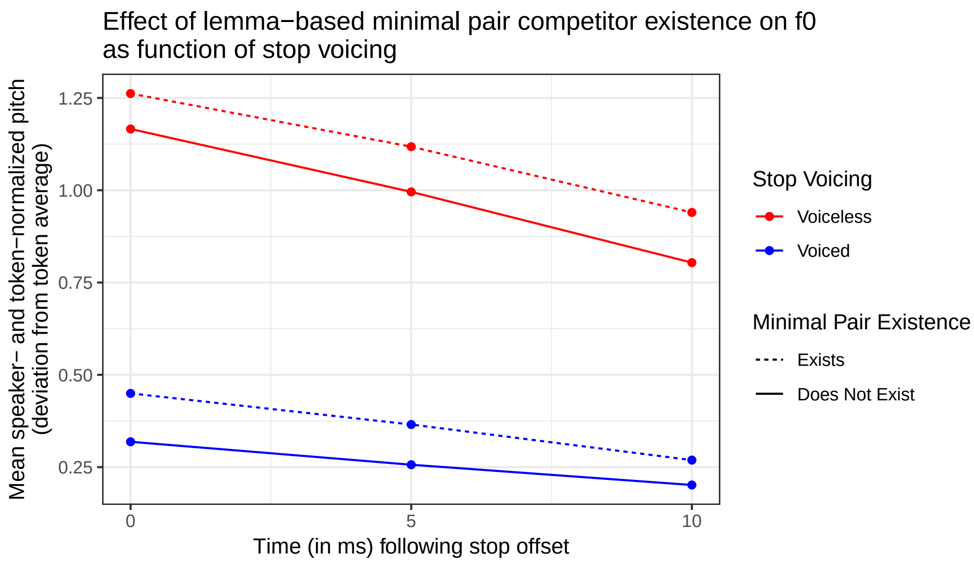

Figure 3.6: The effect of the existence of an initial stop voicing minimal pair competitor by phonological stop voicing on following pitch, token-normalized. The y-axis is the deviation of stop offset pitch from the average pitch of the rest of the syllable, in speaker-normalized semitones. The x-axis represents the time (in ms) elapsed since the offset of the stop. The red lines (top of plot) represent voiceless stops. The blue lines (bottom of plot) represent voiced stops. Solid lines represent tokens of lemmas that do not have an initial stop voicing minimal pair competitor in the lexicon. Dotted lines represent tokens of lemmas that do have an initial stop voicing minimal pair competitor.

3.4 Statistical Analysis Methods

Statistical analysis was conducted by fitting separate linear mixed effects models for voiced and voiceless stops including a variety of independent control predictors expected to influence F0 either directly or indirectly. We chose to fit separate models to the voiced and voiceless stop data instead of including voicing as a covariate in a single model per dependent measure (see Chapter 2 and Wedel et al., 2018 for the same approach). This was done because we have different predictions about how contrastive hyperarticulation might manifest for voiced versus voiceless stops, with contrastive F0 realizations expected to be higher following voiceless stops but lower following voiced stops. Separate models thus allow us to more clearly test the simple effects of initial stop voicing minimal pair lemma existence on the realization of following F0 for each stop type independently.

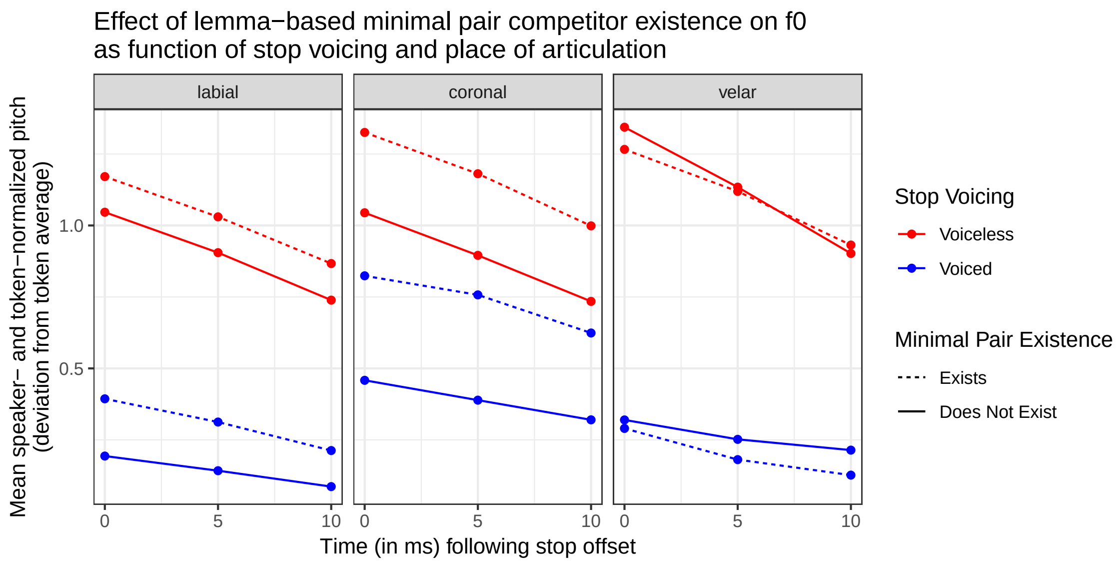

Figure 3.7: The effect of the existence of an initial stop voicing minimal pair competitor by phonological stop voicing and place of articulation on following pitch, token-normalized. The y-axis is the deviation of stop offset pitch from the average pitch of the rest of the syllable, in speaker-normalized semitones. The x-axis represents the time (in ms) elapsed since the offset of the stop. Separate panels of the plot represent data for labial, coronal, and velar stops. The red lines (top of plot panels) represent voiceless stops. The blue lines (bottom of plot panels) represent voiced stops. Solid lines represent tokens of lemmas that do not have an initial stop voicing minimal pair competitor in the lexicon. Dotted lines represent tokens of lemmas that do have an initial stop voicing minimal pair competitor.

We first describe the continuous control predictors used during model fitting, followed by categorical control predictors, then interaction terms among control predictors, and finally the model fitting procedure.

3.4.1 Continuous control predictors

The following continuous measures were included as candidate predictors during model fitting. All continuous variables were centered to assist in the interpretation of model intercepts and scaled to match the relevant dependent variable in order to facilitate model convergence.

Word familiarity

The contextual diversity of the orthographic word form of each token, represented as the percent of films in which the word appears in the SUBTLEX-US database. Lexical frequency is implicated in the realization of prosodic prominence in English (such as focus), which leads to relatively higher F0 within the word and/or relatively lower F0 on the preceding and following words (Breen et al., 2010; Cooper et al., 1985, as cited in Chuoying Ouyang and Kaiser 2014). Word familiarity as measured by contextual diversity has been found to be a better predictor of several behavioral measures than frequency in both the visual (Brysbaert and New, 2009) and auditory domains (Geary, 2019).

Forward and backward contextual probability

Forward and backward conditional bigram probabilities, log-transformed. Like word familiarity, contextual probability is implicated in the realization of prosodic prominence, and therefore F0 (Chuoying Ouyang and Kaiser, 2014). Conditional bigram probabilities were calculated per two-word pair using a combined corpus consisting of both the Buckeye Corpus (Pitt et al., 2007) and the Fisher Part 2 transcript corpus (Cieri et al., 2005), a set of 5849 English telephone conversations containing over 12 million words. A combined corpus was used in order to (i) ensure more realistic bigram probability estimates based on a larger sample of American English conversational speech than the Buckeye Corpus alone provides (note that the Buckeye Corpus contains only about 300,000 words. For use of the Fisher corpus in this way, see, e.g., Arnon and Snider, 2010; Seyfarth, 2014; Wedel et al., 2018), and (ii) ensure that every bigram in the Buckeye Corpus appears at least once in the combined corpus, thus avoiding conditional bigram probabilities of zero (note, however, that this approach likely over-estimates conditional probabilities for infrequent bigrams from the Buckeye Corpus).

Phonotactic probability

The average biphoneme sequence probability of the phonological form of the lemma measured without reference to stress, available through the IPhOD2 database, log-transformed. This measure calculates the average probability of each two-phoneme sequence in the lemma, specific to its position as measured from the start of the lemma (Vitevitch and Luce, 2004). This measure is correlated with general phonetic reduction in speech production (Vitevitch and Chu, 2004), which may affect prominence or pitch.

Speech rate

The number of syllables per second in the phonemic transcription of the stretch of continuous speech surrounding the token. This was calculated as the number of vowels in the phonemic transcript for the region of continuous speech within which the token is found, divided by the number of seconds in that region of continuous speech (Bell et al., 2003). The boundaries for regions of continuous speech were based on the Buckeye transcripts, and included such things as silences, pauses or disfluencies, interviewer speech intervals, and the beginnings and ends of audio files. Although speech rate has not been shown to independently affect stop offset pitch, it has been shown to affect voice onset time for voiceless stops (Yao, 2007; Wedel et al., 2018), which is purportedly correlated with following F0 (Shultz et al., 2012; Dmitrieva et al., 2015; Clayards, 2018, among others).

3.4.2 Categorical control predictors

The following categorical variables were also included as candidate predictors during model fitting. Categorical variables related to the phonological structure of a word were based on the phonemic representations of the word (not the lemma) as represented in CMUDict (i.e., they are not based on the phonemic or phonetic representations from the Buckeye transcripts). All categorical variables were sum-coded with reference to the grand mean to facilitate interpretation of model intercepts.

Place of articulation

The place of articulation for the word-initial stop consonant (`labial`, `coronal`, or `velar`).

Previous mention

Whether or not the lemma had already appeared in the speaker's prior discourse at the time this token was produced (`true` or `false`). Approximately 53% of the tokens in the dataset occurred after a prior use of the same lemma in a speaker's own speech (4,073/7,661 tokens). Repeated use in discourse is correlated with reduced prominence and pitch (Terken, 1984).

High vowel

Whether or not the vowel comprising the syllable nucleus is a high vowel (`true` or `false`). Vowel height is significantly correlated with F0 such that high vowels are typically produced with higher pitch as compared to low or mid vowels (a phenomenon sometimes referred to as "intrinsic pitch"; Hoole and Honda, 2011). The vowels /i, ɪ, u, ʊ/ were coded `true` as high vowels. All other vowels, including diphthongs and the syllabic rhotic /ɝ/, were coded `false`.

Diphthong

Whether or not the syllable nucleus is a diphthongal vowel (`true` or `false`). Diphthongs involve formant contours and may also include more dynamic pitch contours, which may affect our dependent measures (especially the slope of pitch change). The vowels /ej, aj, aw, ɔj, ow/ were coded as `true` as diphthongal vowels, and all other vowels were coded as `false` (note that glides were not treated as part of the vowel, so sequences such as /wɔ, wɛ, ju/ etc. are treated as having the vowels /ɔ, ɛ, u/ etc. See "Following Glide" below).

Following liquid

Whether or not the segment following the initial stop is one of /l, ɹ, ɝ/ (`true` or `false`). While there is no evidence that the presence of a liquid is directly correlated with F0, the presence of a following liquid significantly affects VOT realizations (Nelson and Wedel, 2017; Wedel et al., 2018), which may be correlated with F0 in English. Furthermore, there is a dearth of data regarding the effect of obstruent voicing on following non-vocalic segments. Along with following glide, the following liquid factor allows us to see whether voicing dependent F0 variation is reflected on non-vocalic segments. This combination of factors also allows for the possibility that any effect of vowel height on our dependent measures is only detectable in words with simplex onsets.

Following glide

Whether or not the segment following the initial stop is one of /j, w/ (`true` or `false`). Along with following liquid, this factor allows us to see whether voicing dependent F0 variation is reflected on non-vocalic segments and allows for the possibility that intervening consonants disrupt any potential effect of vowel height on our dependent measures.

Number of syllables

The number of syllables in the phonemic transcription of the word (`1` or `2`). Since all of our word tokens have stress in the first syllable, and pitch is a major correlate of lexical stress (higher pitch is generally perceived as stressed Fry, 1958), the number of syllables was included to control for the possibility that the higher pitch of a word token's initial syllable may decline more rapidly in 1-syllable words than in 2-syllable words, which could potentially impact any of our dependent measures.

Syntactic category

The syntactic category of the word-token (`noun`, `verb`, `adjective`, or `adverb`) as reported in the Buckeye Corpus. Words of certain syntactic categories are more likely to receive focus prominence than others (Calhoun, 2007), and pitch is a key correlate of focus in English (Chuoying Ouyang and Kaiser, 2014).

Speaker gender

The gender of the speaker as reported in the Buckeye Corpus (`female` or `male`). Though our dependent measures are speaker-normalized, they represent changes in pitch between the stop offset and the rest of the syllable nucleus. Females tend to have larger pitch ranges than men, which may affect these dependent measures (though these differences are smaller when pitch is speaker-normalized; see Simpson, 2009 for examples and discussion).

Speaker age

The relative age of the speaker as reported in the Buckeye Corpus (`above 40 years old` or `below 30 years old`). Older speakers tend to have lower pitch than younger speakers (e.g., Baker et al., 2001), which may result in less variability in pitch overall.

3.4.3 Interaction terms among control predictors

The following interactions were also included as candidate control predictors during model fitting.

Speaker gender and speaker age

In a large normative study, Goy et al. (2013) found that effects of age on vocal pitch were different for male and female speakers. In light of this possibility, and to allow for the possibility that gender-dependent pitch effects are smaller among older speakers than younger speakers (e.g., Bang et al., 2018), we included an interaction term between speaker gender and speaker age.

Vowel height and following liquid; vowel height and following glide

To account for the possibility that effects of vowel height on pitch are influenced by intervening segments, both following liquid and following glide were allowed to interact with vowel height separately. In the voiced stop data, there were only three observations of two lemmas which included a following glide (two observations of the lemma *beauty* and one observation of the lemma *dwarf*), so the interaction between vowel height and following glide was not included for voiced stops.

Word familiarity and conditional probabilities

Chuoying Ouyang and Kaiser (2014) showed that pitch modulations associated with focus are partially the result of both lexical frequency (a form of word familiarity) and contextual probability (of which our conditional probability measures are examples). They further showed an interaction between these terms, such that pitch effects associated with new-information focus were only found for high frequency words in low probability contexts, whereas pitch effects associated with corrective focus were only found for low frequency words in high probability contexts. For these reasons, we included a three-way interaction between word familiarity, forward conditional bigram probability, and backward conditional bigram probability.

3.4.4 Backward step-wise model fitting

Model fitting was conducted using the lmerTest package (Kuznetsova et al., 2017) in R (R Core Team, 2018). Separate models were fit for voiced and voiceless stops, as well as for each of our three dependent measures (average-normalized initial pitch, midpoint-normalized initial pitch, and pitch slope), resulting in 6 fitted models. Model fitting involved defining a maximal model including fixed effects for all control predictors described above as well as the factor of interest (lemma-based word initial stop voicing minimal pair existence). Speaker and lemma were included as random intercepts, with the factor of interest included as a correlated random slope on the speaker intercept, to help account for the possibility that the effect of minimal pair competitor existence varies across speakers (and following the advice of Barr et al. 2013). Using the lmerTest package's `step()` function, backwards step-wise selection of control predictors was conducted. In this process, predictors that do not significantly improve model fit are removed one-by-one until no more predictors can be removed without significantly affecting model fit. Interaction terms must be removed before their simple effect terms are considered for elimination. Neither any random effects terms nor the factor of interest were considered for elimination[12], unless explicitly stated otherwise.

It should be noted that this modeling procedure involves numerous implicit "hypothesis tests" for each model. Procedures involving multiple tests have an increased rate of Type I error, that is, an increased chance of falsely rejecting the null hypothesis in favor of the alternative hypothesis. It is common in such cases to correct for these inflated Type I error rates by adjusting the α criterion for statistical significance. However, doing so considerably inflates the likelihood of Type II error, that is, the likelihood of failing to reject the null hypothesis when the alternative hypothesis is in fact true. Given that any effect of minimal pair lemma existence on the realization of post-stop F0 is expected to be relatively small, it was considered more important that Type II error be avoided than Type I error. As such, no attempt was made to adjust the α criterion or otherwise reduce the chances of Type I error at any stage in the modeling procedure. Instead, the reader is reminded that multiple tests will increase the likelihood that some effects appear significant by chance alone, and that all results should be considered alongside results from more controlled laboratory experiments[13].

3.5 Does F0 Vary as a Function of Minimal Pair Existence?

As described in section 3.1.5, one of the principle hypotheses under investigation in this study is whether F0 is contrastively enhanced in words with initial stop voicing minimal pairs relative to words without such minimal pairs. Previous work has already shown that VOT is contrastively enhanced as a function of stop voicing minimal pair existence in a

[12]This was achieved by setting parameters within the `step()` function: `reduce.random = FALSE`, preventing the elimination of any random effect terms, and `keep = c(''Factor_Of_Interest'')`, preventing the factor of interest from being eliminated from the fixed effect terms. The actual factor of interest varies between the analyses reported in sections 3.5, 3.6, and 3.7. For details of how these parameters are interpreted by the `step()` function, see Kuznetsova et al. (2017).

[13]Given that most of the significant effects found in the models are reasonable, explainable, and related to control predictors only, risk of Type I error was not deemed a serious concern.

comparable data set from the same corpus (Chapter 2; see also Wedel et al., 2018). We ask here whether F0 is also contrastively hyperarticulated.

We report the results of our model fitting procedure below, beginning with our average-normalized measure of initial pitch, then turning to our midpoint-normalized measure of initial pitch, and finally turning to our measure of pitch slope. In each case, we first report the model results for voiceless stops, then voiced stops.

3.5.1 Average-normalized initial pitch

Our average-normalized measure of initial pitch following stop offset compares pitch immediately following the word-initial stop to the pitch in the rest of the ROI. As such, we expect that the model intercept should be positive for both stop types, but more so for voiceless stops than voiced stops. That is, we expect pitch to be higher at stop offset and decline over time for both voiced and voiceless stops, but that pitch at stop offset will be higher among voiceless stops than it is among voiced stops (as predicted in Fig. 3.3).

Our principle hypothesis, however, is regarding the existence of a word-initial stop voicing minimal pair. If F0 is contrastively enhanced in words with stop voicing minimal pairs relative to words without such minimal pairs, we would expect to find evidence of this fact in the form of a significant positive correlation between the factor of interest (word-initial stop voicing minimal pair competitor lemma existence) and our dependent measure (average-normalized initial pitch) in the voiceless stop data, and/or in the form of a significant negative correlation between the factor of interest (word-initial stop voicing minimal pair competitor lemma existence) and our dependent measure (average-normalized initial pitch) in the voiced stop data.

Voiceless stops

The fitted model for voiceless stops predicting our average-normalized initial pitch measure is presented in Table 3.2. The factor of interest, word-initial stop voicing minimal pair competitor existence, is bolded for ease of reference.

Variable	β	SE	t	$p(\|t\|)$
(Intercept)	1.092	0.082	13.248	< 0.001***
Gender (female-male)	0.204	0.076	2.694	< 0.05*
Following Liquid (true-false)	0.070	0.027	2.649	< 0.01**
Number of Syllables (1-2)	0.075	0.022	3.424	< 0.001***
Forward Probability	-0.003	0.007	-0.403	0.687
Backward Probability	-0.018	0.007	-2.520	< 0.05*
ForProb : BackProb	-0.005	0.002	-2.301	< 0.05*
Min Pair Exist (true-false)	**0.011**	**0.034**	**0.321**	**0.749**

(a) Fixed effects summary

Group	Variable	Variance	SD	Corr
Lemma	(Intercept)	0.043	0.206	
Speaker	(Intercept)	0.211	0.459	
	Min Pair Exist (true-false)	$<$ **0.001**	**0.031**	**-0.04**
Residual		1.152	1.073	

(b) Random effects summary

Table 3.2: Fixed and random effects summaries for the fitted model predicting the difference between initial pitch following stop offset and the average pitch of the rest of the token (in semitones) for voiceless stops.

The fitted model intercept tells us that, on average across speakers and items, pitch at stop offset is approximately one full semitone higher than it is in the rest of the syllable, consistent with a general pattern of raised pitch following voiceless stops (Hoole and Honda, 2011) and/or falling pitch contours across words. A significant gender effect tells us that, despite speaker-normalized pitch values, female speakers tend to exhibit a greater difference in pitch between stop offset and the rest of the syllable than male speakers, consistent with some prior reports that, on average, female speakers have more dynamic pitch contours than men. We also find significant effects based on syllable structure. One-syllable words show a greater difference in pitch at stop offset relative to the average pitch of the token as compared to two-syllable words; this may be indicative of falling pitch contours across the entire duration of a word, which presumably happen more rapidly in shorter words than in longer words. In addition, words with a liquid in the syllable onset are realized with higher pitch at stop offset (i.e., during phonation of the liquid) relative to the token-average pitch as compared to words without a liquid in the syllable onset.

The effect of backward conditional probability tells us that, if the target word is relatively probable given the previous word, onset pitch is slightly but consistently lower relative to the rest of the syllable. We also see a significant effect of the interaction between forward and backward conditional probability, such that the effect of backward conditional probability appears to be stronger when the target word is also relatively probable given the following word.

Of note, our factor of interest, word-initial stop voicing minimal pair lemma existence, did not significantly impact model fit. In addition, the height of the vowel did not significantly impact model fit, possibly because comparing the onset pitch to the token's average washes out any effect of vowel height. It is also worth noting that the largest fixed effect in terms of the estimate coefficients is the intercept, suggesting that none of our model predictors has a larger impact on our dependent measure than the inherent effects of lexical pitch and, most likely, the phonological voicing status of the stop.

Within the random effects structure, we see that the correlated slope for our factor of interest accounts for very little model variance. The random intercept terms suggest that there is greater variability among speakers than among lemmas despite speaker normalization, consistent with prior reports (Clayards, 2018).

Voiced stops

The fitted model for voiced stops predicting our average-normalized initial pitch measure is presented in Table 3.3. The factor of interest, word-initial stop voicing minimal pair competitor existence, is bolded for ease of reference.

As with the voiceless stops, the fitted model intercept shows that pitch is higher at stop offset than it is for the rest of the syllable. Of note, however, this effect is approximately 1/4 the size of the effect for voiceless stops, consistent with a phonological effect of stop voicing on pitch (see also Fig. 3.3). Again, as with the voiceless stops, we find a significant gender effect showing that female speakers tend to have a greater difference in pitch

Variable	$beta$	SE	t	$p(\|t\|)$
(Intercept)	0.292	0.059	4.946	< 0.001***
Gender (female-male)	0.120	0.044	2.729	< 0.01**
Age (older-younger)	-0.126	0.044	-2.873	< 0.01**
POA (labial-coronal)	-0.036	0.048	-0.736	0.463
POA (velar-coronal)	-0.138	0.058	-2.375	< 0.05*
Word Familiarity	0.024	0.012	1.868	0.064
Backward Probability	-0.008	0.010	-0.869	0.385
Phonotactic Probability	-0.113	0.040	-2.848	< 0.01**
WordFam : BackProb	0.006	0.002	2.448	< 0.05*
Min Pair Exist (true-false)	**0.015**	**0.043**	**0.352**	**0.726**

(a) Fixed effects summary

Group	Variable	Variance	SD	Corr
Lemma	(Intercept)	0.089	0.299	
Speaker	(Intercept)	0.056	0.237	
	Min Pair Exist (true-false)	**0.008**	**0.091**	**0.47**
Residual		1.305	1.142	

(b) Random effects summary

Table 3.3: Fixed and random effects summaries for the fitted model predicting the difference between initial pitch following stop offset and the average pitch of the rest of the token (in semitones) for voiced stops.

between stop offset and the rest of the syllable than male speakers.

In the voiced stop model, however, we also find a significant effect of speaker age, with older speakers showing smaller differences between stop offset pitch and the token's average pitch as compared to younger speakers. We also find a significant effect of place of articulation, most notably in the form of smaller pitch difference between stop offset and token average for the velar voiced stop /g/ relative to the coronal voiced stop /d/ (the simple effect of the difference in average-normalized initial pitch for /b/ relative to /d/ was not significant).

Also unlike in the voiceless stop model, backward conditional probability did not significantly improve model fit in its own right. Furthermore, the phonotactic probability of the word correlated significantly with our dependent measure, such that more common phoneme sequences were correlated with smaller pitch differences between stop offset and the rest of the syllable. We also see a significant interaction between word familiarity and backward conditional probability, such that more familiar words that are relatively predictable given the preceding word have a greater difference between initial pitch and the rest of the syllable. This is consistent with Chuoying Ouyang and Kaiser's (2014) finding that local and global measures of predictability interact in the realization of prosodic prominence.

As with the voiceless stops, our factor of interest, word-initial stop voicing minimal pair lemma existence, did not significantly impact model fit.

The random effect structure again attributes little variance to the correlated slope for our factor of interest on the speaker intercept. Unlike in the voiceless stop model, we see less variability between speakers than between items, although the difference is slight.

3.5.2 Midpoint-normalized initial pitch

Our midpoint-normalized measure of initial pitch following stop offset compares pitch immediately following the word-initial stop to the pitch at the midpoint of the ROI. As such, our hypotheses for both the intercept and our factor of interest were, *a priori*, the same as they were for the average-normalized measure of initial pitch.

Recall from section 3.2.3 that the choice to evaluate initial pitch in two different ways was done in large part to validate the data. As such, we expect the fitted models for midpoint-normalized initial pitch to match the fitted models for average-normalized initial pitch in their essential components, but vary in the details. Consequently, given the results from the average-normalized initial pitch models, it is unlikely that our hypotheses regarding the factor of interest will be borne out in these models.

Voiceless stops

The fitted model for voiceless stops predicting midpoint-normalized initial pitch is presented in Table 3.4. The factor of interest, word-initial stop voicing minimal pair competitor existence, is bolded for ease of reference.

Variable	β	SE	t	$p(\|t\|)$
(Intercept)	1.082	0.096	11.318	< 0.001***
Gender (female-male)	0.264	0.086	3.078	< 0.01**
High Vowel (true-false)	-0.092	0.031	-2.922	< 0.01**
Following Liquid (true-false)	0.102	0.029	3.548	< 0.001***
Number of Syllables (1-2)	0.088	0.025	3.566	< 0.001***
Word Familiarity	-0.023	0.009	-2.633	< 0.01**
Forward Probability	-0.006	0.008	-0.702	0.483
Backward Probability	-0.016	0.008	-2.045	< 0.05*
ForProb : BackProb	-0.007	0.002	-2.923	< 0.01**
Min Pair Exist (true-false)	**0.018**	**0.037**	**0.477**	**0.635**

(a) Fixed effects summary

Group	Variable	Variance	SD	Corr
Lemma	(Intercept)	0.047	0.217	
Speaker	(Intercept)	0.274	0.524	
	Min Pair Exist (true-false)	**0.002**	**0.042**	**0.09**
Residual		1.402	1.184	

(b) Random effects summary

Table 3.4: Fixed and random effects summaries for the fitted model predicting the difference between initial pitch following stop offset and the midpoint of the token (in semitones) for voiceless stops.

In line with the expectation that this model would mirror the voiceless stop model for average-normalized initial pitch, both models have matching significant effects of comparable magnitude for the intercept, speaker gender, following liquid, number of syllables, backward conditional probability, and the interaction between forward and backward conditional probability. Similarly, the factor of interest does not significantly improve model fit for either dependent measure of initial pitch.

However, this model differs from the average-normalized initial pitch model in that the simple effect of word familiarity is significant in this model, such that more familiar words have lower onset pitch relative to the midpoint of the token. The models also differ in that the height of the vowel was retained as a significant factor. It is perhaps unsurprising that vowel height was a more significant predictor of pitch relative to the token midpoint than it was relative to the token average, because the midpoint likely falls directly on the vowel itself for many tokens. The token average, on the other hand, will contain additional influences from other segments within the syllable (such as liquids, glides, and codas) as well as additional coarticulatory effects from content outside the syllable.

Within the random effects structure, we see that the random slope for our factor of interest once again accounted for very little variance. As with the voiceless stop model based on token averages, the speaker intercept accounted for considerably more variance than the lemma intercept. Indeed, the random intercept for speaker accounts for a similar amount of variability as in the model based on token averages (compare variance = 0.274, sd = 0.524 and variance = 0.211, sd = 0.459).

Voiced stops

The fitted model for voiced stops predicting midpoint-normalized initial pitch is presented in Table 3.5. The factor of interest, word-initial stop voicing minimal pair competitor existence, is bolded for ease of reference.

The fitted voiced stop model for midpoint-normalized initial pitch matches the fitted voiced stop model for average-normalized initial pitch in that both models have matching significant effects of comparable magnitude for the intercept, speaker gender, speaker age, place of articulation, and phonotactic probability, and our factor of interest accounts for very little model variance and consequently does not improve model fit.

As with the voiceless stops, there is an additional effect of vowel height in this model that was not present in the model based on token averages. Again, we feel this is most likely due to the fact that the token midpoint will usually fall directly on the vowel, while the token average will often include additional noise from other segments. In addition, we find a significant effect of speech rate such that faster speech correlates with smaller differences between onset pitch and token midpoint pitch. This makes sense given that faster speech rates will generally correlate with shorter syllables, thereby decreasing the amount of time that transpires between stop offset and the midpoint of the syllable. Finally, as in the average-normalized initial pitch model for voiced stops, we find a significant

Variable	β	SE	t	$p(\|t\|)$
(Intercept)	0.237	0.078	3.046	< 0.01**
Gender (female-male)	0.172	0.059	2.930	< 0.01**
Age (older-younger)	-0.170	0.058	-2.905	< 0.01**
High Vowel (true-false)	-0.142	0.049	-2.927	< 0.01**
POA (labial-coronal)	-0.023	0.054	-0.423	0.672
POA (velar-coronal)	-0.159	0.065	-2.446	< 0.05*
Word Familiarity	0.021	0.014	1.494	0.137
Backward Probability	-0.004	0.011	-0.316	0.752
Phonotactic Probability	-0.090	0.044	-2.019	< 0.05*
Speech Rate	-0.048	0.022	-2.209	< 0.05*
WordFam : BackProb	0.005	0.003	1.974	< 0.05*
Min Pair Exist (true-false)	**0.008**	**0.048**	**0.175**	**0.862**

(a) Fixed effects summary

Group	Variable	Variance	SD	Corr
Lemma	(Intercept)	0.108	0.329	
Speaker	(Intercept)	0.104	0.323	
	Min Pair Exist (true-false)	**0.009**	**0.094**	**0.25**
Residual		1.665	1.290	

(b) Random effects summary

Table 3.5: Fixed and random effects summaries for the fitted model predicting the difference between initial pitch following stop offset and the midpoint of the token (in semitones) for voiced stops.

interaction between word familiarity and backward conditional probability whereby more familiar words are realized with greater differences between stop offset pitch and token midpoint pitch when those words are also more predictable given the previously uttered word.

Within the random effects structure, we see that the random slope for our factor of interest is once again accounting for very little variance in the data. As with the average-normalized initial pitch model for voiced stops, the random intercept for lemma accounts for slightly more variability than the random intercept for speaker in this model.

3.5.3 Pitch slopes as change in pitch per millisecond

Our measure of pitch slopes evaluates change in pitch per millisecond from stop offset to the midpoint of the ROI. Unlike our measures of initial pitch, the expectation that pitch is higher at stop offset than it is in the rest of the token leads to the hypothesis that the model intercepts for both voiced and voiceless stops will be negative. As such, negative beta estimates in the models represent *more* rapid decreases in pitch over time, whereas positive beta estimates represent *less* rapid decreases in pitch over time. This is perhaps conceptually counter to the interpretation of positive vs. negative beta estimates for our dependent measures of initial pitch.

Our hypotheses regarding our factor of interest for this measure are also different from our hypotheses for the models of initial pitch in one key way: we only hypothesized that speakers may enhance F0 contrasts in the form of maintaining higher F0 following voiceless stops for longer in words with stop voicing minimal pairs relative to words without such minimal pairs (Hoole and Honda, 2011). Evidence in support of this hypothesis would take the form of a positive correlation between minimal pair existence and pitch slope, such that the negative pitch slopes that we expect to be reflected by the model intercept for voiceless stops would be less dramatic in words with stop voicing minimal pairs relative to words without such minimal pairs. Nonetheless we include fitted model results for voiced stops as well as voiceless stops for completeness' sake.

Voiceless stops

The fitted model for voiceless stops predicting pitch slope is presented in Table 3.6. The factor of interest, word-initial stop voicing minimal pair competitor existence, is bolded for ease of reference.

The fitted model intercept tells us that, on average across speakers and items, pitch decreases from stop offset to syllable midpoint at a rate of nearly 3/4 semitone per millisecond. Consistent with the pitch onset data and prior reports that female speakers have more dynamic pitch contours than men, a significant gender effect in our data shows that

Variable	$beta$	SE	t	$p(\|t\|)$
(Intercept)	-0.739	0.060	-12.416	< 0.001***
Gender (female-male)	-0.196	0.047	-4.143	< 0.001***
Diphthongal Vowel (true-false)	0.044	0.019	2.307	< 0.05*
Syntactic Category (verb-noun)	0.007	0.041	0.179	0.858
Syntactic Category (adj-noun)	0.019	0.048	0.385	0.701
Syntactic Category (adv-noun)	-0.135	0.087	-1.563	0.120
Forward Probability	0.001	0.005	0.220	0.826
Backward Probability	0.006	0.005	1.293	0.196
ForwProb : BackProb	0.006	0.001	4.020	< 0.001***
Min Pair Exist (true-false)	$>$ **-0.001**	**0.025**	**-0.003**	**0.997**

(a) Fixed effects summary

Group	Variable	Variance	SD	Corr
Lemma	(Intercept)	0.023	0.151	
Speaker	(Intercept)	0.080	0.282	
	Min Pair Exist (true-false)	**0.003**	**0.056**	**-0.14**
Residual		0.536	0.732	

(b) Random effects summary

Table 3.6: Fixed and random effects summaries for the fitted model predicting pitch slope (as change in semitones per millisecond) for voiceless stops. Note that, as a model of pitch slopes that *typically* are decreasing due to falling lexical pitch (note the intercept), positive estimates for fixed effects are indicative of a shallower slope, and negative estimates are indicative of a steeper slope.

pitch changes more rapidly for female speakers than male speakers. An additional significant effect for whether or not the token's vowel is diphthongal tells us that pitch decreases more rapidly in tokens with diphthongal vowels relative to tokens with monophthongal vowels. Given that the overall formant structure of diphthongal vowels is more dynamic than it is in monophthongal vowels, it is perhaps unsurprising that fundamental frequency is also more dynamic in these tokens.

We also find that syntactic category significantly improves model fit, although this is not reflected in the form of any significant simple effects for relative levels of the syntactic category factor. The beta estimates between levels of this factor suggest that pitch changes most rapidly for adverbs, then nouns, then verbs, and finally adjectives. However, none of these individual comparisons appears to be significant in its own right.

Though both forward and backward conditional probabilities did not significantly improve model fit in their own right, a strong interaction between these factors was found such that more contextually predictable words were realized with smaller rates of pitch change over time. Once again, our factor of interest did not contribute significantly to model fit.

The model for pitch slopes following voiceless stops thus differs from the models for voiceless stop initial pitch in a few key ways. Unlike in both of initial pitch models, neither the presence of a following liquid nor the number of syllables in the word token significantly improved the current model fit, whereas the diphthongal status of the vowel and the syntactic category of the word did.

Within the random effects structure, we again see very little variance accounted for by the random slope for our factor of interest. We also see greater variance accounted for by the random intercept for speaker relative to the random intercept for lemma, consistent with Clayards (2018) and our previous voiceless stop models.

Voiced stops

The fitted model for voiced stops predicting pitch slope is presented in Table 3.7. The factor of interest, word-initial stop voicing minimal pair competitor existence, is bolded for ease of reference. Note that the random slope for the factor of interest on the speaker intercept was removed to allow this model to converge.

The fitted model intercept tells us that, on average across speakers and items, pitch decreases from stop offset to syllable midpoint at a rate of approximately 2/11 semitone per millisecond in our voiced stop data. As with the voiceless stops, the female speakers show a more rapid decline in pitch per millisecond than the male speakers.

Unlike with the voiceless stops, but as in the other voiced stop models, we find a significant effect of age such that pitch decreases less rapidly for older speakers than for younger speakers. We also find an effect of place of articulation realized most prominently in the form of less rapid pitch decreases for the velar voiced stop /g/ relative to the coronal voiced

Variable	β	SE	t	$p(\lvert t \rvert)$
(Intercept)	-0.182	0.033	-5.511	< 0.001***
Gender (female-male)	-0.099	0.028	-3.539	< 0.01**
Age (older-younger)	0.087	0.028	3.139	< 0.01**
POA (labial-coronal)	0.002	0.023	0.096	0.924
POA (velar-coronal)	0.006	0.028	2.233	< 0.05*
Word Familiarity	-0.012	0.006	-2.134	< 0.05*
Min Pair Exist (true-false)	**-0.012**	**0.020**	**-0.608**	**0.544**

(a) Fixed effects summary

Group	Variable	Variance	SD
Lemma	(Intercept)	0.020	0.140
Speaker	(Intercept)	0.024	0.155
Residual		0.325	0.571

(b) Random effects summary

Table 3.7: Fixed and random effects summaries for the fitted model predicting pitch slope (as change in semitones per millisecond) for voiced stops. Note that, as a model of pitch slopes that *typically* are decreasing due to falling lexical pitch (note the intercept), positive estimates for fixed effects are indicative of a shallower slope, and negative estimates are indicative of a steeper slope.

stop /d/, consistent with both voiced stop initial pitch models. Finally, we also see an effect of word familiarity, such that more familiar words are realized with a more rapid decline in pitch per millisecond relative to less familiar words. Our factor of interest, word-initial stop voicing minimal pair lemma existence, did not improve model fit.

Notable differences between the model for pitch slope for voiced stops relative to the models for initial pitch for voiced stops include a lack of effect for phonotactic probability and a lack of any significant interaction terms in the pitch slope model. Also worthy of note is that the factor for whether a vowel is diphthongal did not contribute significantly to model fit for the pitch slope model for voiced stops, as it had done for voiceless stops.

Within the random effects structure, we once again see that, within the voiced stops, there is little difference in the variance accounted for by the random intercepts for speaker and lemma.

3.5.4 Summary of statistical models

While the fitted models for different dependent measures all vary in their details, a few things are consistent across the models. All model intercepts suggest that pitch at stop offset is generally higher than it is in the rest of the token, consistent with downward-sloping lexical pitch contours. This effect is consistently larger for voiceless stops than for voiced stops, consistent with a phonological voicing effect on following F0[14]. All models also show significant effects of speaker gender, with female speakers showing signs of more dynamic pitch contours than male speakers. An effect of speaker age, however, was consistently significant in the voiced stop models, but not the voiceless stop models. Furthermore, in all voiceless stop models a comparatively large portion of variability is attributed to the random effect of speaker identity relative to lemma identity, consistent with prior observations that the use of pitch as a cue to stop voicelessness varies considerably more across speakers than across items (Clayards, 2018). This difference was consistently not found in the voiced stop models, where variances attributed to the random intercepts for lemma and speaker were fairly comparable in all cases. We tentatively suggest that this may be due to the presence or absence of the fixed effect of speaker age, which, when retained in the model, should absorb some portion of the by-speaker variability that the random intercept picks up in its absence.

Differences across the fitted models were, for the most part, explainable. For example, the fixed effect of vowel height was retained in both the voiced and voiceless models of midpoint-normalized initial pitch, likely due to the fact that the token midpoint will often fall on the vowel itself. However, this factor was not retained in the models of pitch slope, which also make explicit reference to the token midpoint, casting some doubt on this assertion. The fact that both following liquid and the number of syllables in the word were significant predictors of voiceless stop initial pitch measures, but not voiceless stop pitch

[14] Although this was not tested directly in any of these models, a model of the effect of phonological voicing on average-normalized initial pitch was fitted to these data during initial data exploration. This model did not include most of the control predictors included in these models, but did show a highly significant effect of stop voicing on the average-normalized initial pitch measure.

slope, may also be explainable. We suggest that both of these factors represent something about the size of the ROI. In the case of midpoint-normalized initial pitch, this also implies something about the amount of time that transpires between stop offset and token midpoint; with more time comes more opportunity for pitch to change. Indeed, the fact that these factors are no longer significant in the pitch slope measure, which is normalized for time, lends support to this claim. In the case of average-normalized initial pitch, a longer ROI means a greater disparity in the proportion of the ROI that is attributed to initial versus token-average pitch during token normalization; the longer the ROI, the more opportunity there is for lower pitch data to bring down the token average, thereby increasing the difference between pitch at stop offset and pitch in the rest of the token. Finally, at least in the voiceless stop data, pitch changed more rapidly in syllables with diphthongal vowels than it did in syllables with monophthongal vowels, consistent with the assumption that diphthongal vowels are produced with more dynamic frequencies than monophthongal vowels. Though we acknowledged that this could affect any of our dependent measures, it was primarily included as a control for the pitch slope measure.

Of note, however, there are some less easily explained nuances to these findings. As already noted, why was the effect of vowel height not found in either of the models of pitch slope, for which the dependent measure also makes reference to the token midpoint? Likewise, why does the diphthongal status of a vowel not significantly impact model fits for pitch slope in the voiced stop data? In addition, we have no systematic explanation for why the effects of place of articulation and speaker age were only ever retained in models for voiced stops. The most likely reasons for these differences may be such factors as the overall variability in the data (and thereby the amount of room that the models have to attribute variance to our assorted predictors) and particulars of the hill-climbing algorithm used by the models.

It is also worth noting that the intrinsic effect of lexical pitch (i.e., the y-intercept estimate) for voiceless stops was slightly greater than one semitone (i.e., pitch at stop offset was about one semitone higher than the token average or midpoint pitch). The effect of lexical pitch for voiced stops, on the other hand, was about 1/4 of a semitone. Similarly, pitch decreased in voiceless stop tokens at an average rate of approximately 3/4 semitone per millisecond, whereas it decreased in voiced stop tokens at an average rate of approximately 2/11 of a semitone. Together these results suggest that F0 at stop offset is roughly four times higher for voiceless stops relative to voiced stops, but that this heightened F0 drops at roughly four times the rate in voiceless stops relative to voiced stops. In other words, there is no evidence from these data that speakers are prolonging pitch differences based on word-initial stop voicing, as the difference in stop offset pitch is corrected at a rate commensurate with the difference itself.

Indeed, given the formula for calculating speaker semitones (3.1), which divides the speaker's pitch range into 12 steps, 3/4 semitone per millisecond is quite rapid, suggesting that speakers may be more interested in returning to a "normal" pitch for the utterance than they are in signaling a lexical contrast. If this interpretation is correct, it is consistent with the fact that our factor of interest did not significantly improve the fit of any of our models. This suggests that speakers are not manipulating F0 to signal lexical contrast in word-initial stops in any way that is detectable in these models, although we wish to emphasize how tenuous conclusions based on null results can be.

3.6 Does F0 Co-Vary with Voice Onset Time?

While some previous studies have reported that VOT itself is not a good predictor of following F0 within stop voicing categories of English (e.g., Clayards, 2018), other studies have reported significant correlations between VOT and F0 within stop voicing categories (e.g., Dmitrieva et al., 2015). In order to understand the extent to which stop offset F0 realizations can be predicted by VOT realizations, we conducted additional statistical analyses on the subset of our data for which we already have hand-annotated VOT measurements (Chapter 2; see also Wedel et al., 2018). These data consist of 24 of the 39 speakers analyzed in section 3.5, and include measurements of stop closure durations as well as combined durations for burst and aspiration (i.e., VOT). For full methodological details regarding these measurements as well as inclusion and exclusion criteria, see Chapter 2 and Wedel et al. (2018).

Statistical analysis of these data was conducted as in section 3.5, using the dependent measures of initial pitch only[15]. Separate models were fit for voiced and voiceless stops for each of the two dependent measures of speaker- and token-normalized initial pitch. The factor of interest in each of these models was the ratio of stop voice onset time to total stop duration, which provides a local control for speech rate (Chapter 2; see also Wedel et al., 2018). This "VOT/Stop-length ratio" is represented by the formula:

$$\frac{burst + aspiration}{closure + burst + aspiration} \tag{3.5}$$

Where closure, burst, and aspiration are all measures of duration in milliseconds for stop closure, stop release burst, and post-release aspiration, respectively. This results in values between 0 and 1 representing the relative proportion of total stop duration that is taken up by the VOT (i.e., burst + aspiration). As with other continuous measures described in section 3.5, the VOT/Stop-length ratio was centered and linearly scaled to match the distributional range of the dependent measures for each model[16].

3.6.1 Average-normalized initial pitch

Although it is clear that VOT and F0 correlate across stop voicing categories in English, results concerning the correlation between VOT and F0 within stop voicing categories of English is less clear (Shultz et al., 2012; Dmitrieva et al., 2015; Clayards, 2018). For voiceless stops, Dmitrieva et al. (2015) report a significant negative correlation between VOT and F0 across speakers (based on data from Shultz et al., 2012), whereas Clayards (2018) reports no correlation between VOT and F0 across speakers. For voiced stops, Dmitrieva et al. (2015) report a significant positive correlation between VOT and F0 across speak-

[15]The rate of pitch change following the stop was not considered for these analyses because there was no theoretical reason to hypothesize that VOT itself would correlate with different rates of pitch change. The reader is referred to section 3.2.3 for the motivation behind the pitch slope measure used in section 3.5.

[16]Specifically, the dependent measures for initial pitch both fall within a range between -10 and +10; as such, the VOT/Stop-length ratio was multiplied by 10 after centering to fit this same range.

ers (see also Shultz et al., 2012), whereas Clayards (2018) reports a significant negative correlation between VOT and F0 across speakers.

Given these drastic differences, it was unclear what relationship, if any, we should expect between VOT and F0 within voiced and voiceless stop categories. In the event that VOT and F0 are correlated, a few possibilities emerge. On the one hand, following the results of Dmitrieva et al. (2015), VOT and F0 in word tokens with voiceless stops may be negatively correlated, signaling a trade-off relationship whereby a weaker articulation of one cue is compensated for by a stronger articulation of the other cue. This would suggest a strongly phonological relationship between these cues that exists within speakers' control. On the other hand, if we were to find evidence of a positive correlation between VOT and F0 for voiceless stops, this would suggest that these cues co-exist in a cooperative relationship, possibly outside of the control of speakers. We did not anticipate this outcome given prior results. For voiced stops, however, prior evidence suggests either a positive or a negative correlation between VOT and F0. As with the voiceless stops, a negative correlation would suggest a trade-off relationship whereas a positive correlation would suggest a more automatic (likely phonetic/physiological) relationship between VOT and F0.

Voiceless stops

The fitted model for voiceless stops predicting average-normalized initial pitch, as a function of the VOT/Stop-length ratio, is presented in Table 3.8. The factor of interest, the VOT/Stop-length ratio, is bolded for ease of reference.

As with the average-normalized initial pitch model for voiceless stops presented in section 3.5, the intercept tells us that initial pitch is more than one semitone higher that pitch in the rest of the syllable, on average. The most noteworthy differences between this model and the corresponding model from section 3.5 are the lack of a significant effect of speaker gender and the significant effects of both following glide and the interaction between following glide and the presence of a high vowel.

The lack of a significant correlation between the VOT/Stop-length ratio and stop offset

Variable	β	SE	t	$p(\|t\|)$
(Intercept)	1.355	0.148	9.186	< 0.001***
High Vowel (true-false)	0.189	0.102	1.849	0.066
Following Liquid (true-false)	0.095	0.034	2.847	< 0.01**
Following Glide (true-false)	0.268	0.102	2.615	< 0.01**
Number of Syllables (1-2)	0.113	0.028	3.963	< 0.001***
Word Familiarity	-0.014	0.012	-1.185	0.238
Backward Probability	-0.011	0.010	-1.108	0.268
HighVow : FollGlide	0.219	0.102	2.142	< 0.05*
WordFam : BackProb	-0.008	0.003	-2.568	< 0.05*
VOT/Stop-length Ratio	**0.018**	**0.022**	**0.836**	**0.412**

(a) Fixed effects summary

Group	Variable	Variance	SD	Corr
Lemma	(Intercept)	0.073	0.269	
Speaker	(Intercept)	0.251	0.501	
	VOT/Stop-length ratio	**0.003**	**0.054**	**0.47**
Residual		1.053	1.026	

(b) Random effects summary

Table 3.8: Fixed and random effects summaries for the fitted model predicting average-normalized initial pitch (in semitones) for voiceless stops, as a function of the VOT/Stop-length ratio.

F0 within these voiceless stop data suggests that VOT is not a good predictor of F0 realizations within the voiceless stop category. This is consistent with the findings of Clayards (2018), but inconsistent with the findings of Dmitrieva et al. (2015).

Voiced stops

The fitted model for voiced stops predicting average-normalized initial pitch, as a function of the VOT/Stop-length ratio, is presented in Table 3.9. The factor of interest, the VOT/Stop-length ratio, is bolded for ease of reference.

Again, the intercept appears comparable to that reported for the model of average-normalized initial pitch for voiced stops in section 3.5. Noteworthy differences in these models include the absence of any effect of speaker age or gender in the present model, and that forward conditional probability interacts with word familiarity in this model, whereas

Variable	β	SE	t	$p(\|t\|)$
(Intercept)	0.342	0.083	4.133	< 0.001***
POA (labial-coronal)	-0.126	0.068	-1.835	0.068
POA (velar-coronal)	-0.114	0.081	-1.409	0.161
Word Familiarity	-0.002	0.017	-0.138	0.890
Phonotactic Probability	-0.160	0.052	-3.059	< 0.001**
Forward Probability	0.031	0.017	1.897	0.058
WordFam : ForwProb	0.009	0.005	1.963	< 0.05*
VOT/Stop-length Ratio	**-0.022**	**0.043**	**-0.520**	**0.607**

(a) Fixed effects summary

Group	Variable	Variance	SD	Corr
Lemma	(Intercept)	0.111	0.333	
Speaker	(Intercept)	0.079	0.282	
	VOT/Stop-length ratio	**0.018**	**0.132**	**0.33**
Residual		1.299	1.140	

(b) Random effects summary

Table 3.9: Fixed and random effects summaries for the fitted model predicting average-normalized initial pitch (in semitones) for voiced stops, as a function of the VOT/Stop-length ratio.

it was backward conditional probability that interacted with word familiarity in the previous model.

The lack of a significant correlation between the VOT/Stop-length ratio and stop offset F0 within these voiced stop data again suggests that VOT is not a good predictor of post-stop F0 realizations. However, it should be noted that the direction of the trend is opposite that of the voiceless stop data, which could point to a possible interaction between stop voicing and VOT-contingent F0 variation.

3.6.2 Midpoint-normalized initial pitch

As in section 3.5, the use of two different dependent measures of initial pitch was done primarily to validate the data. Our hypotheses were therefore the same as they were for the models based on average-normalized initial pitch. Given the results of those models, however, we do not expect to find significant correlations between VOT and F0 for either voiced or voiceless stops in these data.

Voiceless stops

The fitted model for voiceless stops predicting midpoint-normalized initial pitch, as a function of the VOT/Stop-length ratio, is presented in Table 3.10. The factor of interest, the VOT/Stop-length ratio, is bolded for ease of reference.

Variable	β	SE	t	$p(\|t\|)$
(Intercept)	1.120	0.127	8.831	< 0.001***
High Vowel (true-false)	-0.129	0.044	-2.544	< 0.05*
Following Liquid (true-false)	0.041	0.045	0.925	0.357
Number of Syllables (1-2)	0.109	0.030	3.614	< 0.001***
Word Familiarity	-0.020	0.012	-1.611	0.109
Backward Probability	-0.012	0.011	-1.112	0.266
HighVow : FollLiquid	-0.102	0.044	-2.286	< 0.05*
WordFam : BackProb	-0.012	0.004	-3.201	< 0.01**
VOT/Stop-length Ratio	**0.028**	**0.027**	**1.042**	**0.309**

(a) Fixed effects summary

Group	Variable	Variance	SD	Corr
Lemma	(Intercept)	0.062	0.250	
Speaker	(Intercept)	0.327	0.572	
	VOT/Stop-length ratio	**0.007**	**0.082**	**0.45**
Residual		1.305	1.142	

(b) Random effects summary

Table 3.10: Fixed and random effects summaries for the fitted model predicting midpoint-normalized initial pitch (in semitones) for voiceless stops, as a function of the VOT/Stop-length ratio.

Again, the model looks similar to the midpoint-normalized initial pitch model from section 3.5, although with the conspicuous absence of a gender effect and a significant interaction between vowel height and following liquid (note that this mirrors the differences for the models based on average-normalized initial pitch, except in that case the interaction was between vowel height and following glide).

As with the model for stop offset pitch relative to token average pitch, the VOT/Stop-length ratio did not significantly improve model fit. We also see that the predictors retained in the fitted model are comparable to the voiceless stop model of average-normalized initial pitch.

Voiced stops

The fitted model for voiced stops predicting midpoint-normalized initial pitch, as a function of the VOT/Stop-length ratio, is presented in Table 3.11. The factor of interest, the VOT/Stop-length ratio, is bolded for ease of reference.

Variable	β	SE	t	$p(\|t\|)$
(Intercept)	0.376	0.100	3.747	< 0.001***
POA (labial-coronal)	-0.138	0.079	-1.763	0.079
POA (velar-coronal)	-0.128	0.093	-1.372	0.172
Word Familiarity	-0.004	0.020	-0.194	0.846
Phonotactic Probability	-0.156	0.060	-2.593	< 0.05*
Forward Probability	0.046	0.019	2.468	< 0.05*
WordFam : ForwProb	0.013	0.005	2.424	< 0.05*
VOT/Stop-length Ratio	**-0.033**	**0.051**	**-0.641**	**0.527**

(a) Fixed effects summary

Group	Variable	Variance	SD	Corr
Lemma	(Intercept)	0.159	0.399	
Speaker	(Intercept)	0.127	0.357	
	VOT/Stop-length ratio	**0.028**	**0.167**	**0.30**
Residual		1.624	1.274	

(b) Random effects summary

Table 3.11: Fixed and random effects summaries for the fitted model predicting midpoint-normalized initial pitch (in semitones) for voiced stops, as a function of the VOT/Stop-length ratio.

This model looks akin to the model of average-normalized initial pitch following voiced stops reported above. Again, the VOT/Stop-length ratio did not significantly improve model fit, though again the direction of the trend for these voiced stop data is opposite that of the voiceless stop data.

3.6.3 Summary of statistical models

The models presented here differ from the models in section 3.5 in some key ways. Most notably, the effects of speaker age and gender were not significant in any of the models when the VOT/Stop-length ratio was also included in the model, despite gender always being significant when stop voicing minimal pair lemma existence was included in the model, and despite age being significant in such models of voiced stops. This is interesting given that the same dependent measures are being predicted in all of these models, but is perhaps less interesting when we consider that the present models were built on a subset of the data from section 3.5. We offer no explanation for these differences other than to say that (i) the two datasets differ, with 15 more speakers in the dataset from section 3.5, and (ii) the presence of the VOT/Stop-length ratio versus minimal pair existence, two predictors which we already know correlate in these data (see Chapter 2), may affect the hill-climbing algorithm used to fit these models in different ways.

The effect of the VOT/Stop-length ratio was not significant in any of the models investigated here. This suggests that F0 is not co-variant with VOT *per se*. This is consistent with the lack of a predictive relationship between minimal pair existence and F0 despite a predictive relationship between minimal pair existence and VOT. In the case of voiceless stops, this result in also consistent with the findings of Clayards (2018), but inconsistent with the findings of Dmitrieva et al. (2015). In the case of voiced stops, the lack of a correlation between VOT and F0 stands in contrast to the findings of both Dmitrieva et al. (2015) and Clayards (2018). However, given that their results showed opposite correlations, the lack of any correlation in our data may be representative of a generally unreliable relationship between VOT and F0 for voiced stops of English.

3.7 Post-Hoc Analysis: Does VOT Affect F0 for Words with Minimal Pairs Only?

The models for average- and midpoint-normalized initial pitch and for the rate of pitch change following stop offset suggest no correlation between the existence of a word-initial stop voicing minimal pair competitor in the lexicon and the realization of F0 following said stop (cf. section 3.5). This finding runs counter to the expectations of the contrastive hyperarticulation hypothesis, which predicts that cues to lexical-phonological contrasts will be enhanced in tokens of words with minimal pairs for that contrast. In addition, the models for stop offset pitch as a function of VOT suggest no correlation between VOT *per se* and the F0 following stop offset (cf. section 3.6). This latter finding is consistent with prior reports that these cues do not correlate within voicing categories (Kingston and Diehl, 1994; Clayards, 2018), and suggests that post-stop F0 differences in English are not attributable to phonetic or physiological factors alone. Taken together with the fact that F0 is still consistently contrastive between voiced and voiceless stop categories, this suggests that F0 has become phonologized, with speakers doing something to enhance these F0 differences independently of automatic phonetic factors. However, if F0 is truly "secondary" to VOT as a stop voicing cue (e.g., Clayards, 2018), perhaps speakers put extra effort into voicing-dependent F0 contrasts only when this contrast matters most. In other words, maybe there is a relationship between VOT and F0, but it is only realized in words with stop voicing minimal pairs.

To address this question, a post-hoc analyses was conducted to investigate the relationship between VOT and F0 as a function of word-initial stop voicing minimal pair lemma existence, looking for one of two possible relationships. First, under the assumption that VOT and F0 exist in a trade-off relationship (e.g., Dmitrieva et al., 2015), perhaps this trade-off relationship is simply context-dependent. Given that speakers tend to enhance VOT contrasts in words with stop voicing minimal pairs relative to words without such minimal pairs (Chapter 2; see also Wedel et al., 2018), perhaps they only bother to enhance F0 contrasts *specifically when they failed to sufficiently enhance VOT*. In such a case, we would expect to find an interaction between VOT and minimal pair existence such that, within words with stop voicing minimal pairs only, VOT and F0 are negatively correlated (i.e., if a word has an initial stop voicing minimal pair, stop offset F0 is higher in voiceless stops that have relatively shorter VOTs, and is lower in voiced stops that have relatively longer VOTs). Alternatively, if VOT and F0 do not exist in a trade-off relationship, they should be positively correlated, but this correlation may also be context-dependent such that speakers enhance F0 proportionally to VOT *only in words with stop voicing minimal pairs*. Under this view, we would expect to find an interaction between VOT and minimal pair existence such that, within words with stop voicing minimal pairs only, VOT and F0 are positively correlated (i.e., if a word has an initial stop voicing minimal pair, stop offset F0 is higher in voiceless stops that have relatively longer VOTs, and is lower in voiced stops that have relatively shorter VOTs)[17]. This latter outcome would be consistent with the contrastive hypothesis in that both VOT and F0 are enhanced together in words with minimal pairs, despite not generally correlating in the lexicon at large.

While these two possibilities are in direct opposition to one another, both predict an interaction between VOT and the existence of a word-initial stop voicing minimal pair competitor lemma. As such, the model building process was repeated as described in section 3.6, but with the addition of a factor for word-initial stop voicing minimal pair competitor lemma existence (as described in section 3.2.4) and the interaction between this minimal pair factor and the VOT/Stop-length ratio. In order to facilitate the interpretation of the hypothesized effects, the minimal pair factor was dummy coded with reference to the factor level corresponding to the existence of a minimal pair competitor (using R's `as.logical.factor()` function; R Core Team, 2018). This ensures that the estimates reported for the interaction term in the model summaries reflect the effect of VOT on the realization of F0 *for words with minimal pairs* specifically.

Following Barr et al. (2013), the full interaction between minimal pair existence and the VOT/Stop-length ratio was included as a correlated random slope on the speaker intercept of the maximal model used for model fitting. However, as this made for a fairly complicated random effects structure with limited theoretical basis, the LmerTest package's (Kuznetsova et al., 2017) `step()` function was allowed to remove random effect

[17]Note that we would not expect to find this correlation between F0 and VOT in our data at large, but rather only in the limited context of word-initial stop voicing minimal pairs. As such, this hypothesis is independent of the previous hypothesis that VOT would be correlated with F0 within each stop voicing category, and is instead dependent on the interaction between VOT and minimal pair existence.

terms that did not significantly improve model fit, provided that neither random intercept (speaker or lemma) was removed from the model (to avoid violating the independence assumption).

Before we turn to our models, it should be noted that the conflicting results of previous studies regarding the correlations between VOT and F0 came from laboratory elicitation studies for which the target stimuli were all initial stop voicing minimal pairs (Dmitrieva et al., 2015; Clayards, 2018). Thus, those previous studies were not investigating correlations between VOT and F0 throughout a sample of the entire English lexicon (as was investigated in section 3.6), but rather specifically from the subset of words in the lexicon that have initial stop voicing minimal pairs. That is, the data yielding those previous results correspond most closely to the subset of data under investigation here. Given the different results of those two studies, it may seem unclear why we should expect to find any such correlation in our words with minimal pairs in this dataset. Nonetheless, there is reason to believe that spontaneous speech, such as that from the Buckeye Corpus under study here, and laboratory-elicited speech, such as that from Dmitrieva et al. (2015) and Clayards (2018), come from different speech registers with differing properties that can affect the potential for identifying contrast enhancement (see, e.g., Gahl and Strand, 2016, for discussion). Indeed, in Chapter 2, I argue that clear-speech-like properties of laboratory-elicited speech may obscure signs of contrastive hyperarticulation for some cues, and that some signs of contrastive hyperarticulation may be more readily identifiable under speech conditions that promote general phonetic reduction and thereby greater category overlap, such as spontaneous conversational speech (Ernestus and Warner, 2011; see also Wedel et al., 2018). Since Dmitrieva et al. (2015) did find a correlation between VOT and F0 in their English data for both voiced and voiceless stops, all of which came from stop voicing minimal pairs, we consider it quite possible that such a correlation is present in our spontaneous speech data despite the conflicting laboratory results of Clayards (2018).

3.7.1 Average-normalized initial pitch

Voiceless stops

The fitted model for voiceless stops predicting average-normalized initial pitch, as a function of the interaction between word-initial stop voicing minimal pair lemma existence and the VOT/Stop-length ratio, is presented in Table 3.12. The factor of interest, the interaction between initial stop voicing minimal pair competitor existence and the VOT/Stop-length ratio, is bolded for ease of reference.

Variable	β	SE	t	$p(\|t\|)$
(Intercept)	1.352	0.147	9.180	< 0.001***
High Vowel (true-false)	0.189	0.102	1.844	0.066
Following Liquid (true-false)	0.093	0.034	2.759	< 0.01**
Following Glide (true-false)	0.267	0.102	2.604	< 0.01**
Number of Syllables (1-2)	0.117	0.029	3.983	< 0.001***
Word Familiarity	-0.016	0.012	-1.387	0.167
Backward Probability	-0.010	0.010	-1.001	0.317
HighVow : FollGlide	0.213	0.102	2.094	< 0.05*
WordFam : BackProb	-0.008	0.003	-2.426	< 0.05*
Min Pair Exist (true)	-0.444	0.235	-1.887	0.059
VOT/Stop-length Ratio	0.033	0.204	0.162	0.872
MPexist : VOTratio	**0.882**	**0.420**	**2.101**	$<$ **0.05***

(a) Fixed effects summary

Group	Variable	Variance	SD
Lemma	(Intercept)	0.071	0.266
Speaker	(Intercept)	0.251	0.501
Residual		1.056	1.028

(b) Random effects summary

Table 3.12: Fixed and random effects summaries for the fitted model predicting average-normalized initial pitch (in semitones) for voiceless stops, as a function of the interaction between word-initial stop voicing minimal pair lemma existence and the VOT/Stop-length ratio.

In this model, neither the existence of a word-initial stop voicing minimal pair lemma nor the VOT/Stop-length ratio were predictive of stop offset F0 in their own right. However, the interaction between VOT and minimal pair lemma existence significantly improved model fit such that a larger VOT/Stop-length ratio was significantly predictive of a larger difference between stop offset F0 and the F0 in the rest of the token, but only for words with stop voicing minimal pairs. That the two simple effects are not significant while the interaction term is suggests that there is only a correlation between VOT and F0 among word-initial voiceless stops with voiced-stop minimal pair competitor lemmas. This is inconsistent with the findings of either Dmitrieva et al. (2015) or Clayards (2018), but is instead consistent with the contrastive hypothesis as outlined above.

Voiced stops

The fitted model for voiced stops predicting average-normalized initial pitch, as a function of the interaction between word-initial stop voicing minimal pair lemma existence and the VOT/Stop-length ratio, is presented in Table 3.13. The factor of interest, the interaction between initial stop voicing minimal pair competitor existence and the VOT/Stop-length ratio, is bolded for ease of reference.

Unlike in the corresponding voiceless stop model, not only were the factors for minimal pair lemma existence and the VOT/Stop-length ratio not significant, but the interaction between these two terms was not significant either. This suggests that, while speakers may be manipulating F0 as a cue to stop voicing in words with initial voiceless stops that have voiced stop minimal pairs, they do not manipulate F0 in a corresponding fashion for voiced stops. This stands in contrast to previous reports that VOT is contrastively enhanced for both voiced and voiceless stops in these same data (Chapter 2; Wedel et al., 2018). Furthermore, this is consistent with the view that voicing-contingent F0 variation in aspiration languages is manifested as raised F0 following voiceless stops only (Hoole and Honda, 2011).

3.7.2 Midpoint-normalized initial pitch

Voiceless stops

The fitted model for voiceless stops predicting midpoint-normalized initial pitch, as a function of the interaction between word-initial stop voicing minimal pair lemma existence and

Variable	β	SE	t	$p(\lvert t\rvert)$
(Intercept)	0.310	0.094	3.296	< 0.01**
POA (labial-coronal)	-0.117	0.069	-1.693	0.092
POA (velar-coronal)	-0.124	0.081	-1.533	0.127
Word Familiarity	-0.005	0.018	-0.280	0.780
Forward Probability	0.030	0.017	1.843	0.066
Phonotactic Probability	-0.167	0.053	-3.157	< 0.01**
WordFam : ForwProb	0.010	0.005	2.087	< 0.05*
Min Pair Exist (true)	0.174	0.168	1.039	0.301
VOT/Stop-length Ratio	-0.013	0.481	-0.026	0.979
MPexist : VOTratio	**-0.507**	**0.603**	**-0.841**	**0.400**

(a) Fixed effects summary

Group	Variable	Variance	SD	Corr	
Lemma	(Intercept)	0.113	0.336		
Speaker	(Intercept)	0.082	0.286		
	VOT/Stop-length ratio	1.626	1.275	-0.40	
	Min Pair Exist (true)	0.099	0.314	-0.07	-0.28
Residual		1.277	1.130		

(b) Random effects summary

Table 3.13: Fixed and random effects summaries for the fitted model predicting average-normalized initial pitch (in semitones) for voiced stops, as a function of the interaction between word-initial stop voicing minimal pair lemma existence and the VOT/Stop-length ratio.

the VOT/Stop-length ratio, is presented in Table 3.14. The factor of interest, the interaction between initial stop voicing minimal pair competitor existence and the VOT/Stop-length ratio, is bolded for ease of reference.

As in the model for average-normalized initial pitch, the interaction between minimal pair existence and the VOT/Stop-length ratio was significantly positive, suggesting that VOT and F0 are positively correlated in voiceless stop-initial words with voiced stop minimal pairs. Again, the simple effect of the VOT/Stop-length ratio was not significant, reinforcing that this correlation between VOT and F0 only exists amongst words with minimal pairs and not in the rest of the lexicon.

However, unlike in any of our previous models, we also find a significant main effect of

Variable	β	SE	t	$p(\|t\|)$
(Intercept)	1.110	0.127	8.751	< 0.001***
High Vowel (true-false)	-0.107	0.044	-2.411	< 0.05*
Following Liquid (true-false)	0.038	0.044	0.860	0.392
Number of Syllables (1-2)	0.109	0.031	3.497	< 0.001***
Word Familiarity	-0.023	0.012	-1.845	0.067
Backward Probability	-0.012	0.011	-1.038	0.299
HighVow : FollLiquid	-0.099	0.044	2.249	< 0.05*
WordFam : BackProb	-0.011	0.004	-3.097	< 0.01**
Min Pair Exist (true)	-0.556	0.257	-2.162	< 0.05*
VOT/Stop-length Ratio	0.033	0.284	0.115	0.909
MPexist : VOTratio	**1.171**	**0.461**	**2.544**	$<$ **0.05***

(a) Fixed effects summary

Group	Variable	Variance	SD	Corr
Lemma	(Intercept)	0.059	0.244	
Speaker	(Intercept)	0.326	0.571	
	VOT/Stop-length ratio	0.007	0.081	0.46
Residual		1.303	1.142	

(b) Random effects summary

Table 3.14: Fixed and random effects summaries for the fitted model predicting midpoint-normalized initial pitch (in semitones) for voiceless stops, as a function of the interaction between word-initial stop voicing minimal pair lemma existence and the VOT/Stop-length ratio.

minimal pair competitor existence, such that words with such minimal pairs have significantly lower F0 at stop offset relative to the rest of the token as compared to words without such minimal pairs. It is difficult to interpret this effect in light of the previous findings that F0 is not predicted by minimal pair existence. However, it should be noted that the effect of minimal pair existence was marginally significant for the model of voiceless stop average-normalized initial pitch, with the trend for the effect going in the same direction as the significant effect reported here (note also that these data are from a subset of the data analyzed in section 3.5).

Voiced stops

The fitted model for voiced stops predicting midpoint-normalized initial pitch, as a function of the interaction between word-initial stop voicing minimal pair lemma existence and the VOT/Stop-length ratio, is presented in Table 3.15. The factor of interest, the interaction between initial stop voicing minimal pair competitor existence and the VOT/Stop-length ratio, is bolded for ease of reference.

Variable	β	SE	t	$p(\|t\|)$
(Intercept)	0.353	0.110	3.221	< 0.01**
POA (labial-coronal)	-0.143	0.079	-1.827	0.070
POA (velar-coronal)	-0.128	0.093	-1.382	0.169
Word Familiarity	-0.007	0.020	-0.330	0.742
Forward Probability	0.046	0.019	2.467	< 0.05*
Phonotactic Probability	-0.152	0.060	-2.518	< 0.05*
WordFam : ForwProb	0.012	0.005	2.380	< 0.05*
Min Pair Exist (true)	0.042	0.171	0.245	0.807
VOT/Stop-length Ratio	-0.355	0.571	-0.621	0.538
MPexist : VOTratio	**0.128**	**0.666**	**0.193**	**0.847**

(a) Fixed effects summary

Group	Variable	Variance	SD	Corr
Lemma	(Intercept)	0.155	0.394	
Speaker	(Intercept)	0.127	0.356	
	VOT/Stop-length ratio	0.028	0.167	0.30
Residual		1.625	1.275	

(b) Random effects summary

Table 3.15: Fixed and random effects summaries for the fitted model predicting midpoint-normalized initial pitch, as a function of the interaction between word-initial stop voicing minimal pair lemma existence and the VOT/Stop-length ratio.

As in the model for average-normalized initial pitch, none of the effects of minimal pair existence, the VOT/Stop-length ratio, nor their interaction were significant, reinforcing the suggestion that VOT and F0 are not correlated for voiced stops in this dataset, regardless of stop voicing minimal pair competitor existence. Indeed, that even the directions of these three effects are not consistent across these two models further strengthens the view that any correlations between these predictors and stop offset F0 are completely spurious.

3.7.3 Confirming the interaction by subsetting the data

The models looking for an interaction between voice onset time and the existence of a stop voicing minimal pair in the lexicon suggest two things. First, the models suggest that this interaction does exist for words beginning with voiceless stops, but not for words beginning with voiced stops. This stands in opposition to findings that voice onset time itself is contrastively enhanced for both stop voicing categories (Chapter 2; Wedel et al., 2018). Second, the model estimates support the hypothesis that VOT and F0 do not exist in a trade-off relationship in these English speech data (contra Dmitrieva et al., 2015), but instead are positively correlated in words with minimal pair competitors, consistent with the contrastive hypothesis.

To confirm these findings, we repeated the analysis separately on the voiceless stop data with minimal pairs and without minimal pairs. If our interpretation of the interaction reported above is correct, VOT should significantly predict F0 within the voiceless stop data with minimal pair competitors, but it should not significantly predict F0 within the voiceless stop data without minimal pair competitors. We report newly fitted models to these data subsets to get a more complete picture of how the data may differ across the subsets, but significant findings were confirmed using the previously-fit models from above as well (see Appendix B for summaries of those models).

Voiceless stop-initial words that have a stop voicing minimal pair in the lexicon

Average-normalized initial pitch

The fitted model for tokens of voiceless stop-initial lemmas with existing minimal pair lemmas for stop voicing, predicting average-normalized initial pitch as a function of the VOT/Stop-length ratio, is presented in Table 3.16. The factor of interest, the VOT/Stop-length ratio, is bolded for ease of reference.

Variable	β	SE	t	$p(\|t\|)$
(Intercept)	1.174	0.120	9.802	< 0.001***
VOT/Stop-length Ratio	**0.109**	**0.035**	**3.141**	$<$ **0.05***

(a) Fixed effects summary

Group	Variable	Variance	SD	Corr
Lemma	(Intercept)	< 0.001	0.021	
Speaker	(Intercept)	0.293	0.541	
	VOT/Stop-length ratio	**0.004**	**0.064**	**-0.49**
Residual		0.863	0.929	

(b) Random effects summary

Table 3.16: Fixed and random effects summaries for the fitted model predicting average-normalized initial pitch (in semitones) for voiceless stop-initial words with stop voicing minimal pairs only, as a function of the VOT/Stop-length ratio.

Midpoint-normalized initial pitch

The fitted model for tokens of voiceless stop-initial lemmas with existing minimal pair lemmas for stop voicing, predicting midpoint-normalized initial pitch as a function of the VOT/Stop-length ratio, is presented in Table 3.17. The factor of interest, the VOT/Stop-length ratio, is bolded for ease of reference.

Variable	β	SE	t	$p(\|t\|)$
(Intercept)	1.286	0.143	9.010	< 0.001***
VOT/Stop-length Ratio	**0.157**	**0.037**	**4.219**	$<$ **0.01****

(a) Fixed effects summary

Group	Variable	Variance	SD	Corr
Lemma	(Intercept)	0.006	0.075	
Speaker	(Intercept)	0.416	0.645	
	VOT/Stop-length ratio	**0.002**	**0.047**	**0.31**
Residual		1.052	1.026	

(b) Random effects summary

Table 3.17: Fixed and random effects summaries for the fitted model predicting midpoint-normalized initial pitch (in semitones) for voiceless stop-initial words with stop voicing minimal pairs only, as a function of the VOT/Stop-length ratio.

Of note, fitting new models to these more restricted data resulted in the elimination of all other predictors besides the VOT/Stop-length ratio for both versions of our initial pitch measure (average- and midpoint-normalized initial pitch). This suggests that variability in stop offset F0 for voiceless stops with voiced stop minimal pairs is, according the lme4 package's hill-climbing algorithm (Bates et al., 2014), better attributed to variation in VOT than any of the other factors under consideration. In both cases, the estimated coefficient for the VOT/Stop-length ratio was positive, suggesting that, among voiceless stops with initial stop voicing minimal pairs, longer VOTs correlate with higher initial pitch. As already noted, the same results obtain if we use the fitted models reported in tables 3.12 and 3.14, but with the terms for minimal pair existence and the interaction between minimal pair existence and VOT removed; that is, such models also result in a significant positive correlation between VOT and F0, with no other significant predictors besides the model intercept (Appendix B).

Voiceless stop-initial words that do not have a stop voicing minimal pair in the lexicon

Average-normalized initial pitch

The fitted models for tokens of voiceless stop-initial lemmas with no existing minimal pair lemmas for stop voicing, predicting average-normalized initial pitch as a function of the VOT/Stop-length ratio, is presented in Table 3.18. The factor of interest, the VOT/Stop-length ratio, is bolded for ease of reference.

Midpoint-normalized initial pitch

The fitted models for tokens of voiceless stop-initial lemmas with no existing minimal pair lemmas for stop voicing, predicting midpoint-normalized initial pitch as a function of the VOT/Stop-length ratio, is presented in Table 3.19. The factor of interest, the VOT/Stop-length ratio, is bolded for ease of reference.

As predicted given the interaction reported above, VOT is no longer significantly predictive of stop offset F0 amongst voiceless stop-initial words with no stop voicing minimal pair competitor in the lexicon. This further demonstrates that, while both VOT and F0 vary

Variable	β	SE	t	$p(\lvert t \rvert)$
(Intercept)	1.362	0.151	8.996	< 0.001***
High Vowel (true-false)	0.191	0.107	1.779	0.077
Following Liquid (true-false)	0.093	0.038	2.423	< 0.05*
Following Glide (true-false)	0.283	0.107	2.635	< 0.01**
Number of Syllables (1-2)	0.116	0.034	3.395	< 0.001***
Word Familiarity	-0.012	0.014	-0.852	0.395
Backward Probability	-0.016	0.011	-1.383	0.167
HighVow : FollGlide	0.227	0.107	2.122	< 0.05*
WordFam : BackProb	-0.011	0.004	-2.776	< 0.01**
VOT/Stop-length Ratio	$>$ **-0.001**	**0.027**	**-0.024**	**0.981**

(a) Fixed effects summary

Group	Variable	Variance	SD	Corr
Lemma	(Intercept)	0.090	0.300	
Speaker	(Intercept)	0.252	0.502	
	VOT/Stop-length ratio	**0.006**	**0.080**	**0.50**
Residual		1.088	1.043	

(b) Random effects summary

Table 3.18: Fixed and random effects summaries for the fitted model predicting average-normalized initial pitch (in semitones) for voiceless stop-initial words without stop voicing minimal pairs only, as a function of the VOT/Stop-length ratio.

as a function of stop voicing, within the voiceless stop category the two cues do not co-vary except in words with stop voicing minimal pairs. This is consistent with the hypothesis that F0 is contrastively hyperarticulated as a function of stop voicing minimal pair competitor existence. The present study examined the role of F0 in signaling voicing contrasts following word-initial stop consonants of spontaneous English speech. This involved two primary investigations, each of an independent research question. First, we asked whether F0 contrasts following voiced and voiceless stops are enhanced in words with stop voicing minimal pairs relative to words without such minimal pairs. No such effect was found for either stop type in a corpus of 39 speakers from the Buckeye Corpus (Pitt et al., 2007). Second, we asked whether F0 was predicted by the voice onset time of the stop preceding it. We found no signs of a predictive relationship between VOT and post-stop F0 in a corpus of 24 speakers from the Buckeye Corpus. For both of these research questions,

Variable	β	SE	t	$p(\|t\|)$
(Intercept)	1.112	0.128	8.725	< 0.001***
High Vowel (true-false)	-0.113	0.049	-2.291	< 0.05*
Following Liquid (true-false)	0.034	0.049	0.693	0.490
Number of Syllables (1-2)	0.109	0.036	3.031	< 0.01**
Word Familiarity	-0.016	0.015	-1.095	0.275
Backward Probability	-0.018	0.013	-1.388	0.165
HighVow : FollLiquid	-0.105	0.049	-2.147	< 0.05*
WordFam : BackProb	-0.015	0.004	-3.464	< 0.001***
VOT/Stop-length Ratio	**0.004**	**0.032**	**0.122**	**0.904**

(a) Fixed effects summary

Group	Variable	Variance	SD	Corr
Lemma	(Intercept)	0.078	0.280	
Speaker	(Intercept)	0.320	0.566	
	VOT/Stop-length ratio	**0.011**	**0.103**	**0.43**
Residual		1.353	1.163	

(b) Random effects summary

Table 3.19: Fixed and random effects summaries for the fitted model predicting midpoint-normalized initial pitch (in semitones) for voiceless stop-initial words without stop voicing minimal pairs only, as a function of the VOT/Stop-length ratio.

we fit mixed effects models to three different dependent measures of post-stop F0, each normalized with respect to the speaker and token that they came from, and our factors of interest were not predictive of any of these dependent measures. These results initially suggested that, though F0 varies as a function of phonological stop voicing, this variation is not dependent on the VOT differences between voiced and voiceless stops, and speakers do nothing to enhance these voicing-dependent F0 contrasts as a function of initial stop voicing minimal pair competitor existence.

3.8 Discussion

Prior evidence had strongly suggested that a variety of cues to phonological contrasts are enhanced in words with minimal pair competitors for those contrasts (e.g., Baese-Berk and Goldrick, 2009; Buz et al., 2016; Wedel et al., 2018). What is more, evidence suggests that speakers can simultaneously enhance multiple cues to a given contrast in these minimal pairs (Seyfarth et al., 2016). As such, that VOT is contrastively enhanced in words with stop voicing minimal pairs (Chapter 2) but F0 should not be, despite both cues being strongly dependent on phonological stop voicing, was a surprising result. However, prior evidence also suggests that listeners may only capitalize on F0 differences between voiced and voiceless stops when the VOT of those stops is not sufficiently contrastive (e.g., see Hall, 2011, and citations therein). Consequently, a post-hoc investigation asked whether speakers do enhance F0 contrasts in words with stop voicing minimal pairs, but do so relative to the realized VOT of the initial stop. In other words, we asked whether an interaction exists between the relationship of VOT with F0 and the existence of a stop voicing minimal pair competitor. We found that this interaction term significantly improved our voiceless stop models of initial pitch following stop offset, such that VOT was positively correlated with initial pitch for words with stop voicing minimal pair competitors only. No correlation between VOT and F0 was evident in words without initial stop voicing minimal pairs, and no correlation between VOT and F0 was evident at all in the voiced stop data. This evidence was suggestive of a cooperative but context-dependent relationship between VOT and F0 for voiceless stops only.

The results of this study support a number of prior claims in the literature. Firstly, Seyfarth et al. (2016) tested the effect of coda voicing minimal pairs on the realization of vowel duration and active voicing of the coda in one-syllable words with coda fricatives. They found that both cues were dynamically enhanced in words with minimal pairs relative to words without minimal pairs. Likewise, we found that, in addition to previously reported VOT contrast enhancement, F0 is also contrastively enhanced in words with stop voicing minimal pairs. This is further evidence that speakers can and do enhance multiple phonetic cues to phonemic contrasts in words with minimal pairs for those contrasts.

Secondly, a number of researchers have suggested that, though both VOT and F0 are important cues to stop voicing contrasts in a variety of languages (Lisker and Abramson, 1964; Kirby et al., 2015; Bang et al., 2018; Coetzee et al., 2018), F0 is secondary to VOT in English (Abramson and Lisker, 1985; Lisker, 1986). We found support for this notion in the fact that, while VOT is contrastively enhanced for both voiced and voiceless stops of English in words with stop voicing minimal pairs, F0 is only enhanced in voiceless stops, and only as a function of the VOT. If F0 were an equally important cue to stop voicing contrasts, we would expect to find signs of F0 enhancement independently of VOT enhancement.

Thirdly, a number of researchers have suggested that voicing-dependent F0 variation is driven by phonological rather than phonetic factors (Kingston and Diehl, 1994; Shultz et al., 2012; Dmitrieva et al., 2015). Indeed, Hoole and Honda (2011) suggest that, even though an underlying articulatory (phonetic) source is responsible for higher F0 following voiceless obstruents (specifically, cricothyroid muscle tension), some speakers of German capitalize on this articulatory-driven difference to further enhance the effect. In other words, for at least some speakers of German, voicing-dependent F0 variation is controlled (and therefore at least partially phonological). Our findings support the general notion that voicing-dependent F0 variation is more phonologically driven than phonetically driven. We believe this to be supported by the fact that the existence of a stop voicing minimal pair competitor significantly affects the relationship between VOT and F0 at all. If the existence of a lexical-phonological competitor drives otherwise unaccounted-for differences in the realization of F0, it becomes harder to argue that these differences are entirely due to phonetic or physiological factors[18]. What is more, if voicing-dependent F0 variation were driven by predominantly phonetic factors related to the realization of VOT, we believe that we would have found evidence of this fact in the form of a simple correlation between VOT and F0 in the full voiceless stop dataset. That we only find a correlation between these cues as a function of minimal pair existence is more suggestive of phonological influences.

Finally, Hoole and Honda (2011) also suggested that the effect of stop voicing on following F0 in aspiration languages (such as German and English) takes the form of higher F0 following voiceless stops, and not lowering of F0 following voiced stops. We find evi-

[18]Of course, it may be that there are phonetic factors at play for these voiceless stops other than VOT that we have not accounted for, and which are correlated with the existence of a voiced stop minimal pair.

dence of this in the fact that only voiceless stops show signs of a minimal pair effect, albeit only as a function of VOT. Nonetheless, that no such effect is evident for voiced stops supports the overall picture that F0 is heightened beyond what is phonetically necessary after voiceless stops, but is generally affected in voiced stops by more phonetic or physiological factors, such as the phonetic voicing of the stop closure.

Thus, we have found evidence that F0 is enhanced alongside VOT following word-initial voiceless stops in productions of lemmas with an initial stop voicing minimal pair competitor in a corpus of spontaneous conversational English. This finding improves our understanding of the relationship between VOT and F0 in English. While both VOT and F0 vary systematically as a function of stop voicing, speakers manipulate these cues very differently as sources of enhancement of lexical-phonological contrasts. Prior studies have shown that, in the aggregate, the VOT of voiced stops in words with voiceless stop minimal pairs is shorter than it is in words without voiceless stop minimal pairs; likewise, the VOT of voiceless stops in words with voiced stop minimal pairs is longer than it is in words without voiced stop minimal pairs (Chapter 2; see also Wedel et al., 2018). Here, however, we found that, in the aggregate, F0 does not vary as a function of stop voicing minimal pair existence in any of the ways we looked for it (initial pitch following the stop relative to the token's average pitch, initial pitch following the stop relative to the pitch of the token's midpoint, or the change in pitch per millisecond from the initial pitch following the stop to the token's midpoint).

Nonetheless, the relationship between F0 and VOT does vary as a function of stop voicing minimal pair existence for voiceless stops. Whereas there was no relationship between VOT and F0 within each stop category in the majority of our data, there was such a relationship specifically for tokens of voiceless stop-initial lemmas with stop voicing minimal pairs. For such lemmas, VOT and F0 were positively correlated such that tokens with longer VOTs were also produced with higher F0 following the stop. We take this as evidence that speakers are enhancing a phonologized correlation between VOT and F0 in precisely those words for which the phonological contrast is lexically relevant.

It is important to note that previous researchers have found different correlations between VOT and F0 in English than that reported here. Dmitrieva et al. (2015) found a negative correlation between VOT and F0 among voiceless stops of English from a previous elicitation study (Shultz et al., 2012)[19]. In the study, speakers produced monosyllabic words, including both /p/- and /b/-initial minimal pairs and fillers beginning with other consonants, one at a time. A correlational analysis indicated that, among participants' productions of /p/-initial words from /p/ ~ /b/ minimal pairs, VOT was weakly but significantly negatively correlated with F0. Clayards (2018), on the other hand, found no correlation between VOT and F0 among voiceless stops of English, also in an elicitation study. In her study, speakers also produced monosyllabic words, including both /p/ ~ /b/ minimal pairs and fillers beginning with other consonants, one at a time. As part of analyses of correlations between VOT, following F0, and following vowel duration, Clayards found no correlation between VOT and F0 following the stop. It should be noted that both of these studies examined tokens of words that do have initial stop voicing minimal pair competitors. As such, their findings relate closely to our own post-hoc analysis of the interaction between VOT and minimal pair existence in section 3.7. In contrast to the findings of either Dmitrieva et al. (2015) or Clayards (2018), we found evidence of a positive correlation between VOT and F0 among words with initial stop voicing minimal pairs (Dmitrieva et al., 2015 also report no correlation between VOT and F0 in their Spanish data. See also Kirby et al., 2015 for relevant data from French and Italian showing no correlation and a positive correlation, respectively, between VOT and F0 for voiceless stops).

This difference in findings means that our results should be considered with a degree of caution. While a number of reasons could potentially be responsible for these different results, we believe that two possible reasons stand out. The first is related to differences in speech registers between laboratory elicitation studies and corpus studies of spontaneous speech. Some researchers have suggested considerable differences in the speech styles found in these two types of study that can affect a variety of factors (e.g., Gahl and Strand,

[19](Shultz et al., 2012) also found a negative correlation between VOT and F0 in their complete data, including both voiced and voiceless stops.

2016). For example, I argue in Chapter 2 that laboratory-elicited speech may more closely resemble clear speech registers than the speech found in more spontaneous or conversational speech such as that in the Buckeye Corpus under investigation here (see also Wedel et al., 2018). Clear speech is associated with such phonetic properties as longer durations (including VOT in voiceless stops) and more dynamic pitch than other speech registers (see, e.g., Smiljanić and Bradlow, 2008, 2009). This could help to explain why our results differ from those of previous laboratory studies.

The second possible source of this different finding is related to the materials used in the study. Both Dmitrieva et al. (2015) and Clayards (2018) included only monosyllabic target words beginning with labial stops (for the English data). The present study, by contrast, includes both mono- and di-syllabic words beginning with labial, coronal, and velar stops. We note that both the number of syllables and the place of articulation of the stop were significant predictors in a number of our statistical models, suggesting that these factors may influence F0 realizations in their own right or relative to VOT (which is influenced by these factors; see Chapter 2 and Wedel et al., 2018). In addition, while Dmitrieva et al. (2015), Clayards (2018), and Kirby et al. (2015) all normalized their F0 measurements with respect to the speaker, they did so with reference to each speaker's initial F0 measurements following target stops only, and they performed no token-relative normalization of pitch other than that related to speech rate. In the present study, however, we performed our speaker normalization with respect to each speaker's productions of entire target syllables *except for the initial pitch measurements*, and we also normalized each initial pitch measurement with respect to the pitch of the token it came from. These differences may have had a variety of influences on the specific measures used to represent F0 following target stops. For example, in these previous studies, speaker-normalization using post-stop initial pitch values of roughly equal numbers of voiced and voiceless stop productions should have generally resulted in positive F0 values for voiceless stops and negative values for voiced stops. While our method of speaker normalization also leads to generally positive F0 values for voiceless stops and negative values for voiced stops, additional token normalization leads to generally positive initial pitch values for both voiced and voiceless stops (i.e., it represents how much higher the pitch is after the stop compared to the rest of the syllable). This would change the relative scale of the dependent measure, which may affect the correlation between this measure and VOT.

3.9 Conclusion

We conducted analyses of F0 following voiced and voiceless stops from the Buckeye Corpus of Conversational Speech (Pitt et al., 2007). We examined influences of two different variables on the realization of these F0 values using linear mixed effects models. In section 3.5, we examined the effect of the existence of an initial stop voicing minimal pair on these F0 values for 39 speakers from the corpus. In section 3.6, we examined the effect of the relative voice onset time of the stop on these F0 values for 24 speakers from the corpus. In both cases, our factor of interest did not significantly improve model fits for either voiced or voiceless stops. However, in a post-hoc analysis, we investigated the relationship between these two variables. Specifically, we examined whether voice onset time and following F0 correlate for words with stop voicing minimal pairs only. We found this to be the case for our voiceless stop data, where we found a positive correlation between the voice onset time of the stop and the F0 immediately following that stop in tokens of lemmas with voiced stop minimal pairs. We found no such correlation in tokens of voiceless stop-initial words without voiced stop minimal pairs. We also found no correlation between voice onset time and F0 in our voiced stop data.

These results support the view that both primary and secondary phonetic cues to lexical-phonological contrasts can be enhanced by speakers, but that enhancement of secondary cues may be contingent upon the primary cue. However, further study is needed to understand the full extent of this relationship, especially given that the results reported here stand in opposition to some previous findings (Dmitrieva et al., 2015; Clayards, 2018; Kirby et al., 2015).

CHAPTER 4

SUMMARY AND GENERAL DISCUSSION

In this dissertation, I aimed to clarify the nature and role of contrastive hyperarticulation in English. Contrastive hyperarticulation is a form of hyperarticulation whereby the existence of competing phonological neighbors in the lexicon triggers enhancement of the phonetic cues that signal the lexical-phonological contrast between target and competitor. As such, contrastive hyperarticulation is an example of both speech hyperarticulation and lexically conditioned phonetic variation. Of particular interest in the present research is the implication that contrastive hyperarticulation entails a different kind of phonetic variation than many other hyperarticulation phenomena, such as clear speech hyperarticulation. Specifically, most *hyper*-articulation phenomena involve longer articulatory durations or more extreme articulatory gestures. These cases of hyperarticulation are typically considered in opposition to reduction phenomena, where shorter articulatory durations or less pronounced articulatory gestures are categorized as *hypo*-articulation (generally classified as the H & H Theory: Lindblom, 1990). However, contrastive hyperarticulation is specifically geared toward enhancing phonetic contrasts, not articulatory gestures. Although these enhancements are measured in terms of phonetic-acoustic properties (such as voice onset time and fundamental frequency), the implication is that the perceptual distance between target and competitor is of greater priority than the extent of any particular articulatory gesture. As such, it has been argued that contrastive hyperarticulation can lead to phonetic outcomes that are more typically treated as examples of reduction (or hypoarticulation. See, e.g., Seyfarth et al., 2016; Wedel et al., 2018).

Although support in the literature for contrastive hyperarticulation, thus defined, has grown substantially over the past decade (beginning most notably with Baese-Berk and Goldrick, 2009), the phenomenon itself is not clearly differentiated in the literature from other types of hyperarticulation or lexically conditioned phonetic variation. This has perhaps led to a growing amount of seemingly conflicting research in the domain of lexically conditioned hyperarticulation. For example, Fricke (2013), Fox et al. (2015), and Wedel et al. (2018) each found evidence that word-initial voiceless stop VOTs are predicted by different forms of lexical competition. For Fricke (2013), the number of phonological neighbors that have a different onset from the target was the best predictor of voiceless stop VOTs; for Fox et al. (2015), it was the total number of phonological neighbors; and for Wedel et al. (2018), it was the existence of the voiced stop minimal pair. These three different outcomes occurred despite all three studies investigating, at a minimum, the effect of neighborhood density and minimal pair existence. One possible reason for the different outcomes is related to methodological choices. For example, while both Fricke (2013) and Wedel et al. (2018) analyzed spontaneous speech data, Fox et al. (2015) analyzed elicited speech in a sentence reading task; while both Fricke (2013) and Fox et al. (2015) included only mono-syllabic words with simplex onsets, Wedel et al. (2018) also included di-syllabic words and words with complex onsets; and within the analyses of spontaneous speech (Fricke, 2013; Wedel et al., 2018), the lexicons used to determine which words exist or do not exist were constructed in different ways. However, an important additional possibility is related to the fact that many researchers, including Fricke (2013) and Fox et al. (2015), limit their investigations to only look for hyperarticulation in the form of longer durations or articulatory extents, whereas others, including Wedel et al. (2018), look for hyperarticulation in the form of greater phonetic distance between targets and competitors (see also Schertz, 2013; Clopper and Tamati, 2014; Seyfarth et al., 2016).

In this dissertation, I have attempted to clarify some of these issues by framing contrastive hyperarticulation as a distinct phenomenon, independent of other types of hyperarticulation or lexically conditioned phonetic variation[1]. Specifically, two studies were conducted to clarify (i) the *triggers* of contrastive hyperarticulation, i.e., the kinds of lexical competition that are most correlated with contrast-enhancing phonetic realizations (Chapter 2, investigating VOT as the primary cue to word-initial stop voicing contrasts), and (ii) the *outcomes* of contrastive hyperarticulation, specifically whether secondary or redundant cues to a contrast are contrastively enhanced in addition to primary cues (Chapter 3, investigating F0 as a secondary cue to word-initial stop voicing contrasts).

4.1 Summary of Study 1

In Study 1 (Chapter 2), the nature of lexical competitors that trigger contrastive hyperarticulation was investigated. In this study, voice onset time (VOT), the primary cue to the word-initial stop voicing contrast, was measured for tokens of word-initial stops from 24 speakers in the Buckeye Corpus of Conversational English (Pitt et al., 2005, 2007). Various lexical-phonological competition metrics were calculated to measure the number of lexical-phonological competitors for each target word. These competition metrics were crafted to systematically sample the hypothesis space of potential lexical competition that may affect the realization of word-initial VOT, including overall lexical-phonological neighborhood density, initial stop voicing minimal pair competitor existence, and a variety of intermediary neighborhood measures. Linear mixed effects regression (LMER) models were fit to the VOT data along with a number of control predictors. Each model contained one competition metric, and models were compared according to their corrected Akaike's Information Criteria (AIC_c), with the effect of the competition metric in top-performing models tested for significance using Log-Likelihood Ratio Tests (LLRTs) of nested models. For lexical competition measures based on raw competition counts, the existence of a word-initial stop

[1]This is not to say that other kinds of hyperarticulation or lexically conditioned phonetic variation are non-existent, or even that they do not potentially occur alongside contrastive hyperarticulation – to the contrary, my position is that contrastive hyperarticulation is simply one of numerous phenomena interacting in a complex system of influences affecting the phonetic realizations of speech targets.

voicing minimal pair competitor outperformed all other lexical competition measures for both voiced and voiceless stops. For voiceless stops, the existence of a stop voicing minimal pair competitor correlated with longer VOTs, and significantly improved model fit. For voiced stops, the existence of a stop voicing minimal pair competitor correlated with shorter VOTs, and significantly improved model fit.

The process was repeated for lexical competition measures that were weighted according to the frequency of competitor lemmas. For voiced stops, the frequency-weighted stop voicing minimal pair competitor existence measure outperformed all other models, again predicting shorter VOTs. For voiceless stops, the frequency-weighted measure of overall neighborhood density was the top-performing model. However, after removing a small number of lemmas that had no lexical-phonological neighbors, it was found that this measure did not significantly improve model fit, whereas the existence of a stop voicing minimal pair competitor did. These results were taken as evidence of contrastive hyperarticulation, with phonetic cues to lexical-phonological contrasts being contrastively hyperarticulated to maximize the phonetic distance between the target word and the minimal pair competitor.

4.2 Summary of Study 2

In Study 2 (Chapter 3), the phonetic outcomes of contrastive hyperarticulation were investigated to see whether a secondary/redundant cue to a contrast is hyperarticulated alongside the primary cue. In this study, fundamental frequency (F0), a secondary cue to the word-initial stop voicing contrast in English, was measured for tokens of word-initial stops from 39 speakers in the Buckeye Corpus. LMER models were fit to these data along with a number of control predictors. It was ultimately found that F0 was enhanced alongside the primary VOT cue in words beginning with voiceless stops *that have initial voiced stop minimal pairs only*. This finding was interpreted as limited evidence of contrastive hyperarticulation. The reader is reminded, however, that this result was found in a post-hoc analysis following two planned analyses.

In the first planned analysis, evidence of contrastive hyperarticulation was sought in the form of an effect of stop voicing minimal pair existence on three different dependent F0 measures. Existence of the stop voicing minimal pair competitor was not found to affect model fits for any of the three F0 measures. In the second planned analysis, the relationship between VOT and F0 was investigated for the 24 speakers from Study 1 with VOT measurements (Chapter 2). VOT was included as a predictor in new models to see whether within-stop-voicing-category F0 values are predicted by VOT or if they vary independently of VOT. VOT was not found to predict within-category F0 for voiced or voiceless stops, suggesting that the effect of phonological voicing on following stop F0 is not an automatic by-product of VOT itself (see Dmitrieva et al., 2015, for related argumentation).

It was in light of these two null results that a third, post-hoc investigation was conducted to determine whether a relationship between VOT and F0 exists for words with minimal pairs only. F0 data from the 24 speakers with VOT measurements were reanalyzed with an interaction term included in the models between stop voicing minimal pair existence and VOT. It was found that VOT was positively correlated with F0 for voiceless stops with voiced stop minimal pairs only. VOT was not significantly correlated with F0 for voiceless stops without minimal pairs, or for voiced stops with or without minimal pairs. These results were taken as evidence of contrastive hyperarticulation and for a controlled, phonological effect of stop voicing on following F0. Specifically, since the correlation between VOT and F0 was only found for voiceless stops with stop voicing minimal pairs, this was considered evidence of enhancement of a redundant cue to a lexical-phonological contrast. As voiceless stops of English in word-initial position are realized with both longer VOTs and higher following F0s than their voiced stop counterparts, a positive correlation between these cues suggests that, when one cue is hyperarticulated, so is the other cue. This follows conventional wisdom regarding redundant cues: though the secondary cue often does not provide additional information, it does provide the listener with an additional means of correctly identifying the intended signal (e.g., Kingston and Diehl, 1994; Hall, 2011). In the present case, higher F0 following voiceless stops (which are primarily cued by longer

VOTs) provides an additional signal to the listener that the preceding stop should be interpreted as voiceless. However, that this correlation only exists in tokens of words with stop voicing minimal pairs suggests that the correlation is not driven by purely automatic phonetic/physiological factors; if it were, we would expect to find that the two cues are correlated in all word tokens, not just those with minimal pairs. Instead, this pattern suggests that speakers are redundantly enhancing the F0 cue *only in cases where the voicing value of the preceding stop is meaningful* (see Hall et al., 2016, for discussion of the importance of meaning in the realization of phonological contrasts). Furthermore, that this apparently controlled effect was only found for voiceless stops, and not for voiced stops, suggests that the pattern that speakers are enhancing in these word tokens is the raising of F0 following voiceless stops relative to some baseline (as opposed to either lowering F0 following voiced stops relative to a baseline or a mutual process of raising F0 following voiceless stops and lowering F0 following voiced stops). This was taken as support for the idea that the effect of phonological voicing on following F0 takes the form of higher F0 for voiceless stops associated with the suppression of voicing, which can be enhanced beyond what is physiologically necessary to suppress voicing during production of the stop (Hoole and Honda, 2011).

4.3 Interpreting Results

In the two studies conducted in this dissertation, I sought to clarify the triggers (or "inputs") and outcomes (or "outputs") of contrastive hyperarticulation and to reduce the hypothesis space for future investigations into these and related phenomena. To this end, the results of both studies should be taken as suggestions for more targeted future investigations as opposed to clear or definitive tests of a single specific hypothesis. Nonetheless, the results of these studies shed substantial light on the phenomenon of contrastive hyperarticulation and some of the ways that it may differ from other, related phenomena.

One important outcome of this research is the finding that contrastive hyperarticulation is most reliably predicted by measures of lexical competition that are defined in terms of the specific phonetic contrast in question. In Study 1 (Chapter 2), it was found that the existence of the single lexical-phonological minimal pair competitor defined primarily by the VOT contrast was correlated with contrastive hyperarticulation of VOT for both voiced and voiceless stops. Of note, this hyperarticulation was truly contrastive, taking the form of longer VOTs in voiceless stops and shorter VOTs in voiced stops. This lexical competition measure was found to be a better predictor of VOT realizations for both stop types than a number of other lexical competition measures in almost all cases. This finding was considered as evidence of contrastive hyperarticulation, in particular because the effect went in opposite directions for the two stop types, which is not easily explained by other kinds of hyperarticulation. The finding also supports a number of past studies suggesting that minimal pair competitor existence correlates with contrastive hyperarticulation of primary phonetic cues to the relevant contrast (Baese-Berk and Goldrick, 2009; Peramunage et al., 2011; Schertz, 2013; Buz et al., 2016; Seyfarth et al., 2016; Wedel et al., 2018).

Another important outcome of this research is the finding that not all phonetic cues are created equal in the eyes of contrastive hyperarticulation. While prior research has shown that multiple phonetic cues to a contrast can be manipulated contrastively by speakers of English (Seyfarth et al., 2016), the relative importance of a given phonetic cue may be critically important to whether or how that cue is contrastively hyperarticulated. Specifically, although it was shown in Study 1 (Chapter 2) that VOT, the primary cue to word-initial stop voicing contrasts in English, is contrastively hyperarticulated in both voiced and voiceless stops, it was also shown in Study 2 (Chapter 3) that F0, another cue to word-initial stop voicing contrasts in English, is not contrastively hyperarticulated *independently* of VOT hyperarticulation. However, it was found that F0 is positively correlated with VOT for voiceless stops *with voiced stop minimal pairs only*, suggesting that F0 is enhanced alongside VOT in precisely those cases where the stop voicing contrast is most critical. This was interpreted as limited evidence of contrastive hyperarticulation, but with the understanding that cues of lesser importance are only enhanced in more limited contexts.

One final outcome of this research is the finding that small methodological differences may have considerable effects on the results of our studies in this domain. As an example, it was found in Study 1 (Chapter 2) that minimal pair competitor existence was a better predictor of VOT realizations than either phonological neighborhood density or the number of onset competitors (among other lexical competition metrics). By contrast, in a similar study that also made use of the Buckeye Corpus, Fricke (2013) found that the number of onset competitors was a better predictor of VOT than either neighborhood density or minimal pair competitor existence. It was suggested that these differences may be due to automatic versus manual annotation of VOT, inclusion of only one-syllable words versus both one- and two-syllable words, and/or merging versus maintaining the /ɑ ~ ɔ/ vowel contrast. If these methodological differences are responsible for differences in results, this highlights the need for standardized practices that are as generalizable as possible.

Collectively, the results of Studies 1 and 2 suggest that phonetic cues to lexical-phonological contrasts are contrastively hyperarticulated in tokens of words that have minimal pair competitors that are defined by the phonetic cues in question. However, while some cues are contrastively hyperarticulated in relatively straight-forward terms (as was seen with VOT in Chapter 2), the realization of other cues can be contingent on primary cue hyperarticulation (as was seen with F0 in Chapter 3). It is possible that these differences in how cues are contrastively hyperarticulated may be related to the relative weight or importance assigned to each cue in perception. In the case of VOT and F0, there is substantial evidence that both speakers and listeners assign more weight to VOT than to other cues, including F0 (e.g., Clayards, 2018). This difference in relative cue weight may explain why F0 is not contrastively hyperarticulated independently of VOT hyperarticulation. By comparison, Seyfarth et al. (2016) found that both preceding vowel duration and voicing of a coda fricative were contrastively hyperarticulated in words with final fricative voicing minimal pair competitors. If the relative weight assigned to these two cues is more balanced than the relative weight assigned to VOT and F0, this may help to explain why the hyperarticulation of F0 found in Study 2 is so much more restricted than the hyperarticulation of cues found in their study.

It is also possible, however, that the differences in how cues are contrastively hyperarticulated are attributable to the range of signals that each cue is involved in. For example, VOT is generally used in English as a cue to voicing distinctions in a variety of contexts, alongside other cues such as preceding or following vowel durations and F0 realizations. As a durational measure, VOT could also be considered a potential cue to speech rate, at least relative to other segment durations, but otherwise VOT is largely constrained to signaling voicing contrasts. F0, by contrast, is not only implicated as a cue to voicing contrasts, but is heavily recruited in English to signal intonation, prosody, and prominence. As such, it is possible that F0 as a phonetic signal is "spread too thin" to be a robust and independent cue to voicing contrasts in English, instead surfacing as a contrastive cue in more limited contexts than the otherwise more dedicated cue of VOT.

4.4 Next Steps

The studies undertaken in this dissertation provide a number of answers to questions surrounding contrastive hyperarticulation. However, a number of outstanding questions remain, and provide promising avenues for future research.

4.4.1 Replication and confirmation

The most notable next step is to confirm the findings of the post-hoc investigation from Study 2 (section 3.7). As a post-hoc analysis, it is important to test whether the positive correlation between VOT and F0 for voiceless stops with minimal pairs, and only voiceless stops with minimal pairs, is reproducible in another corpus of spontaneous speech. The finding that minimal pairs are contrastively hyperarticulated even in spontaneous speech, where the minimal pair competitor is unlikely to co-occur (Wedel et al., 2018), is a critical link in understanding how contrastive hyperarticulation affects long term language change. Wedel et al. (2013b) found in a longitudinal cross-language corpus study that phoneme contrasts with few lexical minimal pairs are more likely to undergo merger and be lost from the language. Contrastive hyperarticulation of precisely these lexical minimal pairs in natural speech provides a synchronic mechanism whereby phoneme contrasts represented by larger numbers of lexical minimal pairs are realized in ways that enhance the phonemic contrast and thereby minimize the chances of merger. If the findings of this dissertation cannot be replicated in another corpus of spontaneous English, that would severely restrict the power of these results and call into question the role of contrastive hyperarticulation in language change.

In addition to confirming these findings using other corpora of spontaneous English speech, it is also important to replicate these findings in corpora of spontaneous speech in other languages. Such cross-linguistic confirmation is crucial to demonstrating that contrastive hyperarticulation plays a role in the development of language in general, and not just in English. Of particular interest, studying other languages would further allow for a test of the hypothesis that contrastive hyperarticulation of a phonetic cue is partly dependent on the range of signals for which the cue is recruited. As noted above (and in section 3.8), F0 is recruited as a signal to a variety of linguistic and para-linguistic phenomena, such as intonation, prosody, prominence, and preceding obstruent voicing. It was suggested that this wide array of uses may restrict the availability of F0 as a candidate for hyperarticulation. This hypothesis can be tested by looking for contrastive hyperarticulation of F0 in languages such as Korean or Afrikaans, where the F0 cue has become the dominant cue to preceding stop voicing contrasts for at least some speakers/contexts (Bang et al., 2018; Coetzee et al., 2018). In such languages, it might be expected that, as a primary cue, F0 is contrastively hyperarticulated for both voiced and voiceless stops as the VOT cue is in English (Chapter 2).

Another important next step is to replicate these studies in a controlled, laboratory environment. Although there is some reason to assume that various hyperarticulation phenomena may be partly dependent on speech style (see Wedel et al., 2018, for some limited discussion), the spontaneous speech data analyzed in this dissertation required substantial statistical controls that could be avoided in more controlled contexts. Replicating these findings in such contexts is critical to understanding the validity of the findings[2].

[2]There are substantial data already gathered by other researchers in which contrastive hyperarticulation of VOT was found for voiceless stops with voiced stop minimal pairs, providing partial validation of the results

4.4.2 Languages with few lexical minimal pairs

An additional step to take in this research is to investigate what role, if any, contrastive hyperarticulation plays in languages that have relatively few minimal pairs as traditionally defined. Many languages are more agglutinative or polysynthetic than English and other Indo-European languages. Consequently, there are fewer opportunities for true lexical minimal pairs. Consider, for example, the Hiaki language, spoken in various parts of the Sonoran desert in North America. Words in Hiaki are often built through processes of successive derivation, where, for example, lexical roots may be verbalized, participialized, nominalized, and verbalized again. Such a final word form is unlikely to have a lexical minimal pair that differs in only one phoneme of the word. To what extent can contrastive hyperarticulation help explain patterns of language change in such a language, given that we have established a critical role for lexical minimal pairs in the realization of contrastive hyperarticulation? One possibility is related to the use of lemmas in defining our "lexical" minimal pair relationships. Although processes of derivation in English derive new lemmas, the concept of a lemma-based lexicon could potentially be extended to a lexicon of roots, which may prove useful in languages such as Hiaki, where surface lexical forms will include few minimal pairs, but lexical roots may involve more. This extension of the concept of minimal pair becomes perhaps even more important in some other languages, such as Navajo or Georgian, where many lexical roots are *never* produced as surface word forms. Does it make sense to consider lexcial-phonological competition between theorized units of speech that are never realized on the surface? Consider also Semitic languages that exhibit so-called "templatic" morphology, where consonantal roots are inserted into morphological templates specifying vowels and syllable structure. How should we define the concept of lexical minimal pairs in such a language, and can we expect contrastive hyperarticulation to be extent at all? Additional work in this area is not only important for our understanding of contrastive hyperarticulation as a phenomenon, but also for our un-

of Study 1 (Chapter 2). This data could be reanalyzed to test the post-hoc hypothesis of Study 2 (Chapter 3), namely that F0 is higher following voiceless stops in words with initial stop voicing minimal pairs than it is following voiceless stops in words without such minimal pairs.

derstanding of different language structures and the implications these structures have for the human language lexicon.

4.4.3 Testing key theoretical assumptions

Finally, the theoretical assumption that contrastive hyperarticulation forms a critical step in processes of sound change (as discussed in Wedel et al., 2013b, 2018, among others) is largely rooted in Exemplar Theory (Blevins, 2004; Wedel, 2006; Blevins and Wedel, 2009). In an exemplar model of sound change, perceived tokens of speech production targets (e.g., phonemes or words) are stored in memory for an extended period of time, forming so-called "clouds" of exemplars. In addition to storing perceived tokens in these clouds, the exemplar clouds are also the source that speakers pull from to get their production targets, and speakers' own productions contribute to their exemplar clouds as perceived tokens. This perception-production feedback loop provides an explanatory mechanism for contrastive hyperarticulation as both a synchronic and diachronic phenomenon. A common assumption in such models is that there are language-internal pressures to keep these clouds distinct. If two exemplar clouds become too similar, speakers select targets from the outer edges of the clouds in a sub-conscious effort to maintain the intended contrast. By such a model, this is how contrastive hyperarticulation comes into being, as production targets from exemplar clouds that represent more common contrasts are repeatedly drawn from more extreme exemplars in the cloud, gradually shifting these exemplar clouds farther apart.

One missing link in this explanatory chain is the role of frequency and predictability. Under this exemplar model, higher frequency words, which are necessarily produced and perceived more often, should participate in this perception-production feedback loop more often than lower frequency words. Presumably, this disparity in frequency of participation in the loop should have an impact on the rate at which the loop results in sound change. This has already been demonstrated for word durations in New Zealand English, where changes in informativity or frequency of occurrence of words (e.g., becoming more frequent overall, occurring more often utterance-finally, or becoming more informative) lead to compounding changes in word durations over time (Sóskuthy and Hay, 2017). However, in the studies reported here, as well as in Wedel et al. (2018), lexical frequency (or, as appropriate, contextual diversity) did not seem to significantly impact the minimal pair effect (with the single exception of frequency-weighted neighborhood density affecting voiceless stop VOT realizations in Chapter 2 of this dissertation). Additional targeted research on the effects of lexical frequency and informativity on the existence or size of contrastive hyperarticulation effects should be undertaken to confirm that contrastive hyperarticulation participates in these patterns of sound change through the perception-production loop.

4.5 Conclusion

In this dissertation, I set out to clarify the nature of contrastive hyperarticulation. To this end, I conducted two studies investigating (i) the lexical-phonological *triggers* of contrastive hyperarticulation and (ii) the phonetic *outcomes* of contrastive hyperarticulation. Findings supported the hypothesis that the existence of the lexical-phonological competitor that is primarily defined by the specific phonetic cue under investigation correlates with enhancement of that phonetic cue. Indeed, evidence was found to support the idea that even a secondary cue to a lexical-phonological contrast can be hyperarticulated in ways that enhance that contrast. Evidence in favor of this view was found in the form of (i) better model fits to VOT data when the existence of the initial stop voicing minimal pair competitor is included in the model, relative to a large array of alternative competition metrics, and (ii) a significant positive correlation between VOT and F0 *only in tokens of voiceless stop-initial lemmas with voiced stop minimal pair competitors in the lexicon*. I argued that these findings reflect contrastive hyperarticulation of a pair of phonetic cues that play very different roles in signaling the voicing contrast.

Appendix A

Study 1: Additional model summaries

Variable	β	SE	t
(Intercept)	0.558	0.015	37.13
Overall Neighb. Freq.	**-0.043**	**0.026**	**-1.68**
Stop Phoneme (p-t)	-0.132	0.008	-15.71
Stop Phoneme (k-t)	-0.013	0.008	-1.64
Speech Rate	-0.030	0.016	-1.92
Following Liquid (true-false)	0.049	0.007	6.66
Word Familiarity	-0.018	0.014	-1.33
Phonotactic Probability	-0.015	0.037	-0.40
Forward Probability	0.006	0.009	0.68
Backward Probability	-0.015	0.009	-1.65
Number of Syllables (2-1)	0.009	0.006	1.41
Previous Mention (true-false)	0.001	0.004	0.20
Syntactic Category (Noun-Adjective)	-0.003	0.011	-0.25
Syntactic Category (Adverb-Adjective)	-0.001	0.025	-0.04
Syntactic Category (Verb-Adjective)	-0.014	0.012	-1.12

(a) Fixed effects summary.

Group	Variable	Variance	SD	Corr
Lemma	(Intercept)	0.0010	0.032	
Speaker	(Intercept)	0.0013	0.037	
	Overall Neighb. Freq.	**0.0014**	**0.037**	**-0.95**
Residual	0.0103	0.102		

(b) Random effects summary.

Table A.1: Fixed and random effects summaries for the fitted model including overall neighborhood frequency for voiceless stops when words with no single-phoneme edit distance neighbors are removed.

Appendix B

Study 2: Additional model summaries

Variable	β	SE	t	$p(\|t\|)$
(Intercept)	1.113	0.165	6.747	< 0.001***
High Vowel (true-false)	-0.040	0.102	-0.392	0.696
Following Liquid (true-false)	0.034	0.056	0.597	0.0566
Number of Syllables (1-2)	0.082	0.061	1.340	0.181
Word Familiarity	-0.013	0.020	-0.617	0.543
Backward Probability	-0.003	0.023	-0.128	0.899
Word Fam : Back Prob	-0.003	0.007	-0.512	0.609
VOT/Stop-length ratio	**0.112**	**0.036**	**3.130**	$<$ **0.01****

(a) Fixed effects summary

Group	Variable	Variance	SD
Lemma	(Intercept)	0.002	0.049
Speaker	(Intercept)	0.292	0.541
Residual		0.862	0.928

(b) Random effects summary

Table B.1: Fixed and random effects summaries for the fitted model predicting average-normalized initial pitch (in semitones) for voiceless stops with minimal pairs only, as a function of the VOT/Stop-length ratio. Note that the minimal pair and the following glide factors had to be removed from the previously-fit model specification because this subset of data does not include contrasts for those factors.

Variable	β	SE	t	$p(\|t\|)$
(Intercept)	1.256	0.215	5.833	< 0.001***
High Vowel (true-false)	-0.015	0.153	-0.096	0.924
Following Liquid (true-false)	0.139	0.151	0.924	0.357
Number of Syllables (1-2)	0.080	0.068	1.190	0.235
Word Familiarity	-0.038	0.023	-1.637	0.115
Backward Probability	0.005	0.025	0.218	0.828
HighVow : FollLiquid	0.084	0.148	0.569	0.570
WordFam : BackProb	-0.004	0.008	-0.480	0.632
VOT/Stop-length Ratio	**0.152**	**0.041**	**3.682**	$<$ **0.01****

(a) Fixed effects summary

Group	Variable	Variance	SD	Corr
Lemma	(Intercept)	0.007	0.082	
Speaker	(Intercept)	0.417	0.646	
	VOT/Stop-length ratio	0.003	0.050	0.24
Residual		1.039	1.019	

(b) Random effects summary

Table B.2: Fixed and random effects summaries for the fitted model predicting midpoint-normalized initial pitch (in semitones) for voiceless stops with minimal pairs only, as a function of the VOT/Stop-length ratio. Note that the minimal pair and the following glide factors had to be removed from the previously-fit model specification because this subset of data does not include contrasts for those factors.

Appendix C

Previously published work

REMAINDER OF THIS PAGE INTENTIONALLY LEFT BLANK

The phonetic specificity of competition: Contrastive hyperarticulation of voice onset time in conversational English

Noah Richard Nelson[a]*, Andrew Wedel[a]

*Corresponding author: NoahNelson@email.arizona.edu, Tel: +1 520 621 6897
[a] Department of Linguistics, University of Arizona, P.O. Box 210025, Tucson, AZ, 85721, USA

Abstract

Competition between words in the lexicon is associated with hyperarticulation of phonetic properties in production. This correlation has been reported for metrics of competition varying in the phonetic specificity of the relationship between target and competitor (e.g., neighborhood density, onset competition, cue-specific minimal pairs). Sampling a systematic array of competition metrics, we tested their ability to predict voice onset times in both voiced and voiceless word-initial stops of conversational English. Linear mixed effects models were compared according to their corrected Akaike's Information Criterion (AIC_c) values. High-performing models were evaluated using evidence ratios, with the competition metrics of top-performing models tested for significance using nested model comparisons. Words with a minimal pair defined for initial stop voicing were contrastively hyperarticulated, with shorter voice onset times for voiced stops and longer voice onset times for voiceless stops. No other competition metric reliably predicted hyperarticulation for both stop types. These results suggest that contrastive hyperarticulation is phonetically specific, increasing the perceptual distance between target and competitor.

Keywords

Competition, hyperarticulation, voice onset time, conversational speech, minimal pairs, neighborhood density

1 Introduction

A number of experimental and observational studies have reported that competition at the lexical level is correlated with hyperarticulation of phonetic properties in the target word. This correlation has been reported for English in a number of studies investigating the realization of vowel formants (e.g., Wright, 1997; 2004; Munson & Solomon, 2004; Munson, 2007; Scarborough, 2012), vowel durations (Schertz, 2013; Seyfarth, Buz, & Jaeger, 2016; but see Goldrick, Vaughn, & Murphy, 2013), degree of coarticulation (Scarborough, 2012, 2013), perseveration of voicing in coda fricatives (Seyfarth et al., 2016; Kharlamov, 2014), and initial stop voice onset time (Baese-Berk & Goldrick, 2009; Peramunage et al., 2011; Kirov & Wilson, 2012; Schertz, 2013; Fricke, 2013; Buz, Tanenhaus, & Jaeger, 2016; Fox, Reilly, & Blumstein, 2015; Fricke, Baese-Berk, & Goldrick, 2016). In each of these cases, some form of lexical competition has been found to correlate with hyperarticulation of phonetic properties of individual segments (e.g., Wright, 2004; Fricke, 2013; Buz et al., 2016; but see Goldrick et al., 2013; Gahl, 2015).

This work on competition-associated hyperarticulation is complicated, however, by the fact that 'competition' can be operationalized in a variety of ways. The two most common approaches to operationalizing competition are in terms of lexical-phonological *neighborhood density*, defined as the number of words that can be formed by adding, deleting, or substituting any single segment of the target word (often weighted for frequency; e.g., Luce & Pisoni, 1998), or in terms of *minimal pair relationships* defined for a specific phonetic cue (e.g., Baese-Berk & Goldrick, 2009). These two general approaches differ widely in terms of the specificity of the competition measure. For neighborhood density measures, competition anywhere in the word contributes to the measure, regardless of the phonetic relationship between the target and the competitor (though see Strand & Sommers, 2011, and Gahl & Strand, 2016, for a version of neighborhood density weighted according to perceptual similarity). On the other hand, the cue-specific minimal pair measure identifies the neighbor that differs from the target solely in the cue of interest. Consequently, we can think of the cue-specific minimal pair measure as being more phonetically specific than the neighborhood density measure.

Given the differences in the way these two competition metrics are defined, there is a great deal of intermediate ground between them. One way to think about this intermediate ground is in terms of a continuum of specificity in lexical competition (but see section 5.1 for discussion of alternatives). For simplicity, we conceptualize this continuum relative to a given phonetic cue in terms of two parameters that define lexical neighborhoods of varying specificity: the relative *type* of competition between target and competitor, and the relative *position* of that competition (Fig. 1). The most specific type and position on this continuum is defined by a cue-specific minimal pair competitor, which differs from the target word only in the measured cue, in the same segmental position that the cue is realized. For example, voice onset time is the primary cue distinguishing word-initial stop voicing in English (Lisker & Abramson, 1964; Lisker, 1986). Given the reference word *bill*, *pill* is a neighbor differing in the same cue (voice onset time) in the same segmental position. We can now instead define a more relaxed neighborhood for neighbors that share, for example, a manner of articulation in the same segment as the measured cue (e.g., *pill, kill*, *dill*…). Similarly, we can define the neighborhood to include any onset-competitor (e.g., *pill*, *will*, *mill*…). These two examples vary the phonetic *type* of competition, but hold the *position* of competition steady. We can also vary the relative position of competition. As an example, we can consider only those neighbors that differ in the second position of the word; given the target word *bill*, neighbors differing only in the second segment include such words as *bell*, *bowl*, and *ball*.

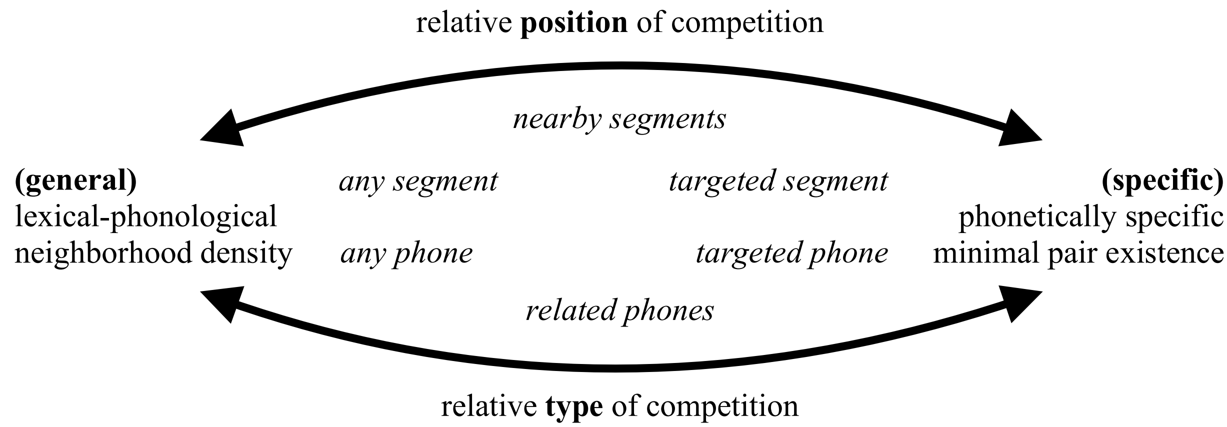

Fig. 1. Schematization of a continuum of specificity in competition. Single-edit competition metrics can lie at various positions along this continuum according to the relative type or position of competition with the target, in terms of neither (neighborhood density), or in terms of both (phonetically specific minimal pair competitor existence).

The present study compares a number of lexical competition metrics based on this continuum of specificity. Based on a systematic sample of alternatives from this continuum, we used mixed effects regression to predict the realization of voice onset time in a corpus of conversational English speech. We tested this sample of alternative metrics separately for both voiced and voiceless word-initial stops. While we compared a substantial set of competing hypotheses, we note that this range of metrics does not exhaust the set of plausible ways of operationalizing lexical competition. Indeed, a number of alternative metrics can be found in the literature, including neighborhood measures weighted for (i) the proportion of segmental overlap (Goldrick, Folk, & Rapp, 2010), (ii) variable Levenshtein edit distance between target and competitors (Yarkoni, Balota, & Yap, 2008), or (iii) perceptual similarity (Strand & Sommers, 2011; Gahl & Strand, 2016). Our results are considered with regard to these alternatives in the discussion.

This paper takes the following course. First, we briefly review cognitive mechanisms that have been proposed to underlie competition-associated hyperarticulation. Next we review prior research on competition-associated hyperarticulation, and the implications for each of these cognitive mechanisms. We then present results of analyses comparing a range of competition metrics for their ability to predict voice onset time in conversational speech. Finally, we discuss the implications of these results for the proposed cognitive mechanisms behind competition-associated hyperarticulation and for diachronic sound-change.

1.1 Accounts of competition-associated hyperarticulation

Why does competition among lexical items lead to hyperarticulation? One proposal appeals to processes of lexical selection and/or speech planning during production. In these production-internal approaches, lexical items compete for activation (e.g., Dell, 1986; see Fricke, 2013, ch. 2, for a review). This competition leads to overall higher activation levels for the target word, which in turn leads to enhancement of articulatory gestures in production (see Baese-Berk

& Goldrick, 2009, and Fricke, 2013, chs. 2 & 6, for discussion). A common implication of this hypothesis is that competition should only result in *increases* in phonetic durations such as voice onset time. Further, these accounts generally assume that competition-associated hyperarticulation should be relatively insensitive to the precise phonetic relationship between target and competitor (e.g., Baese-Berk & Goldrick, 2009; Fricke, 2013; Watson et al., 2015; see Jaeger & Buz, 2016, and Buz & Jaeger, 2016, for reviews).

Two alternative proposals predict that competition can result in targeted hyperarticulation of cues that distinguish a word from a specific competitor. According to communicative accounts (reviewed in Jaeger & Buz, 2016), speakers are implicitly aware that some words are more confusable than others, and are able to hyperarticulate the cues that maximally differentiate a given word from plausible alternatives. According to this approach, hyperarticulation is a tool that can be employed *online* by speakers in order to promote communicative efficiency. Over time, this online hyperarticulation can result in shifts in long-term lexical representations (reviewed in Hall et al., submitted). In a listener-internal approach, however, *hyper*-articulated tokens of potentially ambiguous words are more likely to be recognized, and therefore stored in memory, than *hypo*-articulated tokens. As a result, more ambiguous tokens contribute less to the exemplar cloud representing a given word or category, shifting it toward the hyperarticulated variant (see, e.g., Wedel 2006). Because this same exemplar cloud serves as the source for future productions, this process can, over time, lead to notable differences in the phonetic forms of words. Such an account does not predict that hyperarticulation happens online, but that, through filtering by listeners, words should come to be articulated in a way that enhances the specific phonetic contrasts that distinguish them. It is worth noting that all of these proposed mechanisms operate at different levels, and are therefore mutually compatible. We consider it a priori possible that all of these mechanisms contribute to the range of hyperarticulation effects found in human language.

1.2 The phonetic specificity of competition

The three cognitive mechanisms outlined above are rooted in very different assumptions about the nature of lexical competition. One critical question is how sensitive competition is to the phonetic or phonological relationship between target and competitor. According to most production-internal accounts, competition is relatively abstract, occurring among phonological representations prior to phonetic encoding (e.g., Baese-Berk & Goldrick, 2009; see also Goldrick & Rapp, 2007). For example, Fricke's (2013) Articulate As Soon As Possible Principle (AASAPP) posits that articulation unfolds on a segment-by-segment basis, and begins as soon as competition in a given segmental position is resolved. According to this model, the relative position of difference between the target and its competitors is therefore of paramount importance, but the phonetic relationships among those competitors is not relevant – a greater number of competitors defined for a particular segmental position leads to increased activation of the target segment regardless of phonetic relationship (Fricke, 2013; see also Vitevitch, Ambrüster, & Chu, 2004; Fricke et al., 2016). In contrast, both communicative and listener-internal accounts are based on perceptual confusability, such that hyperarticulation is targeted to those cues that maximize perceptual distinctiveness between competitors.

This question of how sensitive competition is to phonetic relationships is still open. In the case of word-initial stops, the focus of this paper, a number of researchers have argued that voice onset time is hyperarticulated in response to phonetically specific minimal pair competitors. This correlation has been reported for voiceless stops (Baese-Berk & Goldrick, 2009; Peramunage et al., 2011; Kirov & Wilson, 2012; Schertz, 2013; Buz et al., 2016), but also for voiced stops, for which voice onset times have been found to *decrease* (Schertz, 2013; Wedel, Nelson, & Sharp, submitted; but see Ohala 1994, and Goldrick et al., 2013). Similar results have also been found for contrastive cues to coda voicing (vowel length and perseveration of voicing in English fricatives: Seyfarth et al., 2016; duration of fricatives and glottal pulsing in both stops and fricatives in Russian: Kharlamov, 2014). These results suggest that competition-driven hyperarticulation is *contrastive*, increasing the phonetic distance between target and competitor, and may be quite phonetically specific (e.g., Schertz, 2013; Seyfarth et al., 2016). Furthermore, Baese-Berk and Goldrick (2009) found a smaller, but significant effect of the mere lexical existence of the phonetically specific competitor, despite not being present in the experimental context (see also Peramunage et al., 2011). This additional result suggests that these effects may be partially mediated by processes that are not dependent on the presence of the competitor in the immediate context.

However, other researchers have reported evidence that this kind of minimal pair competition does not correlate with contrastive hyperarticulation. Goldrick, Vaughn, and Murphy (2013) used a word list reading task to elicit productions of voiced stop-initial words as well as voiced and voiceless stop-final words. For some of these words, the minimal pair defined for the voicing of their stop consonant existed (e.g., *bun ~ pun*, *coat ~ code*) and for others it did not (e.g., *bum ~ *pum*, *thud ~ *thut*). They found no significant effect of initial stop voicing minimal pair existence on the realization of voice onset time in voiced stops, nor did they find a significant effect of final stop voicing minimal pair existence on the realization of any cues to final voicing for voiceless stops (vowel duration, stop release, closure duration, or perseveration of voicing). They did find that final stop voicing minimal pair existence correlated significantly with vowel duration for final voiced stops, but the effect was opposite what would be expected under contrastive hyperarticulation. Rather than getting longer, vowels before voiced stops in words with minimal pair competitors in coda voicing were *shorter* than those in words without such a minimal pair, indicating a reduction of the contrast with the minimal pair competitor.

Finally, a number of researchers have argued that hyperarticulation results from more general competition. Fox et al. (2015) reported that phonetically specific minimal pair existence did not significantly predict voice onset time realizations in word list and sentence reading tasks, while neighborhood density did. Fricke (2013) and Fricke, Baese-Berk, and Goldrick (2016) reported that hyperarticulation of word-initial voice onset time is most robustly predicted by the number of competitors in the onset position. Fricke and colleagues argued that both cue-specific minimal pair competition and more general neighborhood density correlate with hyperarticulation of word-initial voice onset time because they *also* correlate with this position-based measure (see also Caselli, Caselli, & Cohen-Goldberg, 2015, and Vitevitch et al., 2004). Similarly, Kirov and Wilson (2012) reported that hyperarticulation of voice onset time in word-initial stops correlated with the presence of a minimal pair competitor for either stop voicing (e.g., *cap ~ gap*) or place of articulation (e.g., *cap ~ tap*), both in the initial segment. They further found that positions other than the initial segment, i.e. the vowel or coda of CVC monosyllables, did not affect initial stop voice onset time realizations (e.g., *cat ~ kit*; *cat ~ cap*). In a related study, Schertz (2013) reported contrastive voice onset time hyperarticulation in both voiced and voiceless word-initial stops associated with the presence of a voicing competitor. However, she found no effect of either place or manner of articulation competitors.

In summary, lexical competition, broadly construed, has been found to be consistently associated with contrastive hyperarticulation of voice onset time, but the nature of the associated competition has varied substantially across studies. One possible explanation for these differences is methodological. The majority of studies exploring contrastive hyperarticulation have used laboratory elicitation of defined materials in order to control for the many factors correlated with phonetic realization. Depending on experimental conditions, elicited speech may be prone to particularly clear or slow articulations that have been associated with increased phonetic durations and exaggerated articulatory gestures (e.g., Picheny et al., 1986; de Jong et al., 1993; Smiljanić & Bradlow, 2008). This "clear speech" may represent a conceptually distinct form of hyperarticulation from competition-driven contrastive hyperarticulation of the type reviewed above (for similar terminology, see Ohala, 1994; Seyfarth et al., 2016; for other discussions of different kinds of hyperarticulation, see Cho, Lee, & Kim, 2011; Schertz, 2013). We refer to this kind of generalized hyperarticulation as *clear-speech hyperarticulation*, which has been associated with increased voice onset times, particularly in voiceless stops (Smiljanić & Bradlow, 2008). As most studies examining the effects of competition on the realization of voice onset time have focused on voiceless stops (with the exceptions of Ohala, 1994, Schertz, 2013, and Goldrick et al., 2013), the predictions of both clear-speech and contrastive hyperarticulation have coincided. In these studies, clear-speech hyperarticulation may raise voice onset times toward a ceiling that makes contrastive hyperarticulation effects more difficult to detect (Kirov & Wilson, 2012; see also Wedel et al., submitted). Indeed, most of the studies that have found evidence of contrastive hyperarticulation have used task environments designed to amplify these potential effects, for example by including the minimal pair competitor in the immediate context (Baese-Berk & Goldrick, 2009, study 2; Kirov & Wilson, 2012; Seyfarth et al., 2016; Buz et al., 2016; but see Baese-Berk & Goldrick, 2009, study 1; Peramunage et al., 2011), or by explicitly indicating that the speaker's production was misperceived as the minimal pair competitor (Schertz, 2013; Buz et al., 2016).

One approach to minimize the influence of clear-speech hyperarticulation is to study speech in contexts promoting greater reduction, such as conversation (see, e.g., Gahl, Yao, & Johnson, 2012). A complementary strategy is to study phonetic cues for which the predictions of clear-speech and contrastive hyperarticulation diverge, such as in word-initial voiced stops (see Ohala, 1994, Schertz, 2013, and Goldrick et al., 2013). Contrastive hyperarticulation of voiced stops should lead to *shorter* voice onset times relative to non-hyperarticulated words (Schertz, 2013), while clear-speech conditions are not associated with shorter voice onset times (e.g., Miller, Green, & Reeves, 1986; Kessinger & Blumstein, 1997). Here, we study both voiced and voiceless stops in conversational speech, because contrastive hyperarticulation should move voice onset times in opposite directions for these two stop types (e.g., Schertz, 2013).

1.3 The present study

The present study was designed to clarify the specificity of the relationship among targets and competitors that leads to hyperarticulation. To this end, we compared a sample of competition measures from the continuum of specificity (Fig. 1) for their ability to predict the realizations of word-initial stop voice onset times in conversational English. In particular, we explored (i) the role of segmental position and (ii) the role of the phonetic relationship between segments within a given position. Though these two dimensions do not represent all possible hypotheses about how lexical competition should be operationalized, both of these dimensions have important theoretical consequences for accounts of competition effects. For example, Fricke's (2013) AASAPP, among other production-internal accounts, predicts that hyperarticulation is targeted only to specific segmental positions, and can only lead to increased durations. Both trade-off and perception-based approaches, on the other hand, predict that hyperarticulation is targeted to specific cues and is contrastive, i.e., can be realized as either increased or decreased durations. Here, we look for contrastive hyperarticulation in the form of both longer voice onset times in voiceless stops, and shorter voice onset times in voiced stops. We investigate a sample of competing hypotheses regarding the relationship between target and competitor that gives rise to hyperarticulation, looking for these effects in conversational speech, where reduction is more likely to occur. Our study must therefore be considered alongside additional research testing alternative hypotheses, as well as investigating these phenomena in elicited speech paradigms.

2 Methods

We examined the effects of competition on the realization of voice onset time in conversational English based on a sampling of different competition metrics. These metrics took the form of modified neighborhood densities designed to sample two dimensions of relatedness in competition characterizing the differences between overall neighborhood density and cue-specific minimal pair competition. These modified neighborhood density measures targeted either the relative *position* of competition between target and competitor (i.e., *where* in the word the two words differ) or the relative *type* of competition between target and competitor (i.e., *how* the two words differ). We then used these metrics to predict voice onset time realizations of voiced and voiceless stops in the Buckeye Corpus of Conversational Speech (Pitt et al., 2005; Pitt et al., 2007). The use of conversational speech allows us to study the effects of competition in a generally more reduced context, limiting the potential for clear-speech hyperarticulation to mask the effects of contrastive hyperarticulation. The hypothesis that competition drives contrastive hyperarticulation predicts opposite effects for voiced versus voiceless stops: for voiced stops, voice onset times should decrease, while for voiceless stops, voice onset times should increase (Schertz, 2013). Including both stop types provides two distinct tests of this hypothesis.

We used linear mixed effects models to analyze the predictive relationship between competition and the realization of voice onset time. Due to the large number of hypotheses and the exploratory nature of the study, we evaluated these models using corrected Akaike's Information Criterion (AIC_c) comparisons and evidence ratios (Burnham & Anderson, 2004; Lukacs et al., 2007; Richards et al., 2011). This approach does not test for significance as in null hypothesis testing approaches to model comparison such as log-likelihood ratio tests, making it better suited to multiple hypothesis testing, in which multiple comparisons increase the risk of family-wise error (see Chamberlain, 1890, and Shadish, 1993, for discussion of the merits of testing multiple working hypotheses). In addition, evidence ratios allow for quantified statements about the relative support in favor of one model over another, ideal for comparing competing hypotheses (Burnham & Anderson, 2004; Richards et al., 2011). After comparing the models using AIC_c, we tested the top-performing model for both voiced and voiceless stops to ensure that the competition measure was contributing significantly to model fit using log-likelihood ratio tests of nested models. If the competition measure in the top-performing model does not contribute significantly to model fit, the relative ranking of models is not clearly interpretable.

2.1 Materials

2.1.1 The data source

We used natural speech data from the Buckeye Corpus of Conversational Speech (Pitt et al., 2005; Pitt et al., 2007). The Buckeye Corpus includes 40 hours of conversation, spread over 1-hour long interviews with 40 individuals. Half of the interviewees are male, and half are female; half are over the age of 40 and half are under the age of 30. Interviews were conducted in Columbus, Ohio, and all speakers are from Columbus or surrounding regions of Ohio. The Buckeye Corpus is annotated for utterances as well as words and their syntactic category, and includes phonetic transcriptions with segment-level durations, but no sub-segmental measurements such as voice onset time.

2.1.2 Voice onset time measurements

We used measurements of the voice onset times of voiced and voiceless stops from a balanced set of 24 out of the 40 speakers in the corpus. Measurements were made by hand for stop-initial content words (labeled in the corpus as noun, verb, adjective, or adverb) of one or two syllables produced by these speakers. Our study thus differs from much previous work on competition effects in that our data includes two-syllable words. In addition, unlike in many other studies where only CV-initial words were included (e.g., Baese-Berk & Goldrick, 2009; Kirov & Wilson, 2012; Schertz, 2013; Fricke, 2013; Fox et al., 2015), we expanded our word types to include complex onsets. Because of the phonotactics of English, this meant the inclusion of words with an initial stop followed by a liquid or a glide.

Each word was annotated by hand in Praat (Boersma & Weenink, 2013, ver. 5.3.39) for the beginning of the stop closure, the beginning of the burst, and the beginning of the following sonorant. Based on those measurements, the total stop duration, the closure duration, and the voice onset time were calculated. We excluded tokens with pre-voicing, tokens with no identifiable burst, and tokens with closure durations or voice onset times that were more than 3 standard deviations from the specific speaker's mean for that stop consonant. For the remaining tokens, we calculated the proportion of the total stop duration that was taken up by the voice onset time (VOT). Henceforth, we refer to this measurement as the "VOT-length ratio." We elected to use the VOT-length ratio rather than raw voice onset times because it provides a very local control for speech rate. Speech rate is highly correlated with voice onset time in voiceless stops of English (Kessinger & Blumstein, 1998; Yao, 2007), but this correlation is largely attributable to an effect on the entire stop duration. However, an increase or decrease in the VOT-length ratio necessarily reflects a change in the voice onset time independently of any change to the total stop duration (see Smiljanić & Bradlow, 2008, for use of this and a related proportional measure of voice onset time).

We excluded high frequency discourse markers and any content words homophonous with function words (Bell et al., 2009; Gahl et al., 2012; Seyfarth, 2014). Because we based our neighborhood measures on the lemma forms of words (see section 2.1.3 below), we further excluded verbs with stem-vowel changes in their morphology (e.g., *buy ~ bought*, *come ~ came*), as the specific minimal pair competitor lemma is not consistent across the paradigm (e.g., *pie ~ buy* versus *pot ~ bought*). Of the remaining words included in our dataset, tokens were excluded if there was no identifiable burst, if the token was immediately preceded by another stop consonant (due to unreliability of assigning the beginning of the stop closure), if the token followed an annotated pause or disfluency, or if either the aspiration or closure length of the token was more than three standard deviations from the speaker's mean.

2.1.3 Sources for competition metrics

All of the competition metrics used in both studies were calculated based on lemma forms. We used the Corpus Of Contemporary American English (COCA: Davies, 2012) to filter the Carnegie-Melon University pronouncing dictionary (CMU: Carnegie-Melon University, 2015) such that only distinct lemma forms remained. The CMU pronouncing dictionary, however, includes multiple pronunciations of many words, and so in order to prevent the same lemma from counting in any neighborhood measure more than once, we included only the first pronunciation of each distinct lemma. Only unique phonemic forms were retained, with the result that homophonic lemmas (e.g., *bear*, *bare*) correspond to the same entry. Because the CMU pronouncing dictionary is organized by orthographic form, a set of English homographs corresponding to distinct lemmas were manually retained (e.g., <tear> can be either the verb /tɛɹ/ or the noun /tɪɹ/). The full list of these homographs can be found in Appendix A. We retained the /ɑ ~ ɔ/ contrast (Durian, 2012), but collapsed /w ~ ʍ/ (Labov et al., 2006), consistent with pronunciation norms in Central Ohio.

Finally, we filtered this lemmatized CMU pronouncing dictionary based on contextual diversity, excluding any forms appearing in less than 0.5% of films in the SUBTLEX-US database. This was done to reduce the contribution of jargon or uncommon words to the resulting lexicon. Contextual diversity has been shown to be a better predictor of lexical decision accuracy and reaction time than frequency (Brysbaert & New 2009). The 0.5% cutoff results in a lexicon of 11 692 lemmas; for comparison, a frequency cut-off of 1 per million produces a lexicon of similar size (12 811 lemmas). This lexicon served as the source file when calculating all of our competition metrics.

2.2 Competition metrics

We created a systematic sample of competition metrics from the continuum of specificity as conceptualized in Figure 1. This continuum is defined by the two metrics that have dominated the literature on competition-induced hyperarticulation: lexical-phonological neighborhood density and cue-specific minimal pair competitor existence. For comparability with prior work, we included both of these competition measures in our study. Lexical-phonological neighborhood density was defined as the tally of all words that can be derived from the target by adding, deleting, or replacing any single phoneme of the target word (Luce & Pisoni 1998). Henceforth, we will refer to this measure as the 'overall neighborhood density.' Phonetically specific minimal pair competitor existence was coded as a binary neighborhood density. If switching the voicing value of the initial stop consonant of the target lemma resulted in a unique lemma in our lexicon (e.g., *bat* ~ *pat*), then this measure was coded as 1; otherwise, it was coded as 0. We refer to this metric as 'minimal pair competitor existence.'

To explore the space between these measures on the continuum, we used a sample of intermediate neighborhood density measures targeting either the relative position or type of competition. For the relative position of competition, we calculated three modified neighborhood densities, each targeting a different segmental position, or set of segmental positions, within the word (Fig. 2). These three neighborhood densities targeted the first segmental position, the second segmental position, and the rest of the word (from the third segment to the last segment). We chose to include a metric for competition in the onset position because competition in this position has been found to be predictive in a number of studies (Vitevitch et al., 2004; Goldrick et al., 2010; Fricke, 2013; Caselli et al., 2015; Fricke et al., 2016). We included a metric for competition in the second segment because, as stop bursts often contain cues to the formants of following sonorants (Suchato & Punyabukkana 2005), hyperarticulation of bursts might arise when there are many competitors in the following segment. Finally, we included a metric for competition elsewhere in the word to evaluate whether phonetic variation is insensitive to segmental position. Note that all of these neighborhoods include only neighbors formed by substituting a single segment in the relevant position or set of positions; none of the modified neighborhood densities used in this study included additions or deletions of segments.

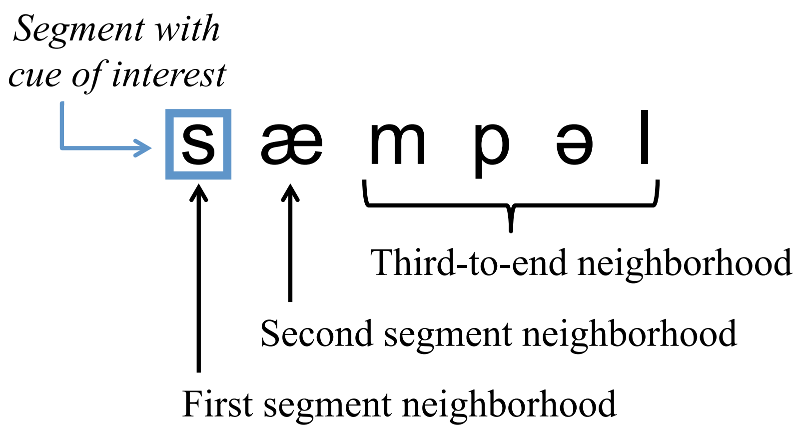

Fig. 2. Neighborhoods for the relative position of competition in the word *sample*. Each neighborhood density references a segmental position or set of segmental positions. The tally of the density of each of these neighborhoods is increased by one if replacing the segment in one of that neighborhood's segmental positions results in a unique lemma.

For the relative type of competition, we calculated modified neighborhood densities targeting three types of competition, each within the first segmental position: the place of articulation, manner of articulation, and voicing value of the competitor relative to the target. For each of these, we calculated two neighborhood densities, one for competitors that share the feature with the target and one for competitors that have a different value for that feature relative to the target. This resulted in a total of six modified neighborhood densities targeting the relative type of competition (Fig. 3). This set of position- and type-based competition measures was not intended to provide an exhaustive survey of all possible competition relationships, but rather was chosen to provide a reasonable sample of possible phonological relationships between target and competitor (for similar approaches, see Kirov & Wilson, 2012; Schertz, 2013).

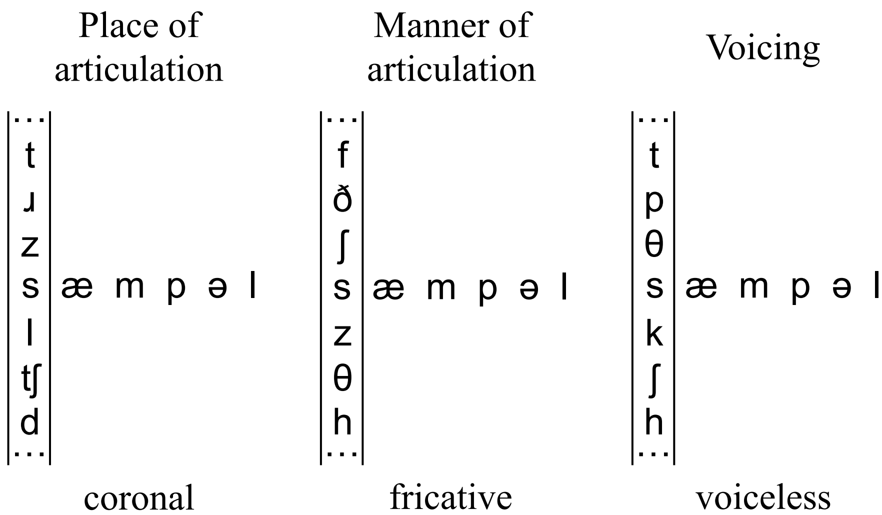

Fig. 3. The three types of same-feature competition included in our study, for the word *sample*. Each neighborhood represented here targets competitors with the same featural value as the target. Three other neighborhoods targeting competitors with different values for each of these three features were also calculated. The density of each of these neighborhoods is increased by one if replacing the first segment with another segment fitting that neighborhood's targeted featural type results in a unique lemma.

In order to calculate these competition metrics, we sorted the various phonemes of English into a simplified set of possible feature values. As our dependent measure was voice onset time, we sorted all consonants into one of three places of articulation: labial, coronal, or dorsal[1]. If a consonant could not be fully described by one of these feature values, that consonant was coded as the closest approximate place of articulation of these three (e.g., /f/ was coded as labial, as it is at least partly labial in articulation; /h/ was coded as dorsal, as dorsal is the closest place of articulation in our list to glottal). All consonants were additionally coded as either plosive or non-plosive, where "plosive" was defined as all oral stops and affricates (/p|b|t|d|k|g|t͡ʃ|d͡ʒ/), and "non-plosive" contained every other consonant, including nasal stops, approximates, and fricatives. Finally, all consonants were coded as either voiced or voiceless according to their standard phonemic transcription.

For a given target word, all neighborhood density measures were calculated by extracting the lemma form of the word and identifying the phonemic form of that lemma in our lexicon (sec. 2.1.3). We then identified each distinct lemma in that lexicon that fit the criteria of each of our competition metrics using regular expressions, and added it to the appropriate metric's neighborhood. Thus, each neighborhood density measure received a count of +1 for each unique lemma entry in our lexicon that could be formed by replacing the appropriate phoneme of the target word with another single phoneme, provided that the substitution fit the criteria of the neighborhood in question. In an effort to further test the range of possibilities in lexical competition, we also created versions of each of these neighborhood density measures that were weighted for frequency. We will turn to the description of these frequency-weighted measures in section 4.

[1] The single exception to this is /w/, which was coded as both labial and dorsal.

Competition metric	Segment(s) and feature(s) targeted	Examples for ***pat*** **(/pæt/)**	
Overall ND	Any segment Any feature	*spat* *at* *pit*	(/spæt/) (/æt/) (/pɪt/)
First segment ND	First segment only Any feature	*cat* *fat*	(/kæt/) (/fæt/)
Second segment ND	Second segment only Any feature	*pit* *pot*	(/pɪt/) (/pɔt/)
Third-to-end ND	Any segment after the second segment Any feature	*pan* *patch* *pack*	(/pæn/) (/pæt͡ʃ/) (/pæk/)
Same place of articulation ND	First segment only Same place of articulation	*fat* *mat*	(/fæt/) (/mæt/)
Different place of articulation ND	First segment only Different place of articulation	*cat* *sat*	(/kæt/) (/sæt/)
Same manner of articulation ND	First segment only Same manner of articulation	*cat* *bat*	(/kæt/) (/bæt/)
Different manner of articulation ND	First segment only Different manner of articulation	*fat* *that*	(/fæt/) (/ðæt/)
Same voicing ND	First segment only Same voicing value	*cat* *sat*	(/kæt/) (/sæt/)
Different voicing ND	First segment only Different voicing value	*that* *rat*	(/ðæt/) (/ræt/)
Minimal pair competitor existence	First segment only Same place of articulation Same manner of articulation Different voicing value	*bat*	(/bæt/)

Table 1. The competition metrics with descriptions and examples. Examples are non-exhaustive, provided for the target word *pat*.

2.3 Mixed effects models

The competition metrics were included in linear mixed effects models conducted in R (R Core Team 2016) using the lme4 package (Bates et al. 2015). Models were evaluated separately for voiced and voiceless stops due to different expectations for the effects of competition on the realization of voice onset times for each stop type (e.g., Schertz, 2013). For each of our stop types, each competition metric was included in a separate linear mixed effects regression model along with a consistent set of control predictors that have been shown to influence the realization of either voice onset time specifically, or word or segment durations more generally. For each subset of the data (voiced and voiceless stops), an additional model with only the control predictors and no competition metric was included as a baseline to which the other models could be compared (the "base model"). Thus, for both voiced and voiceless stops, the models in each set differed only in terms of the particular competition metric included (when one was included at all).

2.3.1 Factors of interest

VOT-length ratio: The dependent measure used in this study was the proportion of total stop duration taken up by the voice onset time (VOT-length ratio). This was calculated as the total voice onset time (from release of the closure to the end of aspiration) divided by the total stop duration (from initialization of the closure to the end of aspiration; Smiljanić & Bradlow, 2008; see also Wedel et al., submitted). As a proportion, the VOT-length ratios all lay on a scale between 0 and 1.

Competition metrics: Each model either included one of the competition metrics described above, or did not include any competition metric (base models). If the model included a competition metric, that value was centered, and then linearly transformed to a scale between -1 and 1 (note that this does not apply to the logical minimal pair existence factor).

2.3.2 Control predictors (fixed effects)

We included the following control predictors in all models. All continuous variables were centered, and then linearly transformed to a scale between -1 and 1 to facilitate model convergence.

Stop phoneme: The identity of the initial stop phoneme (/p|t|k/, /b|d|g/).

Speech rate: The number of phonemic vowels (i.e. syllables) per second (Bell et al., 2003). Speech rate is highly correlated with durational cues, including voiceless stop voice onset time (e.g., Yao, 2007).

Contextual diversity: The percent of films in which the word appears in the SUBTLEX-US database. Word familiarity measures correlate with reduction (e.g., word frequency: Bell et al., 2009), and contextual diversity has been found to be a better predictor of behavioral data than other familiarity measures such as frequency (Brysbaert & New, 2009).

Following liquid: Logical factor for whether or not the initial stop was followed by a liquid (/l|ɹ|ɚ/). Visual inspection of the data indicated that words followed by a liquid had considerably longer voice onset times.

Number of syllables: Either 1 or 2, coded as a factor. This factor was included under the assumption that bisyllabic words would have systematically shorter voice onset times on average.

Conditional (forward/backward) bigram probability: The log-transformed conditional bigram probability of each word given the preceding word (backward conditional bigram probability) and the following word (forward conditional bigram probability), based on a combination of the Fisher English Training Part 2 corpus (Cieri et al., 2005) and the Buckeye Corpus of Conversational Speech (Pitt et al., 2005; Pitt et al., 2007). Conditional probability is predictive of durations in the Buckeye corpus (Seyfarth, 2014).

Previous mention: Logical factor for whether or not the lemma (not necessarily the specific word) occurred earlier in the relevant transcript. Repetitions of words are more reduced than initial productions (Bell et al., 2009).

Phonotactic probability: The log-transformed average two-phoneme sequence probability, based on relative position within the word, calculated using the IPhOD2 database's unstressed, unweighted biphoneme probability calculator (Vaden et al., 2009).

2.3.3 Random effects

All models included random intercepts for both Speaker and Lemma. In addition, for each model with a competition metric, we included a correlated random slope for that metric on the Speaker intercept (Barr et al., 2013). We did not include random slopes for our control predictors as this would lead to problems with model convergence, and we had no principled reason to include any particular slopes over others.

2.4 Model evaluation procedure

Models were first evaluated according to their corrected Akaike's Information Criterion (AIC_c), an information-theoretic approach to model comparison based on entropy (see Burnham & Anderson, 2004), using the AICcmodavg package (Mazerolle, 2016). The AIC_c value can be thought of in terms of information loss, where a lower AIC_c value corresponds to less information loss and therefore a more accurate model. However, AIC_c values by themselves are essentially meaningless, and must be interpreted relative to alternative AIC_c values. To do this, models are included in a candidate set and ranked according to their AIC_c value. These models are given weights based on their normalized log-likelihood, and compared based on these weights (Burnham & Anderson, 2004). This relative evaluation is operationalized in terms of the change in AIC_c value between the current model and the top-performing model (Δ AIC_c). In general, models with a Δ $AIC_c \leq 2$ are considered to have substantial support relative to the top-performing model (i.e., they are not deemed considerably inferior to that model), and models with a Δ $AIC_c \geq 10$ are taken to have very little support (Burnham & Anderson, 2004). In addition, AIC_c comparison allows evidence ratios to be calculated. Evidence ratios compare two models directly, and are calculated as the ratio of the AIC_c weights of the two models (Richards et al., 2011). These evidence ratios allow us to describe the models in terms of the amount of evidence in favor of the better model with respect to the other.

In addition to the AIC_c model comparison, we further tested the top-performing model of each set (voiced and voiceless stop models) for the significance of the competition metric. To test for statistical significance, we subjected these models to nested model comparison using the log-likelihood ratio test. In this case, the models in question were compared to the corresponding model with the fixed effect of the competition metric removed. Thus, for both voiced and voiceless stops separately, the model with the lowest AIC_c value was compared to a restricted version of itself without the fixed effect of the relevant competition metric.

3 Results

Results are presented for voiced (section 3.1) and voiceless (section 3.2) stops separately. Summaries of the top-performing models can be found in Appendices B (voiced) and C (voiceless).

3.1 Voiced stops

The voiced stop data included 2 267 observations meeting our criteria from 24 speakers, distributed over 293 lemmas. Variance inflation factors (VIFs) for all of the predictors of interest were less than 1.9, indicating low multi-collinearity between the factors of interest and control predictors. The overall neighborhood density metric had the highest VIF of all factors of interest (1.80), while minimal pair existence had the smallest VIF (1.21). VIFs less than 2 are generally not cause for concern (see O'Brien, 2007 and Belsley, Kuh, & Welsch, 1980/2005, ch. 3, for discussion).

3.1.1 AIC_c comparison

The AIC_c comparison table for all voiced stop models is presented in Table 2. Models are ranked according to their AIC_c value, with a lower AIC_c value corresponding to a better model fit. The model including a factor for minimal pair existence had the lowest AIC_c value (Δ $AIC_c = 0$), and therefore the best overall fit to the data. The model including a factor for neighbors with the same place of articulation as the target was the second-best model (same place ND: Δ $AIC_c = 4.49$), but evidence ratios indicate that there was about 9 times more evidence in favor of the minimal pair model. The only other model with a Δ $AIC_c < 10$ was the neighborhood density for second segment competitors (second segment ND: Δ $AIC_c = 8.13$), with about 58 times more evidence in favor of the minimal pair model. AIC_c comparisons suggest virtually no evidence in favor of any of the other competition metrics in the voiced stop dataset (AIC_c weight < 0.01; Δ $AIC_c > 10$). These results suggest that the existence of the phonetically specific minimal pair competitor in the lexicon was the best predictor of voice onset time values for voiced stops in this dataset.

Rank	Model	K	AIC_c	Δ AIC_c	AIC_cWt	Cum.Wt	LL
1	Minimal pair exist	20	-4131.63	0.00	0.88	0.88	2086.00
2	Same place ND	20	-4127.14	4.49	0.09	0.97	2083.76
3	Second segment ND	20	-4123.49	8.13	0.02	0.99	2081.93
4	Overall ND	20	-4121.26	10.36	0.00	0.99	2080.82
5	First segment ND	20	-4118.50	13.13	0.00	1.00	2079.43
6	Base model	17	-4118.22	13.41	0.00	1.00	2076.24
7	Different manner ND	20	-4117.89	13.74	0.00	1.00	2079.13
8	Third-to-end ND	20	-4117.28	14.35	0.00	1.00	2078.83
9	Same voicing ND	20	-4116.93	14.70	0.00	1.00	2078.65
10	Different voicing ND	20	-4116.49	15.14	0.00	1.00	2078.43
11	Same manner ND	20	-4115.87	15.76	0.00	1.00	2078.12
12	Different place ND	20	-4115.83	15.80	0.00	1.00	2078.10

Table 2. AIC_c comparison table for all models predicting word-initial voiced stop VOT-length ratios. Models are ranked in order of AIC_c value. K = the number of estimable parameters in the model. AIC_c = the corrected AIC value. ΔAIC_c = the change in corrected AIC value between each model and the top performing model. *AIC_cWt* = the relative likelihood that the present model is the best model, presented as a proportional weight. *Cum.Wt* = the cumulative AIC_c weight of the present model and all higher-ranked models. *LL* = the log-likelihood of the model.

3.1.2 The top model: minimal pair existence

The minimal pair existence model is summarized in Appendix B. As predicted, minimal pair existence correlates with a *decrease* in the VOT-length ratio, suggesting that these words are contrastively hyperarticulated away from their voice onset time competitor. As the top-performing model, we evaluated whether the minimal pair existence factor significantly contributed to model fit using nested model comparison. Removal of the fixed effect of minimal pair existence significantly affected model fit ($\chi^2(1) = 16.319$, $p < 0.001$), indicating that minimal pair existence is a significant predictor of voiced stop VOT-length ratios.

3.2 Voiceless stops

The voiceless stop data included 3 690 observations meeting our criteria from 24 speakers, distributed over 417 lemmas. VIFs for all of the predictors of interest were less than 1.8, indicating low multi-collinearity between the factors of interest and control predictors. The overall neighborhood density metric had the highest VIF of all factors of interest (1.70), while minimal pair existence had the smallest VIF (1.10).

3.2.1 AIC_c comparison

The AIC_c comparison table for all voiceless stop models is presented in Table 3. The model including a factor for minimal pair existence had the lowest AIC_c value ($\Delta\ AIC_c = 0$), and therefore the best overall fit to the data. The model including a factor for the second segment neighborhood density was the second-best model ($\Delta\ AIC_c = 2.56$), and evidence ratios indicate that there was only about 4 times more evidence in favor of the minimal pair model. Relative to all other models, the minimal pair model had limited support ($\Delta\ AIC_c > 2$). However, the minimal pair model had more substantial support relative to the base model ($\Delta\ AIC_c = 5.82$), with about 18 times more evidence in favor of the minimal pair existence model over the base model. These results suggest that the existence of the phonetically specific minimal pair competitor in the lexicon was the best predictor of voiceless stop voice onset times, but that the support for this model over many of the other models was limited.

Rank	Model	K	AIC_c	$\Delta\ AIC_c$	AIC_cWt	Cum.Wt	LL
1	Minimal pair exist	20	-6124.87	0.00	0.55	0.55	3082.55
2	Second segment ND	20	-6122.31	2.56	0.15	0.70	3081.27
3	Same place ND	20	-6121.32	3.55	0.09	0.79	3080.77
4	Same manner ND	20	-6120.24	4.63	0.05	0.85	3080.24
5	Different voice ND	20	-6120.24	4.63	0.05	0.90	3080.24
6	Overall ND	20	-6119.39	5.48	0.04	0.94	3079.81
7	Base model	17	-6119.05	5.82	0.03	0.97	3076.61
8	First segment	20	-6117.66	7.21	0.01	0.98	3078.95
9	Different manner ND	20	-6116.30	8.57	0.01	0.99	3078.27
10	Different place ND	20	-6115.79	9.08	0.01	0.99	3078.01
11	Third-to-end ND	20	-6114.44	10.43	0.00	1.00	3077.34
12	Same voice ND	20	-6113.72	11.15	0.00	1.00	3076.98

Table 3. AIC_c comparison table for all models predicting word-initial voiceless stop VOT-length ratios. Models are ranked in order of AIC_c value. *K* = the number of estimable parameters in the model. AIC_c = the corrected AIC value. $\Delta\ AIC_c$ = the change in corrected AIC value between each model and the top performing model. AIC_cWt = the relative likelihood that the present model is the best model, presented as a proportional weight. *Cum.Wt* = the cumulative AIC_c weight of the present model and all higher-ranked models. *LL* = the log-likelihood of the model.

3.2.2 The top model: minimal pair existence

The minimal pair existence model is summarized in Appendix C. As predicted, minimal pair existence correlates with an *increase* in the VOT-length ratio, suggesting that these words are contrastively hyperarticulated away from their voice onset time competitor. As the top-performing model, we evaluated whether the minimal pair existence factor significantly contributed to model fit using nested model comparison. Removal of the fixed effect of minimal pair existence significantly affected model fit ($\chi^2(1) = 5.25$, $p < 0.05$), indicating that minimal pair existence is a significant predictor of voiceless stop VOT-length ratios.

3.3 Summary and discussion

These results support the claim that minimal pairs are hyperarticulated away from each other in a manner that enhances the phonetic contrast between them. We found evidence of this contrastive hyperarticulation in the form of both longer voice onset times in voiceless stops, and shorter voice onset times in voiced stops. For both voiced and voiceless stops, the best model in terms of AIC_c included a factor for cue-specific minimal pair existence that contributed significantly to model fit. This suggests that competition defined for the cue of interest is the best predictor of contrastive hyperarticulation in word-initial stops of conversational English. We also found limited evidence that competition among neighbors that share a place of articulation in their initial segment (same place ND), as well as competition in the following segment (second segment ND), can influence voice onset time in word-initial stops (these were the second- and third-best models in terms of AIC_c for both voiced and voiceless stops); however, the evidence in support of these models was relatively weak (all Δ $AIC_c > 2$).

In this dataset, we find evidence of contrastive hyperarticulation in both voiced and voiceless stops. To our knowledge, only three prior investigations of contrastive hyperarticulation have examined voiced stops (Ohala, 1994; Schertz, 2013; Goldrick et al., 2013). Schertz (2013) found that voiced stops were realized with shorter voice onset times in clarifications following misperception as their voiceless counterpart, paralleling our results here. Ohala (1994) found the same trend, but it was not significant. Goldrick, Vaughn, and Murphy (2013), however, found that word-initial voiced stops were not articulated differently depending on whether they had an initial stop voicing minimal pair in a word list reading task. These differences in results may be due, in part, to the use of different tasks. Studies employing word or sentence reading tasks to look for contrastive hyperarticulation of minimal pairs have sometimes found small effects (Baese-Berk & Goldrick, 2009, study 1; Peramunage et al., 2011; but see Fricke et al., 2016), but have sometimes failed to find any effect (e.g., Goldrick et al., 2013; Fox et al., 2015). On the other hand, studies in which cue-specific minimal pairs compete directly in communicative tasks have reported contrastive hyperarticulation effects more consistently or more robustly (Baese-Berk & Goldrick, 2009, study 2; Kirov & Wilson, 2012; Schertz, 2013; Seyfarth et al., 2016; Buz et al., 2016; but see Ohala, 1994). This suggests that contrastive hyperarticulation effects of the type reported here may be more easily detected under conditions promoting communication (Buz et al., 2016). That we find evidence of contrastive hyperarticulation of voiced stops may partly reflect the communicative nature of the speech in this dataset.

4 Alternative analyses based on frequency-weighted neighborhood measures

We repeated our analyses using frequency-weighted versions of the competition metrics described above. Following Luce and Pisoni (1998), these frequency-weighted measures were constructed as the ratio of the log-transformed target word frequency over the total neighborhood frequency, defined as the sum of the log-transformed frequencies of every member of the neighborhood (including the target). This method of frequency-weighting neighborhood density is commonly used in studies of competition-induced hyperarticulation (e.g., Munson, 2007; Scarborough, 2013; Buz & Jaeger, 2016). This measure is typically used in the choice of experimental materials to bin test items into ‘hard’ and ‘easy’ categories. Hard words have low frequency relative to their (relatively numerous) neighbors, while easy words have high frequency relative to their (relatively less numerous) neighbors (e.g., Wright, 2004). Log-transformed frequencies were extracted directly from SUBTLEX-US (Brysbaert & New, 2009). Henceforth, we will refer to these frequency-weighted ratios as 'neighborhood frequency' measures.

As an illustration of the simplest case, the minimal pair neighborhood frequency consists of the log-transformed word frequency of the target divided by the sum of the log-transformed frequencies of the target and its voicing minimal pair competitor. Thus, if the target does not have a voicing minimal pair competitor (e.g., *bright*), the minimal pair neighborhood frequency is 1 (log-transformed target frequency divided by log-transformed target frequency + 0). If the target does have a minimal pair competitor, however, the denominator will include the frequency of that competitor, will therefore be larger than the numerator, and the resulting value of the ratio will be less than 1. Consequently, these neighborhood frequency measures range in value from 0 to 1, where 1 indicates that the target has no neighbors as defined for that neighborhood[2].

We repeated our mixed effects analyses as reported above for these neighborhood frequency measures. As before, the competition metrics were centered, and each appeared in its own model as both a fixed effect and as a random slope on speaker. Results are presented for voiced (section 4.1) and voiceless (section 4.2) stops separately.

4.1 Voiced stops

As before, the voiced stop data included 2 267 observations meeting our criteria from 24 speakers, distributed over 293 lemmas. VIFs for all of the predictors of interest were less than 1.8, indicating low multi-collinearity between the factors of interest and control predictors. The frequency-weighted metric corresponding to all neighbors with the same voicing value as the target had the highest VIF of all factors of interest (1.73), while the frequency-weighted minimal pair factor had the smallest VIF (1.17).

4.1.1 AIC_c comparison

The AIC_c comparison table for all voiced stop models is presented in Table 4. The model including the frequency-weighted minimal pair factor had the lowest AIC_c value ($\Delta AIC_c = 0$), and therefore the best overall fit to the data. The model including a factor for the second segment neighborhood frequency was the second-best model (second segment ND: $\Delta AIC_c = 4.44$), and evidence ratios indicate that there was about 9 times more evidence in favor of the minimal pair model. The only other model with limited support was the model including a factor for neighbors that share their place of articulation with the target (same place ND: $\Delta AIC_c = 4.52$). All other models had little or no support (all $\Delta AIC_c > 12$). This result suggests that a frequency-weighted measure of cue-specific minimal pair competition is a better predictor of voiced stop voice onset times than other frequency-weighted neighborhood measures, mirroring the results for unweighted neighborhood densities.

[2] Mathematically, a value of 0 is impossible, but the value can come arbitrarily close to 0 in theory.

Rank	Model	K	AIC_c	Δ AIC_c	AIC_cWt	Cum.Wt	LL
1	Minimal pair NF	20	-4132.31	0.00	0.82	0.82	2086.34
2	Second segment NF	20	-4127.86	4.44	0.09	0.91	2084.12
3	Same place NF	20	-4127.78	4.52	0.09	0.99	2084.08
4	Same voicing NF	20	-4120.05	12.26	0.00	1.00	2080.21
5	Different manner NF	20	-4119.53	12.78	0.00	1.00	2079.95
6	Base model	17	-4118.22	14.09	0.00	1.00	2076.24
7	First segment NF	20	-4118.02	14.28	0.00	1.00	2079.20
8	Third-to-end NF	20	-4117.78	14.52	0.00	1.00	2079.08
9	Overall NF	20	-4116.44	15.86	0.00	1.00	2078.41
10	Same manner NF	20	-4115.58	16.73	0.00	1.00	2077.98
11	Different voicing NF	20	-4115.37	16.94	0.00	1.00	2077.87
12	Different place NF	20	-4114.57	17.74	0.00	1.00	2077.47

Table 4. AIC_c comparison table for all models predicting word-initial voiced stop VOT-length ratios. Models are ranked in order of AIC_c value. *K* = the number of estimable parameters in the model. AIC_c = the corrected AIC value. ΔAIC_c = the change in corrected AIC value between each model and the top performing model. AIC_cWt = the relative likelihood that the present model is the best model, presented as a proportional weight. *Cum.Wt* = the cumulative AIC_c weight of the present model and all higher-ranked models. *LL* = the log-likelihood of the model.

4.1.2 The top model: frequency-weighted minimal pair competition

The frequency-weighted minimal pair model is summarized in Appendix D. As predicted, the correlation between this factor and the VOT-length ratio is positive, indicating that voice onset time decreases as competition increases (i.e., as the frequency ratio decreases). This suggests that these words are contrastively hyperarticulated away from their voice onset time competitor. We evaluated whether the minimal pair existence factor significantly contributed to model fit using nested model comparison, and found that removal of the fixed effect of minimal pair existence significantly affected model fit ($\chi^2(1) = 17.13$, $p < 0.001$). As with the unweighted neighborhood densities, this indicates that cue-specific minimal pair competition is a significant predictor of initial voiced stop voice onset times.

4.2 Voiceless stops

As before, the voiceless stop data included 3 690 observations meeting our criteria from 24 speakers, distributed over 417 lemmas. VIFs for all of the predictors of interest were less than 1.6, indicating low multi-collinearity between the factors of interest and control predictors. The second segment neighborhood frequency metric had the highest VIF of all factors of interest (1.53), while the frequency-weighted minimal pair factor had the smallest VIF (1.10).

4.2.1 AIC_c comparison

The AIC_c comparison table for all voiceless stop models is presented in Table 5. The model including the overall neighborhood frequency had the lowest AIC_c value ($\Delta AIC_c = 0$), and therefore the best overall fit to the data. No other model had substantial support (all $\Delta AIC_c > 12$), and evidence ratios indicate that there was about 552 times more evidence in favor of the overall neighborhood frequency model over the next-best model (first segment NF: $\Delta AIC_c = 12.63$). This result stands in stark contrast to the results of our other three analyses, in which cue-specific minimal pair competitor existence or neighborhood frequency provided the most predictive models of voice onset times.

Rank	Model	K	AIC_c	Δ AIC_c	AIC_cWt	Cum.Wt	LL
1	Overall NF	20	-6141.40	0.00	1	1	3090.82
2	First segment NF	20	-6128.78	12.63	0	1	3084.50
3	Different voicing NF	20	-6126.32	15.08	0	1	3083.27
4	Second segment NF	20	-6123.87	17.53	0	1	3082.05
5	Different manner NF	20	-6122.92	18.48	0	1	3081.58
6	Same place NF	20	-6122.75	18.65	0	1	3081.49
7	Minimal pair NF	20	-6122.59	18.82	0	1	3081.41
8	Same manner NF	20	-6122.18	19.22	0	1	3081.21
9	Different place NF	20	-6121.11	20.29	0	1	3080.67
10	Base model	17	-6119.05	22.35	0	1	3076.61
11	Same voicing NF	20	-6117.03	24.38	0	1	3078.63
12	Third-to-end NF	20	-6116.15	25.25	0	1	3078.19

Table 5. AIC_c comparison table for all models predicting word-initial voiceless stop VOT-length ratios. Models are ranked in order of AIC_c value. *K* = the number of estimable parameters in the model. AIC_c = the corrected AIC value. ΔAIC_c = the change in corrected AIC value between each model and the top performing model. AIC_cWt = the relative likelihood that the present model is the best model, presented as a proportional weight. *Cum.Wt* = the cumulative AIC_c weight of the present model and all higher-ranked models. *LL* = the log-likelihood of the model.

4.2.2 The top model: overall neighborhood frequency

The overall neighborhood frequency model is summarized in Appendix E. As predicted, the correlation between this factor and the VOT-length ratio is negative, indicating that voice onset time increases as competition increases (i.e., as the frequency ratio decreases). We evaluated whether the overall neighborhood frequency factor significantly contributed to model fit using nested model comparison, and found that removal of the fixed effect of overall neighborhood frequency significantly affected model fit ($\chi^2(1) = 9.74$, $p < 0.01$).

4.2.3 A closer look at the neighborhood frequency metric

For the analyses of both frequency-weighted and unweighted competition metrics in the voiced stop data, as well as the analysis of unweighted measures in the voiceless stop data, cue-specific minimal pair competition provided the best predictor of voice onset time realizations. However, for frequency-weighted competition metrics in the voiceless stop data, the opposite measure as defined by our continuum of specificity was most predictive: overall neighborhood density.

Why might this result be so different from the others? One possibility lies in the distribution of data along the neighborhood frequency variable. A feature of the neighborhood frequency metric is that items with no neighbors have a neighborhood frequency of 1, because the ratio for words with no neighbors is the frequency of the word divided by itself. When items are included that have no neighbors, this can result in a distribution in which the majority of values are concentrated at the lower range, with an isolated peak at 1 corresponding to words with no neighbors. In the voiceless dataset, the bulk of the data is distributed between overall neighborhood frequency values of 0 and 0.2 (Fig. 4). There is a substantial gap in the observed values in the upper range, followed by a single peak corresponding to a neighborhood frequency ratio of 1. Extreme values can exert undue leverage on regression models, and notably, studies using neighborhood frequency to contrast ‘hard’ vs. ‘easy’ words have excluded words with no neighbors, i.e. those with a neighborhood frequency value of 1 (e.g., Wright, 2004; Munson & Solomon, 2004; Munson, 2007; Scarborough, 2013). To ask whether words with no neighbors contributed to this anomalous result, we removed from the dataset the 337 observations of the 69 lemmas with no neighbors and repeated the analysis. 67 of these lemmas were bisyllabic, and the remaining 2 monosyllabic lemmas were the phonotactically unusual *prompt* and *puke*. The resulting dataset retained 3 353 observations from 24 speakers, distributed over 348 lemmas.

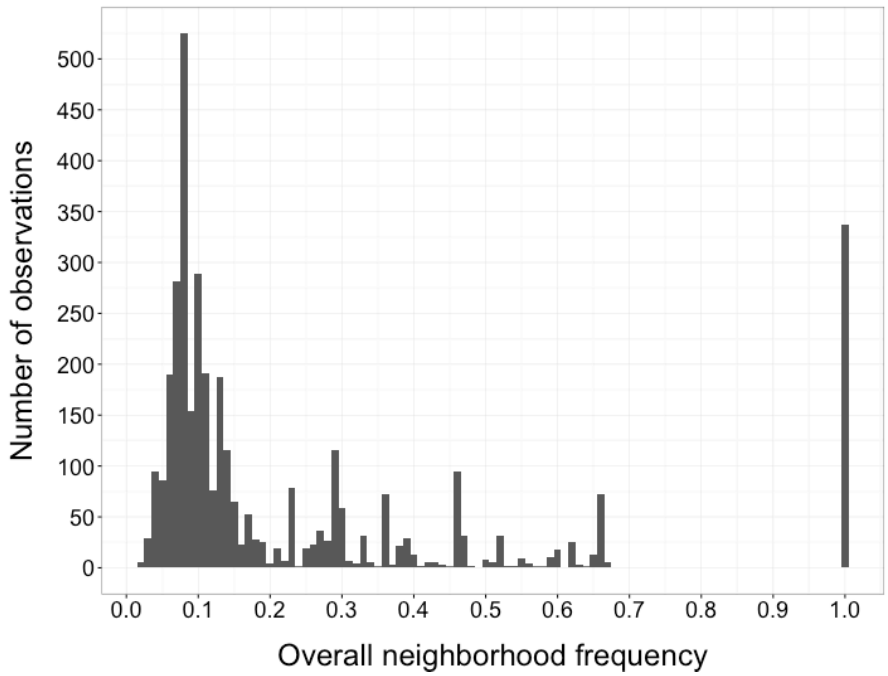

Fig. 4. The distribution of observations in the voiceless stop dataset by overall neighborhood frequency.

The AIC_c comparison table for all voiceless stop models is presented in Table 6. The model including the overall neighborhood frequency still had the lowest AIC_c value ($\Delta\ AIC_c = 0$), and therefore the best overall fit to the data. However, the overall neighborhood frequency factor no longer contributed significantly to model fit ($\chi^2(1) = 2.74$, $p > 0.05$; see Appendix E for details). Furthermore, two other models now varied minimally from the overall neighborhood frequency model: the first segment neighborhood frequency model ($\Delta\ AIC_c = 2.07$) and the minimal pair neighborhood frequency model ($\Delta\ AIC_c = 2.27$). Evidence ratios indicate that there was about 3 times more evidence in favor of the overall neighborhood frequency model relative to both the first segment neighborhood frequency model and the minimal pair neighborhood frequency model. These results indicate that overall neighborhood frequency is still the most predictive of these frequency-weighted competition metrics in this voiceless stop dataset, but the effect of this factor was not significant and the evidence in support of this model over the two subsequent models in the AIC_c ranking is limited. We confirmed that unweighted minimal pair existence was still significantly predictive of voice onset times in this subset of the voiceless stop data ($\chi^2(1) = 5.38$, $p < 0.05$), suggesting that the lack of significance for the overall neighborhood frequency factor is not purely due to a reduction of power.

Rank	Model	K	AIC_c	Δ AIC_c	AIC_cWt	Cum.Wt	LL
1	Overall NF	20	-5555.97	0.00	0.38	0.38	2798.11
2	First segment NF	20	-5553.90	2.07	0.13	0.51	2797.08
3	Minimal pair NF	20	-5553.71	2.27	0.12	0.63	2796.98
4	Different voicing NF	20	-5553.42	2.55	0.11	0.74	2796.84
5	Same place NF	20	-5552.42	3.55	0.06	0.80	2796.34
6	Second segment NF	20	-5552.10	3.87	0.05	0.85	2796.18
7	Base model	17	-5551.72	4.26	0.04	0.90	2792.95
8	Different manner NF	20	-5551.33	4.46	0.04	0.94	2795.79
9	Same manner NF	20	-5551.05	4.92	0.03	0.97	2795.65
10	Different place NF	20	-5549.75	6.23	0.02	0.99	2795.00
11	Same voicing NF	20	-5548.95	7.02	0.01	1.00	2794.60
12	Third-to-end NF	20	-5546.59	9.38	0.00	1.00	2793.42

Table 6. AIC_c comparison table for all models predicting word-initial voiceless stop VOT-length ratios in the subset of data for which overall neighborhood frequency is less than 1. Models are ranked in order of AIC_c value. *K* = the number of estimable parameters in the model. AIC_c = the corrected AIC value. ΔAIC_c = the change in corrected AIC value between each model and the top performing model. AIC_cWt = the relative likelihood that the present model is the best model, presented as a proportional weight. *Cum.Wt* = the cumulative AIC_c weight of the present model and all higher-ranked models. *LL* = the log-likelihood of the model.

4.3 Summary and discussion

The results for voiced stops were nearly identical for both frequency-weighted and unweighted measures. The competition metric corresponding to cue-specific minimal pair competition produced the best models in both cases, where minimal pair competition correlated with shorter voice onset times. The results for voiceless stops, however, varied substantially. While the analysis for unweighted competition metrics indicated that cue-specific minimal pair existence was the best predictor of voiceless stop voice onset times, the analysis for frequency-weighted competition metrics provided more evidence in favor of the overall neighborhood frequency measure. However, we noted that the neighborhood frequency metric results in a value of 1 for words without any neighbors, regardless of their frequency. This cluster of data at one extreme of the distribution creates the possibility for leverage, where a relatively small number of observations can have undue influence on the regression model. What is more, studies using this neighborhood frequency measure to categorize words as 'easy' or 'hard' have excluded words with no neighbors (e.g., Wright, 2004; Munson & Solomon, 2004; Munson, 2007; Scarborough, 2013). We repeated our analysis of the voiceless data with these words excluded and found that the overall neighborhood frequency model had the greatest support, but the effect of overall neighborhood frequency was no longer significant. These results suggest that, at least within this dataset, neighborhood frequency measures are not robustly predictive of voice onset time in voiceless stops.

We also found limited evidence in favor of both the first segment and minimal pair neighborhood frequency measures predicting voice onset times in voiceless stops (first segment: $\Delta AIC_c = 2.07$; minimal pair: $\Delta AIC_c = 2.27$). These findings parallel those of Fricke (2013), who reported that overall neighborhood density, minimal pair competitor existence, and the number of competitors differing in their onset ("rhyme neighbors" in her terminology) all correlated with hyperarticulation of voiceless stop voice onset times in the Buckeye Corpus (see also Fricke et al., 2016). However, Fricke reported that onset competition was a better predictor of voice onset time than either minimal pair competitor existence or overall neighborhood density, concluding that each of these other measures correlate with voice onset time because they also correlate with the position-dependent measure of onset competition. Instead, we found that overall neighborhood density was a better predictor, but was not significant when we excluded words with no neighbors. Some methodological considerations may help to explain this different result. For example, Fricke only included monosyllabic words beginning with simplex onsets, while we included bisyllabic words and words with complex onsets. This difference affects our various neighborhood measures, particularly the neighborhoods corresponding to onset/first segment competition.

5 General Discussion

We identified a set of lexical competition metrics that sampled a conceptual space between the phonetically specific measure of cue-defined minimal pair existence, and the phonetically more general measure of neighborhood density. These measures were tested for their ability to predict voice onset time in voiced and voiceless word-initial stops of conversational English. Previous studies have suggested that competition from a cue-defined minimal pair competitor induces contrastive hyperarticulation of voice onset time (e.g., Baese-Berk & Goldrick, 2009; Schertz, 2013; Buz et al., 2016), and we found support for this hypothesis in the form of shorter voice onset times in voiced stops and longer voice onset times in voiceless stops (see also Wedel et al., submitted). However, previous studies using a variety of speech elicitation paradigms have suggested that less phonetically specific measures of competition also correlate with hyperarticulation of voice onset time in voiceless stops (e.g., Kirov & Wilson, 2012), in some cases more strongly than the cue-specific minimal pair competitor measure (Fricke, 2013; Fricke et al., 2016; Fox et al., 2015). In this natural speech dataset, we found that these other metrics of competition were less predictive of the realization of voice onset time in both voiced and voiceless stops.

We also conducted analyses in which we weighted our competition metrics for the relative frequencies of neighbors. For the voiced stops, this neighborhood frequency approach did not alter the results. The model including a frequency-weighted factor for the existence of an initial stop voicing minimal pair remained the most predictive of our models. For voiceless stops, however, overall neighborhood frequency provided the most predictive model. Upon further inspection, we noted the possibility that the extreme neighborhood frequency values contributed by words with no neighbors could unduly leverage model outcomes. With these data removed, the model including overall neighborhood frequency remained the most predictive in terms of AIC_c, but the contribution of overall neighborhood frequency to the model was no longer significant. Further, the models including factors for first segment and minimal pair neighborhood frequencies were minimally different from the overall neighborhood frequency model in terms of AIC_c.

5.1 Alternative metrics of lexical competition

The competition measures tested in this study represent only a subset of the large and multi-dimensional hypothesis space surrounding lexical competition effects. One way to expand this hypothesis space is to extend the continuum beyond a single-phoneme edit distance between target and competitor, so that similar words with a multi-phoneme edit distance can still contribute to competition. An example of such a measure in the literature is to weight lexical items for their Levenshtein edit distance, so that edit distances greater than 1 contribute to the neighborhood, but less so than more closely related words (Yarkoni et al., 2008). A related, but different approach is to weight the neighbors for proportion of position-dependent segmental overlap with the target (Goldrick et al., 2010). Using either of these approaches, *trap* would not only have such neighbors as *trip*, *tap*, and *track*, but also neighbors such as *trick* and *tarp*, for which more than one segment are different. Proportion of segmental overlap has been found to be predictive of spoken and written errors in subjects with acquired language impairment, especially when additionally weighted for frequency, syntactic category, and onset overlap (i.e., the Lex-Form Composite: Goldrick et al., 2010). However, broader competition metrics such as these are less phonetically specific than the range of neighborhood measures tested in our study. We find that hyperarticulation of word-initial voice onset time in conversational English is primarily predicted by a competitor differing solely in word-initial voicing, suggesting that these broader alternatives should be less predictive of contrastive hyperarticulation in voice onset time.

Another competition metric from the literature compares the target to all words in the lexicon, weighting them for their perceptual similarity (Phi-square density: Strand & Sommers, 2011). Under this approach, *fist* and *fish* would be considered stronger competitors than *fist* and *fit*, because /s/ and /ʃ/ are perceptually similar, while the absence of /s/ is perceptually salient. This measure has been found to be predictive of word recognition accuracy, but whether it contributes predictive power above and beyond overall neighborhood density is unclear, and it was not found to be significantly predictive of word durations in the Buckeye corpus (Gahl & Strand, 2016). Although the phi-square density measure is gradiently sensitive to similarity, it considers perceptual relationships without respect to a particular cue. As such, neighbors that are perceptually similar in the segment or cue of interest, such as the cue-specific minimal pair competitor, are given no priority over neighbors differing in other positions or cues. It is unclear to what extent this measure should correlate with contrastive hyperarticulation of individual cues, but it may provide a fruitful model for studying the role of perceptual similarity in contrastive hyperarticulation (cf. Buz et al., 2016).

5.2 Other cue-types

If these results are representative, we expect to find similar contrastive hyperarticulation of other types of cues, where this hyperarticulation specifically increases phonetic distance between competitors. In another study of conversational speech, Wedel, Nelson, and Sharp (submitted) found evidence of hyperarticulation in vowels in the form of F1-F2 Euclidean distance. Vowels in words with minimal pairs defined by a nearby competitor vowel (e.g., *pit* ~ *pet*) were farther apart in F1-F2 space than the same vowels in words without such minimal pairs (e.g., *ship* ~ **shep*). This difference was not explained by more general competition in the form of overall neighborhood density. In conjunction with the results reported here, this suggests that contrastive hyperarticulation of phonetic cues arises in response to competition with nearby competitors within that cue, and that this holds across two very different cue types (voice onset time and F1-F2 Euclidean distance).

5.3 Cognitive mechanisms

What do our results have to say about the cognitive mechanisms proposed to underlie competition-based hyperarticulation? In production-internal accounts that appeal to interactive cascading activation, the set of competitors has often been defined broadly as the set of lexical neighbors differing by one edit-distance in any position. More recent accounts, however, have been modified to include differential neighborhood effects at different positions within words (e.g., Vitevitch et al., 2004; Goldrick et al., 2010; Fricke, 2013). For example, Fricke (2013) proposed that interactive cascading activation during motor planning is processed on a segment-by-segment basis. Each progressive segment is articulated as soon as competition is resolved, with greater competition leading to higher activation levels and, consequently, longer phonetic durations. According to this Articulate As Soon As Possible Principle, competition is sensitive to segmental position within the word, but is insensitive to the phonetic relationships among competitors within that segmental position. Consequently, in the case of word-initial voice onset time, this model predicts that the overall competition within the word onset should be the best predictor of competition-induced hyperarticulation.

Our findings did not support this hypothesis. In general, the existence of the minimal pair competitor defined by the initial stop voicing contrast was most predictive of voice onset times for both voiced and voiceless stops. This result is difficult to reconcile with any model of competition effects that does not consider the specific phonetic contrast between target and competitor. On the other hand, this result is consistent with communicative and listener-internal approaches, according to which the contrastive cues to phonemic and/or lexical categories can be specifically targeted for hyperarticulation (e.g., Pierrehumbert, 2002; Wedel, 2006; Buz et al., 2016; Hall et al., submitted). Notably, there is evidence that these cues may be especially targeted for hyperarticulation when they are most confusable. Buz, Tanenhaus, & Jaeger (2016) elicited productions of voiceless stop-initial monosyllabic words of English in a cooperative task. Speakers produced longer voice onset times on average when the minimal pair competitor was present in the experimental context, replicating past findings (e.g., Baese-Berk & Goldrick, 2009). Crucially, in a post-hoc analysis they found evidence that speakers were achieving this shift in average voice onset time not by increasing the target voice onset time of their productions, but by reducing the distribution of productions near the category boundary, where small changes to voice onset time have detectable effects on perception (McMurray, Aslin, & Tanenhaus, 2002). This suggests that speakers specifically avoid production variants that would result in greater perceptual similarity with a competitor.

If speakers contrastively hyperarticulate by avoiding more confusable productions, we expect that contrastive hyperarticulation may be less evident in conditions favoring clear-speech hyperarticulation, in which category contrasts may be inherently clearer. In elicited speech paradigms such as those reported above, voice onset times might be distinct enough a priori that the voiced and voiceless categories are completely non-overlapping[3]. In such a case, speakers may not attempt to hyperarticulate further, as the added effort would offer little in communicative benefit. On the other hand, if less formal registers induce greater overall reduction, we would expect phonetically contrastive cues to be less distinctive overall. In such a scenario, increasing articulatory effort to enhance this contrast would be worthwhile in words with minimal pair competitors defined for the relevant cue, but not for words without such a competitor.

What is more, the communicative aspect of this story is likely playing a further role. The studies that have found contrastive effects most consistently have involved (i) communication with a (sometimes simulated) human partner or (ii) clarifications of previous utterances (e.g.,

[3] Unfortunately, most of the relevant studies have investigated only voiced or voiceless stops, but not both. The one exception is Schertz (2013), who reports changes to voice onset time for each condition, but not raw voice onset time values. Consequently, it is difficult to test this hypothesis based on the existing data.

Baese-Berk & Goldrick, 2009, study 2; Kirov & Wilson, 2012; Schertz, 2013; Seyfarth et al., 2016; Buz et al., 2016). These conditions highlight the possibility that hyperarticulation is part of a communicative goal (Buz et al., 2016; Hall et al., submitted). Our conversational data similarly involve communication with an interlocutor (see also Wedel et al., submitted). In contrast, studies that have not found evidence of contrastive hyperarticulation of the kind reported here have largely used word or sentence reading tasks to elicit productions (e.g., Goldrick et al., 2013; Fox et al., 2015; but see Ohala, 1994; for studies finding minimal pair effects using list reading paradigms, see Baese-Berk & Goldrick, 2009, study 1; Peramunage et al., 2011). It is possible that, in such tasks, cue-defined minimal pair competitor existence fails to predict contrastive hyperarticulation as reliably as in more communicative tasks because there would be little communicative benefit to this hyperarticulation[4].

5.4 Implications for sound change

In what follows, we assume that consistent utterance-level bias in phonetic output can shift long-term mental representations, which can in turn shape the trajectory of sound change in a speech community (discussed in Wedel, 2012; Seyfarth, 2014). In this dataset, we observed that only competitors specifically distinguished from target words by the voice onset time cue robustly predicted contrastive hyperarticulation of that cue. This suggests in turn that contrastive hyperarticulation – and associated longer-term sound change – will be best predicted by the finer grained, cue-specific minimal pair relationships in the lexicon, rather than more abstract neighborhood relationships. Conversely, it suggests that more general measures of the functional load of phonetic cues such as phoneme-level entropy (Hockett, 1967; Surendran & Niyogi, 2006) will be predictive of sound change only through their correlation with the probability that a cue defines a minimal pair contrast. This is consistent with crosslinguistic evidence that phoneme contrasts distinguishing more minimal pairs are significantly less likely to merge over time (Wedel et al., 2013).

Further, the finding that competition results in significant shortening of voice onset time for voiced stops supports the notion that contrastive hyperarticulation creates greater perceptual distance to a competitor (Seyfarth et al., 2016; Buz et al., 2016; Wedel et al., submitted; see Hall et al., submitted, for discussion), rather than resulting in a general exaggeration of duration or extent of articulatory gestures. Slow or clear speech conditions, which are associated with increased phonetic durations, have not been found to correlate with shortening of voice onset time for initial voiced stops (Miller et al., 1986; Kessinger & Blumstein, 1997). This in turn suggests that phonetic distinctions that are on average more confusable should be more likely to trigger contrastive hyperarticulation. As reviewed above, this is supported by evidence that, for voiceless stops, contrastive hyperarticulation primarily suppresses shorter voice onset time productions that would be close to the voiced-voiceless category boundary (Buz et al., 2016). Taken together, these two observations predict that contrastive hyperarticulation will exert a greater influence on the trajectory of change for cue-distinctions (i) which are more perceptually confusable, and (ii) which distinguish a greater number of cue-specific minimal pairs. A

---

[4] In the case of Fox et al., (2015), greater phonological neighborhood density was nonetheless predictive of longer voice onset time realizations (see also Fricke et al., 2016). Alongside previous findings correlating neighborhood density with longer phonetic durations (see Fricke, 2013, chs. 2 & 6 for discussion), this raises the question of whether cue-defined minimal pair competition and overall neighborhood density index different kinds of competition and, consequently, different kinds of hyperarticulation.

corollary of this prediction is that contrastive-hyperarticulation is more likely to exert an effect through modulating production of reduced/rapid speech, where phonetic distinctions tend to be reduced, rather than through effects on careful/slow speech, where phonetic distinctions tend to be more robust.

5.5 Conclusion

We have argued that voice onset time is contrastively hyperarticulated in conversational speech in a way that increases the perceptual distance between lexical minimal pairs defined specifically for the voice onset time cue. The effect of minimal pair existence was more robust for voiced stops, for which voice onset times were found to decrease under competition. These results are most consistent with models of hyperarticulation effects that consider the fine-grained phonetic relationships among competitors, such as listener-oriented models that consider the communicative goals of speakers (e.g., to be understood). Furthermore, this is consistent with prior work arguing that fine-grained difference in individual-level productions of lexical minimal pairs can help to explain patterns of sound change such as phoneme merger (Wedel et al., 2013). In addition, we noted that prior work has emphasized (i) voiceless stop hyperarticulation, for which increased voice onset times are predicted under both contrastive and clear-speech hyperarticulation, and (ii) elicited speech paradigms, some of which may not provide motivation for contrastive hyperarticulation in speakers (e.g., word list or sentence reading tasks), and which may instead promote clear-speech hyperarticulation more generally. We argued that such studies are not ideally suited to inducing or detecting contrastive hyperarticulation, potentially explaining why some of these studies find small effects, but others do not. Furthermore, we noted that the pattern of results in the literature and in the data presented here suggest the importance of considering language as a tool for communication, where speakers hyperarticulate cues that facilitate effective communication, but are less likely to hyperarticulate when there is little benefit to doing so.

Acknowledgements

We would like to thank Michael Hammond, Adam Ussishkin, Elise Bell, Benjamin Martin, Rachel Brown, and Amy Fountain for comments and feedback on an earlier version of this paper. We would also like to thank Bodo Winter for his insights and suggestions. Finally, we thank our anonymous reviewers, whose feedback was invaluable and greatly improved the paper.

Appendix A

Homographs retained in final source lexicon

abuse	*advocate*	*allied*	*alternate*	*appropriate*	*articulate*	*bass*
bow	*buffet*	*close*	*combine*	*compact*	*complex*	*compound*
concert	*conduct*	*conflict*	*console*	*content*	*contract*	*convict*
decrease	*desert*	*dove*	*intimate*	*invalid*	*laminate*	*lead*
learned	*live*	*minute*	*moped*	*object*	*polish*	*present*
produce	*progress*	*read*	*rebel*	*record*	*refuse*	*resign*
resume	*separate*	*subject*	*tear*	*use*	*wind*	*wound*

List of English homographs manually retained in our source lexicon. Additional homographs not present on this list either involve inflectional variants (e.g., *confines* (n.) ~ *confines* (v., inflected)), have pronunciations varying only in stress, which in our phonemic representations do not differ (e.g., *exploit* (n.) ~ *exploit* (v.)), or involve a form not represented in CMU at all (e.g., *blessed* (v.) ~ *blessed* (adj., not represented)).

Appendix B

Factor		Estimate	Std. Error	*t*-value
(Intercept)		0.241	0.013	18.654
MPexist = TRUE		-0.036	0.009	-4.146
Stop Phoneme				
	b	-0.107	0.008	-13.511
	g	0.034	0.010	3.417
SpeechRate		0.026	0.017	1.499
FollLiquid = TRUE		0.086	0.008	10.671
ContextDiversity		-0.010	0.015	-0.658
logBiphoneProb		-0.007	0.039	-0.174
logForBigramProb		0.039	0.010	3.864
logBackBigramProb		-0.012	0.010	-1.232
NumSyllables = 2		-0.007	0.007	-0.986
PrevMen = TRUE		0.006	0.004	1.335
Syntactic Category				
	Noun	-0.003	0.009	-0.354
	Adverb	0.017	0.013	1.261
	Verb	0.024	0.011	2.114

B1a. Summary of fixed effects for the model including minimal pair existence for voiced stops.

Groups	Name	Variance	Std. Dev.	Corr.
Lemma	(Intercept)	0.0009	0.030	
Speaker	(Intercept)	0.0012	0.035	
	MPexist = TRUE	0.0001	0.011	-0.71
Residual		0.0086	0.093	

B1b. Summary of random effects for the model including minimal pair existence for voiced stops.

Appendix C

Factor		**Estimate**	**Std. Error**	***t*-value**
(Intercept)		0.557	0.015	37.98
MPexist = TRUE		0.023	0.010	2.34
Stop Phoneme				
	k	-0.018	0.008	-2.26
	p	-0.137	0.008	-17.23
SpeechRate		-0.033	0.015	-2.21
FollLiquid = TRUE		0.039	0.007	5.88
ContextDiversity		-0.019	0.013	-1.46
logBiphoneProb		-0.001	0.036	-0.03
logForBigramProb		0.007	0.009	0.79
logBackBigramProb		-0.012	0.009	-1.41
NumSyllables = 2		-0.0004	0.006	-0.08
PrevMen = TRUE		-0.0004	0.004	-0.12
Syntactic Category				
	Noun	0.004	0.010	0.42
	Adverb	0.015	0.023	0.65
	Verb	-0.006	0.011	-0.50

C1a. Summary of fixed effects for the model including minimal pair existence for voiceless stops.

Groups	**Name**	**Variance**	**Std. Dev.**	**Corr.**
Lemma	(Intercept)	0.0010	0.032	
Speaker	(Intercept)	0.0015	0.039	
	MPexist = TRUE	0.0004	0.020	0.08
Residual		0.0103	0.101	

C1b. Summary of random effects for the model including minimal pair existence for voiceless stops.

Appendix D

Factor		**Estimate**	**Std. Error**	***t*-value**
(Intercept)		0.228	0.012	18.883
MP neighb. freq.		0.084	0.020	4.296
Stop Phoneme				
	b	-0.108	0.008	-13.771
	g	0.033	0.010	3.383
SpeechRate		0.026	0.017	1.493
FollLiquid = TRUE		0.085	0.008	10.603
ContextDiversity		-0.013	0.015	-0.906
logBiphoneProb		-0.008	0.039	-0.202
logForBigramProb		0.039	0.010	3.863
logBackBigramProb		-0.012	0.010	-1.252
NumSyllables = 2		-0.007	0.007	-1.007
PrevMen = TRUE		0.006	0.004	1.342
Syntactic Category				
	Noun	-0.003	0.009	-0.362
	Adverb	0.018	0.013	1.318
	Verb	0.025	0.011	2.218

D1a. Summary of fixed effects for the model including minimal pair neighborhood frequency for voiced stops.

Groups	**Name**	**Variance**	**Std. Dev.**	**Corr.**
Lemma	(Intercept)	0.0009	0.030	
Speaker	(Intercept)	0.0010	0.032	
	MP neighb. freq.	0.0005	0.023	0.64
Residual		0.0086	0.093	

D1b. Summary of random effects for the model including minimal pair neighborhood frequency for voiced stops.

Appendix E

Factor	Estimate	Std. Error	*t*-value
(Intercept)	0.554	0.015	37.90
Overall neighb. freq.	-0.048	0.015	-3.26
Stop Phoneme			
k	-0.015	0.008	-1.91
p	-0.132	0.008	-16.52
SpeechRate	-0.034	0.015	-2.29
FollLiquid = TRUE	0.043	0.007	6.47
ContextDiversity	-0.021	0.013	-1.65
logBiphoneProb	-0.003	0.035	-0.08
logForBigramProb	0.008	0.009	0.92
logBackBigramProb	-0.014	0.008	-1.61
NumSyllables = 2	0.007	0.006	1.20
PrevMen = TRUE	-0.0008	0.004	-0.22
Syntactic Category			
Noun	0.004	0.010	0.37
Adverb	0.012	0.023	0.55
Verb	-0.008	0.011	-0.67

E1a. Summary of fixed effects for the model including overall neighborhood frequency for voiceless stops.

Groups	Name	Variance	Std. Dev.	Corr.
Lemma	(Intercept)	0.0010	0.031	
Speaker	(Intercept)	0.0016	0.040	
	Overall neighb. freq.	0.0013	0.036	-0.16
Residual		0.0102	0.101	

E1b. Summary of random effects for the model including overall neighborhood frequency for voiceless stops.

Factor	Estimate	Std. Error	*t*-value
(Intercept)	0.558	0.015	37.13
Overall neighb. freq.	-0.043	0.026	-1.68
Stop Phoneme			
k	-0.013	0.008	-1.64
p	-0.132	0.008	-15.71
SpeechRate	-0.030	0.016	-1.92
FollLiquid = TRUE	0.049	0.007	6.66
ContextDiversity	-0.018	0.014	-1.33
logBiphoneProb	-0.015	0.037	-0.40
logForBigramProb	0.006	0.009	0.68
logBackBigramProb	-0.015	0.009	-1.65
NumSyllables = 2	0.009	0.006	1.41
PrevMen = TRUE	0.001	0.004	0.20
Syntactic Category			
Noun	-0.003	0.011	-0.25
Adverb	-0.001	0.025	-0.04
Verb	-0.014	0.012	-1.12

E2a. Summary of fixed effects for the model including overall neighborhood frequency for voiceless stops when words with no single-phoneme edit distance neighbors are removed.

Groups	Name	Variance	Std. Dev.	Corr.
Lemma	(Intercept)	0.0010	0.032	
Speaker	(Intercept)	0.0013	0.037	
	Overall neighb. freq.	0.0014	0.037	-0.95
Residual		0.0103	0.102	

E2b. Summary of random effects for the model including overall neighborhood frequency for voiceless stops when words with no single-phoneme edit distance neighbors are removed.

References

Baese-Berk, M. M. & Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes,* 24, 527-554.

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, 255-278.

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software,* 67(1), 1-48. doi:10.18637/jss.v067.i01.

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. *The Journal of the Acoustical Society of America*, *113*(2), 1001-1024.

Bell, A., Brenier, J., 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, 92–111.

Belsley, D. A., Kuh, E. & Welsch, R. E. (1980/2005). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: Wiley.

Boersma, P., & Weenink, D. (1992–2013). Praat: Doing phonetics by computer (Version 5.3.39). Amsterdam.

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, 997-990.

Burnham, K. P. & Anderson, D. R. (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods and Research,* 33(2), 261-304. doi: 10.1177/0049124104268644.

Buz, E., Tanenhaus, M. K., & Jaeger, T. F. (2016). Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers' subsequent pronunciations. *Journal of Memory and Language*, *89*, 68-86.

Buz, E., & Jaeger, T. F. (2016). The (in)dependence of articulation and lexical planning during isolated word production. *Language, Cognition and Neuroscience*.

Carnegie Mellon University. (2015). CMUdict: The CMU Pronouncing Dictionary 0.7. http://www.speech.cs.cmu.edu/cgi-bin/cmudict.

Caselli, N. K., Caselli, M. K., & Cohen-Goldberg, A. M. (2015). Inflected words in production: Evidence for a morphologically rich lexicon. *The Quarterly Journal of Experimental Psychology,* 1-23.

Chamberlin, T. C. (1890). The method of multiple working hypotheses. *Science*, *15*(366), 92-96.

Cho, T., Lee, Y., & Kim, S. (2011). Communicatively driven versus prosodically driven hyper-articulation in Korean. *Journal of Phonetics, 39*(3), 344-361.

Cieri, C., Graff, D., Kimball, O., Miller, D., & Walker, K. (2005). Fisher English Training Part 2. Linguistic Data Consortium, Philadelphia.

Davies, M. (2012). Corpus of Contemporary American English (1990-2012). Available at http://corpus.byu.edu/coca/.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological review*, *93*(3), 283.

Durian, D. (2012). A New Perspective on Vowel Variation across the 19th and 20th Centuries in Columbus, OH. Doctoral Dissertation: The Ohio State University.

Fox, N. P., Reilly, M., & Blumstein, S. E. (2015). Phonological neighborhood competition affects spoken word production irrespective of sentential context. *Journal of Memory and Language*, 83, 97–117.

Fricke, M. D. (2013). *Phonological encoding and phonetic duration* (Doctoral dissertation). University of California, Berkeley.

Fricke, M. D., Baese-Berk, M. M., & Goldrick, M. (2016). Dimensions of similarity in the mental lexicon. *Language, Cognition and Neuroscience*, *31*(5), 639-645.

Gahl. S. (2015). Lexical competition in vowel articulation revisited: Vowel dispersion in the Easy/Hard database. *Journal of Phonetics,* 49, 96–116.

Gahl, S., & Strand, J. F. (2016). Many neighborhoods: Phonological and perceptual neighborhood density in lexical production and perception. *Journal of Memory and Language*, *89*, 162-178.

Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of memory and language*, *66*(4), 789-806.

Goldrick, M., Folk, J. R., & Rapp, B. (2010). Mrs. Malaprop's neighborhood: Using word errors to reveal neighborhood structure. *Journal of Memory and Language, 62*(2), 113-134.

Goldrick, M., & Rapp, B. (2007). Lexical and post-lexical phonological representations in spoken production. *Cognition*, *102*(2), 219-260.

Goldrick, M., Vaughn, C., & Murphy, A. (2013). The effects of lexical neighbors on stop consonant articulation. *Journal of the Acoustical Society of America, 134*, EL172-EL177.

Hall, K. C., Hume, E., Jaeger, T. F., & Wedel, A. (submitted). The Message Shapes Phonology.

Jaeger, T. F. (2013). Production preferences cannot be understood without reference to communication. *Frontiers in psychology*, *4*, 230.

Jaeger, T. F. & Ferreira, V. (2013). Seeking predictions from a predictive framework. *Behavioral and Brain Sciences, 36*(4), 359-360.

Jaeger, T. F. & Buz, E. (2016). Signal reduction and linguistic encoding. In E. M. Fernández & H. S. Cairns (Eds.), *Handbook of psycholinguistics*. Wiley-Blackwell.

de Jong, K., Beckman, M. E., & Edwards, J. (1993). The interplay between prosodic structure and coarticulation. *Language and Speech,* 36, 197-212.

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

Kessinger, R. H., & Blumstein, S. E. (1998). Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies. *Journal of Phonetics*, *26*(2), 117-128.

Kharlamov, V. (2014). Incomplete neutralization of the voicing contrast in word-final obstruents in Russian: Phonological, lexical, and methodological influences. *Journal of Phonetics, 43*, 47-56.

Kirov, C., & Wilson, C. (2012). The specificity of online variation in speech production. *34th Annual meeting of the Cognitive Science Society,* Sapporo, Japan.

Kondaurova, M. V., & Francis, A. L. (2008). The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. *The Journal of the Acoustical Society of America, 124,* 3959–3971. http://dx.doi.org/10.1121/1.2999341.

Labov, W., Ash, S., & Boberg, C. (2006). The atlas of North American English: Phonetics, phonology, and sound change. Berlin: Mouton de Gruyter.

Lisker, L. (1986). "Voicing" in English: a catalogue of acoustic features signaling/b/versus/p/in trochees. *Language and speech*, 29(1), pp??

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

Liu, R. & Holt, L. L. (2015). Dimension-based statistical learning of vowels. Journal of Experimental Psychology: Human Perception & Performance. ePub ahead of print.

Luce, P. A. & Pisoni, D.B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing,* 19, 1-36.

Lukacs, P. M., Thompson, W. L., Kendall, W. L., Gould, W. R., Doherty, P. F., Burnham, K. P., & Anderson, D. R. (2007). Concerns regarding a call for pluralism of information theory and hypothesis testing. *Journal of Applied Ecology,* 44, 456-460.

Mazerolle, M. J. (2016). AICcmodavg: Model selection and multimodel inference base on (Q)AIC(c). R package version 2.0-4. http://CRAN.R-project.org/package=AICcmodavg.

McMurray, B., Tanenhaus, M., and Aslin, R. (2002). Gradient effects of within-category phonetic variation on lexical access, *Cognition,* 86(2), B33-B42. PMCID: PMC2630474.

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*(1-3), 106–115.

Munson, B. (2007). Lexical access, lexical representation, and vowel production. In J. Cole & J. I. Hualde (Eds.), *Laboratory phonology 9: Phonology and phonetics,* 201–227. Berlin: Mouton.

Munson, B., & Solomon, N. P. (2004). The effect of phonological neighborhood density on vowel articulation. *Speech, Language, and Hearing Research,* 47, 1048–1058.

O'brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, *41*(5), 673-690.

Ohala, J. J. (1994). Acoustic study of clear speech: a test of the contrastive hypothesis. In *Proceedings of the International Symposium on Prosody,* 75–89, Yokohama, Japan.

Peramunage, D., Blumstein, S. E., Myers, E. B., Goldrick, M., & Baese-Berk, M. M. (2011). Phonological neighborhood effects in spoken word production: An fMRI study. *Journal of Cognitive Neuroscience, 23*(3), 593–603.

Peterson, R. R., & Savoy, P. (1998). Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 539-557.

Picheney, M., Durlach, N. & Braida, L. (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* (Vol. 7, pp. 101-139). Berlin: Mouton de Gruyter.

Pitt, M., Johnson, K., Hume, E., Kiesling, S., Raymond, W. (2005). The Buckeye Corpus of conversational speech: labeling conventions and a test of transcriber reliability. *Speech Communication,* 45, 90-95.

Pitt, M., Dilley, L., Jonson, K., Kiesling, S., Raymond, W., Hume, E., & Foler-Lussier, E. (2007). *Buckeye Corpus of Conversational Speech (2nd release)*. Department of Psychology, Ohio State University, Columbus, OH.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Richards, S. A., Whittingham, M. J., & Stephens, P. A. (2011). Model selection and model averaging in behavioural ecology: the utility of the IT-AIC framework. *Behavioral Ecology and Sociobiology,* 65(1), 77-89.

Scarborough, R. (2012). Lexical similarity and speech production: Neighborhoods for nonwords. *Lingua*, *122*(2), 164-176.

Scarborough, R. (2013). Neighborhood-conditioned patterns in phonetic detail: Relating coarticulation and hyperarticulation. *Journal of Phonetics*, *41*(6), 491-508.

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

Seyfarth, S. (2014). Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition*, *133*(1), 140-155.

Seyfarth, S., Buz, E., & Jaeger, T. F. (2016). Dynamic hyperarticulation of coda voicing contrasts. *Journal of the Acoustical Society of America,* 139(2), EL31 - EL37.

Shadish, W. R. (1993). Critical multiplism: A research strategy and its attendant tactics. *New directions for program evaluation*, *1993*(60), 13-57.

Smiljanić, R. & Bradlow, A. R. (2008). Stability of temporal contrasts across speaking styles in English and Croatian. *Journal of Phonetics*, 36(1). 91–113.

Strand, J. F., & Sommers, M. S. (2011). Sizing up the competition: Quantifying the influence of the mental lexicon on auditory and visual spoken word recognition. *The Journal of the Acoustical Society of America*, *130*(3), 1663-1672.

Suchato, A., & Punyabukkana, P. (2005, September). Factors in classification of stop consonant place of articulation. In *INTERSPEECH* (pp. 2969-2972).

Vaden, K. I., Halpin, H. R., & Hickok, G. S. (2009). Irvine Phonotactic Online Dictionary, Version 2.0. [Data file]. Available from http://www.iphod.com.

Vitevitch, M. S., Armbrüster, J., & Chu, S. (2004). Sublexical and lexical representations in speech production: effects of phonotactic probability and onset density. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 514.

Watson, D. G., Buxó-Lugo, A., & Simmons, D. C. (2015). The effect of phonological encoding on word duration: Selection takes time. In E. Gibson & L. Frazier (Eds.), *Explicit and implicit prosody in sentence processing,* 85–98. Switzerland: Springer International Publishing. doi:10.1007/978-3-319-12961-7{_}5.

Wedel, A. B. (2006). Exemplar models, evolution and language change. *The linguistic review*, *23*(3), 247-274.

Wedel, A., Kaplan, A., & Jackson, S. (2013). High functional load inhibits phonological contrast loss: A corpus study. *Cognition, 128*, 179-186.

Wedel, A., Nelson, N. R., & Sharp, R. (submitted). The phonetic specificity of contrastive hyperarticulation in natural speech.

Wright, R. (1997). Lexical competition and reduction in speech: A preliminary report. Indiana University research on spoken language processing progress report no. 21 (pp. 471–485).

Wright, R. (2004). Factors of lexical competition in vowel articulation. In Local, J., Ogden, R. & Temple, R. (Eds.), Papers in laboratory phonology VI, 26–50. Cambridge: Cambridge University Press.

Yao, Y. (2007). Closure duration and VOT of word-initial voiceless plosives in English in spontaneous connected speech. *UC Berkeley Phonology Lab Annual Report* (183-225).

Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review, 15*(5), 971-979.

REFERENCES

Abramson, A. S. and Lisker, L. (1985). Relative power of cues: F0 shift versus voice timing. *Phonetic linguistics: Essays in honor of Peter Ladefoged*, pages 25–33.

Arnon, I. and Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, 62(1):67–82.

Baese-Berk, M. and Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes*, 24(4):527–554.

Baker, K. K., Ramig, L. O., Sapir, S., Luschei, E. S., and Smith, M. E. (2001). Control of vocal loudness in young and old adults. *Journal of Speech, Language, and Hearing Research*.

Bang, H.-Y., Sonderegger, M., Kang, Y., Clayards, M., and 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., and 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., Maechler, M., Bolker, B., and Walker, S. (2014). lme4: Linear mixed-effects models using eigen and s4. r package version 1.1-7.

Bates, D., Maechler, M., and Dai, B. (2008). The lme4 package. *Computer software manual]. Retrieved from http://cran. r-project. org/web/packages/lme4/lme4. pdf*.

Bell, A., Brenier, J. M., Gregory, M., Girand, C., and 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., and Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in english conversation. *The Journal of the Acoustical Society of America*, 113(2):1001–1024.

Belsley, D., Kuh, E., and Welsch, R. (1980/2005). Regression diagnostics: Identifying influential data and sources of collinearity.

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

Blevins, J. and Wedel, A. (2009). Inhibited sound change: An evolutionary approach to lexical competition. *Diachronica*, 26(2):143–183.

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In *Proceedings of the institute of phonetic sciences*, volume 17, pages 97–110. University of Amsterdam.

Boersma, P. and Weenink, D. (2010). *Praat: Doing phonetics by computer*.

Bradlow, A. R. (2002). Confluent talker-and listener-oriented forces in clear speech production. *Laboratory Phonology*, 7.

Breen, M., Fedorenko, E., Wagner, M., and Gibson, E. (2010). Acoustic correlates of information structure. *Language and cognitive processes*, 25(7-9):1044–1098.

Brysbaert, M. and New, B. (2009). Moving beyond kučera 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(4):977–990.

Burnham, K. P. and Anderson, D. R. (2004). Multimodel inference: understanding aic and bic in model selection. *Sociological methods & research*, 33(2):261–304.

Buz, E. and Jaeger, T. F. (2016). The (in) dependence of articulation and lexical planning during isolated word production. *Language, Cognition and Neuroscience*, 31(3):404–424.

Buz, E., Jaeger, T. F., and Tanenhaus, M. K. (2014). Contextual confusability leads to targeted hyperarticulation. In *Proceedings of the 36th annual conference of the cognitive science society*.

Buz, E., Tanenhaus, M. K., and Jaeger, T. F. (2016). Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers' subsequent pronunciations. *Journal of Memory and Language*, 89:68–86.

Calhoun, S. (2007). Predicting focus through prominence structure. In *Proceedings of Interspeech, 2007*.

Carnegie-Mellon University (1993-2015). *Carnegie-Mellon University Pronouncing Dictionary*. http://www.speech.cs.cmu.edu/cgi-bin/cmudict (accessed January 16, 2016).

Caselli, N. K., Caselli, M. K., and Cohen-Goldberg, A. M. (2015). Inflected words in production: Evidence for a morphologically rich lexicon. *The Quarterly Journal of Experimental Psychology*, 69(3):432–454.

Chamberlin, T. C. (1890). The method of multiple working hypotheses. *Science*, 15(366):92–96.

Cho, T., Lee, Y., and Kim, S. (2011). Communicatively driven versus prosodically driven hyper-articulation in korean. *Journal of Phonetics*, 39(3):344–361.

Chuoying Ouyang, I. and Kaiser, E. (2014). Prosodic encoding of informativity: Word frequency and contextual probability interact with information structure. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 36.

Cieri, C., Graff, D., Kimball, O., Miller, D., and Walker, K. (2005). Fisher english training part 2. *Linguistic Data Consortium, Philadelphia*.

Clark, E. (1980). Convention and innovation in acquiring the lexicon. *Papers and Reports on Child Language Development*, 19:1–20.

Clayards, M. (2018). Individual talker and token covariation in the production of multiple cues to stop voicing. *Phonetica*, 75(1):1–23.

Clopper, C. G. and Tamati, T. N. (2014). Effects of local lexical competition and regional dialect on vowel production. *The Journal of the Acoustical Society of America*, 136(1):1–4.

Coetzee, A. W., Beddor, P. S., Shedden, K., Styler, W., and Wissing, D. (2018). Plosive voicing in afrikaans: Differential cue weighting and tonogenesis. *Journal of Phonetics*, 66:185–216.

Cooper, W. E., Eady, S. J., and Mueller, P. R. (1985). Acoustical aspects of contrastive stress in question–answer contexts. *The Journal of the Acoustical Society of America*, 77(6):2142–2156.

Davies, M. (2009). The 385+ million word corpus of contemporary american english (1990–2008+): Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics*, 14(2):159–190.

De Jong, K. (1995). On the status of redundant features: the case of backing and rounding in american english. *B. Connell & A. Arvaniti (eds.)*, pages 68–86.

De Jong, K., Beckman, M. E., and Edwards, J. (1993). The interplay between prosodic structure and coarticulation. *Language and speech*, 36(2-3):197–212.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological review*, 93(3):283.

Dmitrieva, O., Llanos, F., Shultz, A. A., and 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.

Durian, D. (2012). *A new perspective on vowel variation across the 19th and 20th centuries in Columbus, OH*. PhD thesis, The Ohio State University.

Ernestus, M. and Warner, N. (2011). An introduction to reduced pronunciation variants. *Journal of Phonetics*, 39(3):253–260.

Ferguson, S. H. and Kewley-Port, D. (2007). Talker differences in clear and conversational speech: Acoustic characteristics of vowels. *Journal of Speech, Language, and Hearing Research*, 50(5):1241–1255.

Fox, N. P., Reilly, M., and Blumstein, S. E. (2015). Phonological neighborhood competition affects spoken word production irrespective of sentential context. *Journal of Memory and Language*, 83:97–117.

Francis, A. L., Ciocca, V., Wong, V. K. M., and Chan, J. K. L. (2006). Is fundamental frequency a cue to aspiration in initial stops? *The Journal of the Acoustical Society of America*, 120(5):2884–2895.

Fricke, M., Baese-Berk, M. M., and Goldrick, M. (2016). Dimensions of similarity in the mental lexicon. *Language, Cognition and Neuroscience*, 31(5):639–645.

Fricke, M. D. (2013). *Phonological encoding and phonetic duration*. PhD thesis, University of California, Berkeley.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and speech*, 1(2):126–152.

Gahl, S. (2015). Lexical competition in vowel articulation revisited: Vowel dispersion in the easy/hard database. *Journal of Phonetics*, 49:96–116.

Gahl, S. and Strand, J. F. (2016). Many neighborhoods: Phonological and perceptual neighborhood density in lexical production and perception. *Journal of Memory and Language*, 89:162–178.

Gahl, S., Yao, Y., and Johnson, K. (2012). Why reduce? phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*, 66(4):789.

Gandour, J. (1974). Consonant types and tone in siamese. *University of California Working Papers in Phonetics*, 27:92–117.

Geary, J. (2019). Contextual diversity, not word frequency, predicts auditory lexical decision performance: an assessment of subtlex norms for use in spoken word recognition research. In preparation.

Goldinger, S. D. and Summers, W. V. (1989). Lexical neighborhoods in speech production: A first report. *The Journal of the Acoustical Society of America*, 85(S1):S97–S97.

Goldrick, M., Folk, J. R., and Rapp, B. (2010). Mrs. malaprop's neighborhood: Using word errors to reveal neighborhood structure. *Journal of Memory and Language*, 62(2):113–134.

Goldrick, M. and Rapp, B. (2007). Lexical and post-lexical phonological representations in spoken production. *Cognition*, 102(2):219–260.

Goldrick, M., Vaughn, C., and Murphy, A. (2013). The effects of lexical neighbors on stop consonant articulation. *The Journal of the Acoustical Society of America*, 134(2):EL172–EL177.

Goy, H., Fernandes, D. N., Pichora-Fuller, M. K., and van Lieshout, P. (2013). Normative voice data for younger and older adults. *Journal of Voice*, 27(5):545–555.

Granlund, S., Hazan, V., and Baker, R. (2012). An acoustic–phonetic comparison of the clear speaking styles of finnish–english late bilinguals. *Journal of Phonetics*, 40(3):509–520.

Hall, D. C. (2007). The role and representation of contrast in phonological theory (2007). *Toronto Working Papers in Linguistics*.

Hall, D. C. (2011). Phonological contrast and its phonetic enhancement: Dispersedness without dispersion. *Phonology*, 28(1):1–54.

Hall, K. C., Hume, E., Jaeger, T. F., and Wedel, A. B. (2016). The message shapes phonology. *under review*.

Hanson, H. M. (2009). Effects of obstruent consonants on fundamental frequency at vowel onset in english. *The Journal of the Acoustical Society of America*, 125(1):425–441.

Hockett, C. F. (1966). The quantification of functional load. *Word*, 23.

Hombert, J. (1977). Consonant types, vowel height and tone in yoruba. *Studies in African Linguistics*, 8(2):173–190.

Hombert, J.-M., Ohala, J. J., and Ewan, W. G. (1979). Phonetic explanations for the development of tones. *Language*, pages 37–58.

Hoole, P. and Honda, K. (2011). Automaticity vs. feature-enhancement in the control of segmental f0. In Clements, G. N. and Ridouane, R., editors, *Where Do Phonological Features Come From?: Cognitive, Physical and Developmental Bases of Distinctive Speech Categories*, chapter 3, pages 131–171. John Benjamins Publishing, Amsterdam/Philadelphia.

House, A. S. and Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *The Journal of the Acoustical Society of America*, 25(1):105–113.

Hyman, L. M. (2013). Enlarging the scope of phonologization. In Yu, A. C. L., editor, *Origins of sound change: Approaches to phonologization*, chapter 1, pages 3–28. Oxford University Press, Oxford, UK.

Hyman, L. M. and Schuh, R. G. (1974). Universals of tone rules: evidence from west africa. *Linguistic inquiry*, 5(1):81–115.

Jaeger, T. F., Buz, E., Fernandez, E., and Cairns, H. (2016). Signal reduction and linguistic encoding. *Handbook of Psycholinguistics. Wiley-Blackwell*.

Kessinger, R. H. and Blumstein, S. E. (1997). Effects of speaking rate on voice-onset time in thai, french, and english. *Journal of Phonetics*, 25(2):143–168.

Kessinger, R. H. and Blumstein, S. E. (1998). Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies. *Journal of Phonetics*, 26(2):117–128.

Kharlamov, V. (2014). Incomplete neutralization of the voicing contrast in word-final obstruents in russian: Phonological, lexical, and methodological influences. *Journal of Phonetics*, 43:47–56.

King, R. D. (1967). Functional load and sound change. *Language*, pages 831–852.

Kingston, J. (2007). Segmental influences on f0: Automatic or controlled? In Gussenhoven, C. and Riad, T., editors, *Tones and Tunes*, volume 2, pages 171–210. Mouton de Gruyter Berlin, Germany.

Kingston, J. and Diehl, R. L. (1994). Phonetic knowledge. *Language*, 70(3):419–454.

Kirby, J. P., Ladd, D. R., et al. (2015). Stop voicing and f0 perturbations: Evidence from french and italian. *The Scottish Consortium for ICPhS*.

Kirov, C. and Wilson, C. (2012). The specificity of online variation in speech production. In *Proceedings of the 34th annual meeting of the cognitive science society. Sapporo, Japan*.

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13):1–26.

Labov, W., Ash, S., and Boberg, C. (2006). Atlas of north american english: Phonology and phonetics. *Berlin: Mouton de Gruyter*.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the h&h theory. In *Speech Production and Speech Modelling*, pages 403–439. Springer.

Lisker, L. (1986). "voicing" in english: A catalogue of acoustic features signaling/b/versus/p/in trochees. *Language and Speech*, 29(1):3–11.

Lisker, L. and Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20(3):384–422.

Lisker, L. and Abramson, A. S. (1967). Some effects of context on voice onset time in english stops. *Language and speech*, 10(1):1–28.

Löfqvist, A. (1975). Intrinsic and extrinsic f0 variations in swedish tonal accents. *Phonetica*, 31(3-4):228–247.

Löfqvist, A., Baer, T., McGarr, N. S., and Story, R. S. (1989). The cricothyroid muscle in voicing control. *The Journal of the Acoustical Society of America*, 85(3):1314–1321.

Luce, P. A. and Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, 19(1):1.

Lukacs, P. M., Thompson, W. L., Kendall, W. L., Gould, W. R., Doherty Jr, P. F., Burnham, K. P., and Anderson, D. R. (2007). Concerns regarding a call for pluralism of information theory and hypothesis testing. *Journal of Applied Ecology*, 44(2):456–460.

Martinet, A. (1952). Function, structure, and sound change. *WORD*, 8(1):1–32.

Mazerolle, M. (2016). Aiccmodavg: Model selection and multimodel inference based on (q) aic (c)[software].

McMurray, B., Tanenhaus, M. K., and Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, 86(2):B33–B42.

Mielke, J. (2004). *The emergence of distinctive features*. PhD thesis, The Ohio State University.

Miller, J. L., Green, K. P., and Reeves, A. (1986). Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast. *Phonetica*, 43(1-3):106–115.

Munson, B. (2007). Lexical access, lexical representation, and vowel production. *Laboratory Phonology*, 9:201–228.

Munson, B. and Solomon, N. (2004). The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language & Hearing Research*, 47(5):1048 – 1058.

Nelson, N. R. and Wedel, A. B. (2017). The phonetic specificity of competition: Contrastive hyperarticulation of voice onset time in conversational english. *Journal of Phonetics*.

Ohala, J. J. (1994). Acoustic study of clear speech: A test of the contrastive hypothesis. In *International symposium on prosody*, volume 18, pages 75–89.

Ohde, R. N. (1984). Fundamental frequency as an acoustic correlate of stop consonant voicing. *The Journal of the Acoustical Society of America*, 75(1):224–230.

Oldfield, R. C. and Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, 17(4):273–281.

O'brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & quantity*, 41(5):673–690.

Peramunage, D., Blumstein, S. E., Myers, E. B., Goldrick, M., and Baese-Berk, M. (2011). Phonological neighborhood effects in spoken word production: An fmri study. *Journal of Cognitive Neuroscience*, 23(3):593–603.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (1986). Speaking clearly for the hard of hearing iiacoustic characteristics of clear and conversational speech. *Journal of Speech, Language, and Hearing Research*, 29(4):434–446.

Pierrehumbert, J. (2002). Word-specific phonetics. *Laboratory phonology VII. Berlin: Mouton de Gruyter*, pages 101–40.

Pitt, M. A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., and Fosler-Lussier, E. (2007). Buckeye corpus of conversational speech (2nd release). *Columbus, OH: Department of Psychology, Ohio State University*.

Pitt, M. A., Johnson, K., Hume, E., Kiesling, S., and Raymond, W. (2005). The buckeye corpus of conversational speech: Labeling conventions and a test of transcriber reliability. *Speech Communication*, 45(1):89–95.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Richards, S. A., Whittingham, M. J., and Stephens, P. A. (2011). Model selection and model averaging in behavioural ecology: the utility of the it-aic framework. *Behavioral Ecology and Sociobiology*, 65(1):77–89.

Scarborough, R. (2012). Lexical similarity and speech production: Neighborhoods for nonwords. *Lingua*, 122(2):164–176.

Scarborough, R. (2013). Neighborhood-conditioned patterns in phonetic detail: Relating coarticulation and hyperarticulation. *Journal of Phonetics*, 41(6):491–508.

Schertz, J. (2013). Exaggeration of featural contrasts in clarifications of misheard speech in english. *Journal of Phonetics*, 41(3):249–263.

Seyfarth, S. (2014). Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition*, 133(1):140–155.

Seyfarth, S., Buz, E., and Jaeger, T. F. (2016). Dynamic hyperarticulation of coda voicing contrasts. *The Journal of the Acoustical Society of America*, 139(2):EL31–EL37.

Shadish, W. R. (1993). Critical multiplism: A research strategy and its attendant tactics. *New directions for program evaluation*, 1993(60):13–57.

Shultz, A. A., Francis, A. L., and Llanos, F. (2012). Differential cue weighting in perception and production of consonant voicing. *The Journal of the Acoustical Society of America*, 132(2):EL95–EL101.

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

Smiljanic, R. and Bradlow, A. R. (2008). Stability of temporal contrasts across speaking styles in english and croatian. *Journal of Phonetics*, 36(1):91–113.

Smiljanić, R. and Bradlow, A. R. (2008). Temporal organization of english clear and conversational speecha). *The Journal of the Acoustical Society of America*, 124(5):3171–3182.

Smiljanić, R. and Bradlow, A. R. (2009). Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Language and Linguistics Compass*, 3(1):236–264.

Sóskuthy, M. and Hay, J. (2017). Changing word usage predicts changing word durations in new zealand english. *Cognition*, 166:298–313.

Strand, J. F. and Sommers, M. S. (2011). Sizing up the competition: Quantifying the influence of the mental lexicon on auditory and visual spoken word recognition. *The Journal of the Acoustical Society of America*, 130(3):1663–1672.

Suchato, A. and Punyabukkana, P. (2005). Factors in classification of stop consonant place of articulation. In *Ninth European Conference on Speech Communication and Technology*.

Surendran, D. and Niyogi, P. (2006). Quantifying the functional load of phonemic oppositions, distinctive features, and suprasegmentals. *Amsterdam Studies in the Theory and History of Linguistic Science Series 4*, 279:43.

Terken, J. (1984). The distribution of pitch accents in instructions as a function of discourse structure. *Language and Speech*, 27(3):269–289.

Tucker, B. V., Brenner, D., Danielson, D. K., Kelley, M. C., Nenadić, F., and Sims, M. (2019). The massive auditory lexical decision (mald) database. *Behavior Research Methods*, 51(3):1187–1204.

Vaden, K. I., Halpin, H., and Hickok, G. S. (2009). Irvine phonotactic online dictionary, version 2.0 [data file]. *Available from www.iphod.com*.

Vitevitch, Michael S., A. J. and Chu, S. (2004). Sublexical and lexical representations in speech production: Effects of phonotactic probability and onset density. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(2):514–529.

Vitevitch, M. S. and Luce, P. A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in english. *Behavior Research Methods, Instruments, & Computers*, 36(3):481–487.

Wardhaugh, R. (1970). The contrastive analysis hypothesis. *TESOL quarterly*, pages 123–130.

Warner, N. (2011). Reduction. *The Blackwell companion to phonology*, pages 1–26.

Watson, D. G., Buxó-Lugo, A., and Simmons, D. C. (2015). The effect of phonological encoding on word duration: Selection takes time. In *Explicit and implicit prosody in sentence processing*, pages 85–98. Springer.

Wedel, A. (2012). Lexical contrast maintenance and the organization of sublexical contrast systems. *Language and Cognition*, 4(4):319–355.

Wedel, A., Jackson, S., and Kaplan, A. (2013a). Functional load and the lexicon: Evidence that syntactic category and frequency relationships in minimal lemma pairs predict the loss of phoneme contrasts in language change. *Language and Speech*, page 0023830913489096.

Wedel, A., Kaplan, A., and Jackson, S. (2013b). High functional load inhibits phonological contrast loss: A corpus study. *Cognition*, 128(2):179–186.

Wedel, A., Nelson, N. R., and Sharp, R. (2018). The phonetic specificity of contrastive hyperarticulation in natural speech. *Journal of Memory and Language*, 100:61–88.

Wedel, A. B. (2006). Exemplar models, evolution and language change. *The linguistic review*, 23(3):247–274.

Whalen, D., Abrahamson, A., Lisker, L., and Moody, M. (1990). Gradient effects of fundamental frequency on stop consonant voicing judgements. *Phonetica*, 1(47):36–49.

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

Wright, R. (1997). Lexical competition and reduction in speech: A preliminary report. *Research on Spoken Language Processing Progress Report*, 2.

Wright, R. (2004). Factors of lexical competition in vowel articulation. *Papers in Laboratory Phonology VI*, pages 75–87.

Yao, Y. (2007). Closure duration and vot of word-initial voiceless plosives in english in spontaneous connected speech. *UC Berkeley Phonology Lab Annual Report*, pages 183–225.

Yarkoni, T., Balota, D., and Yap, M. (2008). Moving beyond coltheart's n: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, 15(5):971–979.