English as a voicing and aspirating language: Evidence from nonnative speech perception

by

Luca Cavasso

B.A. (Spanish, French and Linguistics), University of Texas at Austin, 2016

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Arts

in the
Department of Linguistics
Faculty of Arts and Social Sciences

SIMON FRASER UNIVERSITY
Spring 2020

Approval

Name:	**Luca Cavasso**
Degree:	**Master of Arts (Linguistics)**
Title:	**English as a voicing and aspirating language: Evidence from nonnative speech perception**
Examining Committee:	**Chair:** Suzanne Hilgendorf Associate Professor
	Henny Yeung Senior Supervisor Assistant Professor
	Murray Munro Supervisor Professor
	Chandan Narayan External Examiner Associate Professor Department of Languages, Literatures and Linguistics York University
Date Defended/Approved:	April 8, 2020

Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University

or has conducted the research

c. as a co-investigator, collaborator, or research assistant in a research project approved in advance.

A copy of the approval letter has been filed with the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

Update Spring 2016

Abstract

The English contrast between fortis (i.e., /p t k/) and lenis plosives (i.e., /b d g/) is widely considered to depend on the presence or absence of aspiration. English is thus generally thought of as an “aspirating language,” as opposed to a “voicing language.” This thesis describes a study in which English listeners from across Canada rated Marathi plosives and provides results contradicting the analysis of English as an aspirating language. Native listeners of an aspirating language would be expected to rate Marathi voiceless unaspirated and voiced aspirated stops as respectively lenis and somewhat fortis-like. However, these English listeners rated them as respectively ambiguous and somewhat lenis-like. This suggests that voicing and aspiration are of similar importance to one another, contradicting the view that English is an aspirating language and, further, suggesting that English speakers may be better positioned to learn non-native laryngeal contrasts than has been thought.

Keywords: voicing; aspiration; VOT; non-native speech perception; English; Marathi

Acknowledgements

This thesis would not have been created without the help of others.

Henny Yeung, my supervisor, guided me in pursuing a research topic and sharpened my research skills immensely. It is difficult to express the full extent of his influence on my growth as a researcher, so to keep this page from dragging on I'll leave it at that.

Elise McClay gave me valuable insights into acoustic research and troubleshooting problems in R, and proofread Chapters 1 and 4 of this thesis. She has also instructed me in the creation of artistic, if not informative, graphing techniques (the more avant-garde ones do not appear here but are available upon request).

Feedback from the members of the Language Learning and Development Lab has been extremely helpful in refining the analyses and conclusions of this research.

Without the emotional support of people close to me, I don't know how I would have endured the period in which I produced this thesis. Specifically, I mean Lydia Castro, Xizi Deng, Emmy Francis, Zach Gilkison, and Noah Jozić. Each of them has made a unique, significant, and indelible impact on my life which I hope I have returned.

I also thank the staff of the Linguistics Department for making sure the department runs smoothly, especially Michelle Beninteso, Silvana Di Tosto, Kathy Godson, Debra Purdy Kong, Judi Levang, Roselyn Tam, and Clif Ng.

Table of Contents

List of Tables

List of Figures

List of Acronyms

L1	First language – in this thesis synonymous with native language
NHR	Noise-to-harmonics ratio
PAM	Perceptual Assimilation Model
SLM	Speech Learning Model
VOT	Voice-onset time

Glossary

Aspirates, or aspirated plosives	These terms interchangeably refer to plosives with aspiration (see Aspiration).
Aspiration	Turbulent airflow that is part of the longer release of English fortis stops, such as the “t” at the beginning of “tap.”
Fortis stops	Traditionally described as “voiceless” stops – a laryngeal category that is aspirated for an aspirating language, or voiceless for a voicing language. In English, fortis stops include /p, t, k/.
Inaspirates, or unaspirated plosives, or plain plosives	These terms interchangeably refer to plosives without aspiration (see Aspiration).
Laryngeal category	A category of sounds distinguished by the activity of the larynx (typically, this activity is voicing or aspiration). English laryngeal categories for stops include fortis (/p, t, k/) and lenis (/b, d, g/). Marathi laryngeal stop categories include voiced aspirates, voiced inaspirates, voiceless inaspirates, and voiceless aspirates. These six categories in particular involve distinctions between voicing and aspiration.
Lenis stops	Traditionally described as “voiced” stops – a laryngeal category that is unaspirated for an aspirating language, or voiced for a voicing language. In English, lenis stops include /b, d, g/.
Onset F0	The fundamental frequency at the onset of voicing. Measured after a plosive as a cue to laryngeal category membership.
Prevoicing	Periodic glottal pulses before the release of a plosive
Voice onset time (VOT)	The difference in time between the onset of voicing and the onset of the release of a plosive.
Voicing	Periodic glottal pulses articulated as part of a sound. In this thesis, voicing is usually discussed as happening before or during the release of a plosive.

Chapter 1.

Introduction

Much prior speech research has focused on how people identify speech sounds in their native language. In this study, I investigated the identification of speech from a non-native language to gain insight into English phonetics, offering new possibilities for examining broader patterns of perception in both the native language and in non-native languages, as well as the acquisition of non-native languages. Specifically, I asked how English speakers in Canada use voicing and aspiration – which normally occur in complementary distribution in English, and are not contrastive – to identify non-native sounds that use these two features contrastively.

In the present thesis, I examine the perception of laryngeal contrasts in plosives, like /b/ versus /p/, /d/ versus /t/, or /g/ versus /k/. These sounds are typically divided between fortis plosives /p, t, k/ and lenis plosives /b, d, g/ (Beckman, Jessen, & Ringen, 2013; Hunnicutt & Morris, 2016; Lisker & Abramson, 1964, among others). Traditionally, fortis plosives are referred to as “voiceless,” and lenis plosives as “voiced,” however the presence or absence of aspiration (rather than voicing) is a more reliable indicator of category membership in English. Thus, following Beckman et al. (2013), I will refer to “fortis” and “lenis” categories rather than “voiced” and “voiceless” ones[1]. The target phonetic context for perception tasks used in this thesis, and for much of the prior literature on plosives, is at the beginning of stressed syllables, where the two English laryngeal categories are thought to differ primarily based on voicing and aspiration (Abramson & Whalen, 2017; T. Cho, Whalen, & Docherty, 2019; Lisker & Abramson, 1964). Laryngeal contrasts manifest differently in other positions (see, e.g., Abramson & Whalen, 2017; Iverson & Salmons, 1995), so those phonetic contexts will not be discussed further in this thesis.

The English contrast between fortis and lenis stops involves two laryngeal elements: voicing and aspiration (T. Cho & Ladefoged, 1999; Hunnicutt & Morris, 2016; Lisker & Abramson, 1964). A plosive is considered *voiced* if periodic vocal fold vibration

[1] I will use this terminology for all languages with a two-way laryngeal contrast, regardless of whether voicing, aspiration, or both are involved.

begins before the release of the plosive; in this phonetic context, this kind of voicing is called prevoicing or voicing lead (Beckman et al., 2013). If the onset of periodic vocal fold vibration occurs during or after the release of the plosive, it is considered *voiceless* (Beckman et al., 2013). A plosive is *aspirated* if the vocal folds are held open after release, allowing a burst of turbulent airflow to occur before the onset of periodic voicing which marks the subsequent vowel (Iverson & Salmons, 1995). Otherwise, it is *unaspirated*. In English: Fortis stops (/p, t, k/) are consistently aspirated, and lenis stops (/b, d, g/) are consistently unaspirated (Abramson & Whalen, 2017; T. Cho et al., 2019; Klatt, 1973; Lisker & Abramson, 1964). As for voicing, fortis stops in English are consistently voiceless, while lenis stops are generally voiceless but sometimes voiced (Abramson & Whalen, 2017; Klatt, 1973; Lisker & Abramson, 1964). In English, these cues to stop voicing are typically in complementary distribution, and thus redundant in natural speech (Lisker, 1986; Lisker & Abramson, 1964). For example, no English stop consonant is both voiced and aspirated. Due to this redundancy, and because aspiration is more consistently contrastive than voicing, English is generally considered to be an *aspirating* language, rather than a *true voicing* language (Beckman et al., 2013; Honeybone, 2012; Iverson & Salmons, 1995; Jessen & Ringen, 2002).

Laryngeal contrasts between voicing and aspiration are generally measured using two acoustic cues: voice-onset time (VOT) and onset F0[2] (Abramson & Lisker, 1985; Abramson & Whalen, 2017; T. Cho et al., 2019; Dmitrieva & Dutta, 2019; Dutta, 2007; Flege, 1982; Klatt, 1973; Lisker & Abramson, 1964; Ohde, 1984). VOT is the time interval between the release of the stop and the onset of voicing (Abramson & Whalen, 2017; Lisker & Abramson, 1964). VOT is negative for stops with voicing lead (i.e., prevoicing), as the onset of voicing occurs before the release of the stop; it is short for stops with no aspiration or prevoicing, and it is long for (non-prevoiced) aspirated stops (Abramson & Whalen, 2017; Lisker & Abramson, 1964). Many other acoustic cues to laryngeal categories exist, but most are thought to result from the same gesture and are thus highly cross-correlated (Lisker, 1986). VOT has been used to measure laryngeal contrasts in many languages due to its power and simplicity (Abramson & Whalen, 2017; T. Cho et al., 2019; Lisker & Abramson, 1964). Onset F0, or the fundamental frequency at the onset of the following vowel, has also emerged as an important acoustic cue

[2] Some authors (e.g., T. Cho et al., 2019; Kingston & Diehl, 1994) instead refer to this cue as "CF0," or "consonant-induced F0" (T. Cho et al., 2019, p. 57).

(Abramson & Lisker, 1985; T. Cho et al., 2019; Haggard, Ambler, & Callow, 1970; Ohde, 1984) and is able to cue the English fortis-lenis contrast on its own (Haggard et al., 1970). Specifically, onset F0 is higher for fortis than lenis stops (Dmitrieva, Llanos, Shultz, & Francis, 2015; Haggard et al., 1970; Kingston & Diehl, 1994; Ohde, 1984). Cross-linguistically, this cue has been associated with laryngeal contrasts in both aspirating and true voicing languages (Dmitrieva et al., 2015; Kingston & Diehl, 1994), as well as languages with more complex laryngeal contrasts (Dmitrieva & Dutta, 2019; Dutta, 2007), indicating that it is not a redundant measure to VOT (Dmitrieva & Dutta, 2019; Dmitrieva et al., 2015; Haggard et al., 1970). The relationship between the English fortis-lenis contrast and these two cues is summarized in Table 1.1.

Table 1.1. Acoustic cues to English laryngeal categories

Category	VOT	Onset F0
Lenis (/b, d, g/)	Negative or short	Low
Fortis (/p, t, k/)	Long	High

Synthetically altered speech can be used to manipulate acoustic cues to reveal a pattern of trading relations between them (Repp, 1982). For instance, Repp (1979) found that increasing the amplitude of aspiration noise by 1 dB produces similar perceptual results to increasing the VOT by 0.43 ms, on average. When a plosive with contradictory cues is presented to a listener, they tend to prefer one cue over another; for example, when VOT and onset F0 conflict, English listeners use VOT rather than onset F0 to determine the laryngeal category of a consonant (Abramson & Lisker, 1985; Francis, Kaganovich, & Driscoll-Huber, 2008). Yet, while the relationship between onset F0 and VOT is well-documented, so far prevoicing and aspiration have not been manipulated to contradict one another, so it is unclear how English listeners would resolve such a conflict. This thesis therefore investigates how English listeners perceive sounds that have prevoicing (i.e., negative VOT, associated with lenis stops) and aspiration (specifically, late onset of modal voicing relative to the release, associated with fortis stops). I have used natural Marathi speech to explore this issue as Marathi's phonological inventory includes prevoiced, aspirated plosives, obviating the need to synthesize sounds.

Just as English has a two-way laryngeal contrast for plosives, other languages – including many South Asian languages, such as Marathi, Hindi, or Bengali – have three-

or four-way contrasts that involve many of the same cues that distinguish stop voicing in English (Berkson, 2012; Dmitrieva & Dutta, 2019; Lisker & Abramson, 1964; Mikuteit & Reetz, 2007). Specifically, I explored Canadian English speakers' perception of Marathi oral stops, which have a four-way contrast between prevoicing and aspiration, and which involve the same acoustic cues discussed above in the English voicing contrast. This provides an opportunity to disentangle acoustic cues that are complementary and redundant in English, as well as the phonological features they represent.

In the following sections, I discuss English's status as an aspirating language and compare how English and Marathi use voicing and aspiration. Then I survey acoustic cues to the laryngeal categories of English and Marathi and compare the two systems. Since much of the aforementioned literature is based on production data, but the present thesis is a perceptual study, I also discuss perceptual studies of laryngeal categories. Finally, I outline the motivation for a rating study that assesses how English natives weight different acoustic cues to judge the voicing of Marathi plosives, which has implications for how the English laryngeal contrast operates.

1.1. Laryngeal categories of English and Marathi

English is generally considered to be an aspirating language whose 2-way laryngeal contrast is based on differences in aspiration rather than voicing (Beckman et al., 2013; T. Cho et al., 2019; Lisker & Abramson, 1964). This is based on observations that few English speakers show evidence of consistent prevoicing (i.e., negative VOT) in lenis stops, but do consistently contrast fortis and lenis stops by aspiration (Abramson & Whalen, 2017; Beckman et al., 2013; T. Cho et al., 2019; Lisker & Abramson, 1964). However, Southern American English speakers do consistently voice lenis stops with negative VOT (Hunnicutt & Morris, 2016; Jacewicz, Fox, & Lyle, 2009; Walker, 2020), leading Hunnicutt & Morris (2016) to suggest that Southern American English is not simply an aspirating language but an aspirating and voicing language, which has also been argued for Swedish (Helgason & Ringen, 2008) and Norwegian (Ringen & van Dommelen, 2013). This view challenges the conventionally accepted notion that languages are *either* voicing *or* aspirating.

In the generally accepted view, languages with a two-way laryngeal contrast are typically divided between voicing languages (e.g. Russian, Spanish) and aspirating

languages (e.g. German, Cantonese), depending on how they distinguish fortis and lenis stops (Beckman et al., 2013; T. Cho et al., 2019; Hunnicutt & Morris, 2016). In true voicing languages, plosives are generally unaspirated: lenis stops are voiced while fortis stops are voiceless (Beckman et al., 2013; T. Cho et al., 2019). In aspirating languages, plosives are generally voiceless: lenis stops are unaspirated and fortis stops are aspirated (Beckman et al., 2013; T. Cho et al., 2019; Iverson & Salmons, 1995). However, there is some debate over the phonological feature or features that govern this contrast, and whether voicing and aspirating languages are different at the phonetic or phonological level. Some phonologists have argued that the fortis-lenis contrast in both aspirating and true voicing languages is based on the binary feature [± voice], which is realized differently at the phonetic level in aspirating and voicing languages (Keating, 1984; Kingston & Diehl, 1994; Wetzels & Mascaró, 2001). The consensus in current mainstream phonology, however, is to treat laryngeal features as privative[3], which means that these particular features are either specified or not, rather than always being specified as either a plus or minus value (Beckman et al., 2013; Y. Y. Cho, 1990; Honeybone, 2012; Hunnicutt & Morris, 2016; Iverson & Salmons, 1995; Jessen & Ringen, 2002; Keating, 1993). In this more current account, a true voicing language has lenis stops specified for [voice] and fortis stops without a laryngeal specification, while an aspirating language has fortis stops specified for [spread glottis] (abbreviated [sg]) and lenis stops without a laryngeal specification, meaning that aspirating and voicing languages differ at the phonological level (Beckman et al., 2013; Honeybone, 2012; Iverson & Salmons, 1995; Jessen & Ringen, 2002, among others).

In contrast to this typology, Helgason & Ringen (2008) suggest that some languages may use both [voice] and [sg]. Specifically, they show that Swedish reliably uses both aspiration and voicing in its two-way laryngeal contrast, such that Swedish lenis stops are specified for [voice] and fortis stops are specified for [sg] (Helgason & Ringen, 2008), which has been corroborated by further work (Beckman, Helgason, McMurray, & Ringen, 2011). Similarly, Lesho (2018) reports that Metro Manila Philippine English shows consistent prevoicing in lenis stops and some aspiration in fortis stops. Furthermore, Ringen & van Dommelen (2013) state that [voice] and [sg] are both active in Trøndelag Norwegian; aspiration consistently contrasts lenis and fortis stops, and about half of lenis stops in their dataset are voiced. In light of this, Table 1.2 shows the

[3] There is some ongoing support for non-privative [voice], such as Bennet & Rose (2017).

feature specifications of fortis and lenis stops according to the combined typology of aspirating languages, voicing languages, and voicing and aspirating languages (Beckman et al., 2013).

Table 1.2. Laryngeal feature specification for voicing and aspirating languages

Language type	Example languages	Lenis specification	Fortis specification
Voicing	French, Russian	[voice]	[Ø]
Aspirating	German, Cantonese	[Ø]	[spread glottis]
Voicing and aspirating	Swedish, Norwegian	[voice]	[spread glottis]

[Ø] indicates that no laryngeal feature is specified

English has been analyzed as an aspirating language on the basis that prevoiced lenis stops are rare or marginal (Beckman et al., 2013; T. Cho et al., 2019; Iverson & Salmons, 1995; Lisker & Abramson, 1964), but evidence has emerged that some speakers actually do maintain a reliable voicing contrast (Hunnicutt & Morris, 2016; Jacewicz et al., 2009; Walker, 2020). Hunnicutt & Morris (2016) found that speakers of Southern American English from Alabama and Mississippi reliably contrasted both voicing and aspiration in plosives. Likewise, Jacewicz et al. (2009) found that speakers from North Carolina showed reliable voicing through closure in word-initial utterance-medial stops, while speakers from Wisconsin were less consistent. Both groups from this study (Jacewicz et al., 2009) showed stronger patterns of voicing than were found in word-initial post-pausal stops by Lisker & Abramson (1964). Walker (2020) reports that prevoiced lenis plosives are a marker of Southern dialects based on Southwest Virginia talkers; two of the four native talkers prevoiced more than 75% of their word-initial lenis stops. Based on the prevalence of prevoicing in Southern American English lenis stops, Hunnicutt & Morris (2016) argue that Southern American English shows evidence of being a voicing *and* aspirating language, specifying lenis stops for [voice] and fortis stops for [sg], not unlike the laryngeal stop contrast in Swedish (Beckman et al., 2011; Helgason & Ringen, 2008).

While the languages discussed so far have two-way laryngeal contrasts involving voicing and aspiration, Marathi has a four-way laryngeal contrast between aspiration and voicing (Berkson, 2012; Dmitrieva & Dutta, 2019; Lisker & Abramson, 1964). Although aspiration and voicing are redundant in English, in Marathi they are contrastive. Thus while English has two categories – (sometimes voiced) unaspirated plosives and

voiceless aspirated plosives – Marathi has four: voiced inaspirates, voiced aspirates[4], voiceless inaspirates, and voiceless aspirates (Dmitrieva & Dutta, 2017). Beckman et al. (2013) suggest that Hindi, which has a similar four-way contrast, specifies both [voice] and [sg]: Voiced plosives are specified for [voice] and aspirated plosives are specified for [sg], resulting in the specification shown in Table 1.3.

Table 1.3. Proposed laryngeal feature specification for Marathi

Category	Feature specification
Voiceless aspirate	[sg]
Voiceless inaspirate	[Ø]
Voiced aspirate	[voice] and [sg]
Voiced inaspirate	[voice]

[Ø] indicates that no laryngeal feature is specified

Specifically, I take the contrastive use of voicing and aspiration in Marathi stop production as an opportunity to explore how voicing and aspiration are perceived in English. According to the accepted view of English as an aspirating (and not a true voicing) language, Canadian English listeners should base their perception of Marathi plosives on aspiration, that is, the perceived specification of [sg]. But if the Canadian English laryngeal contrast involves both [voice] and [sg], in other words voicing and aspiration, Canadian English listeners should not show a preference in their perception of Marathi plosives for voicing over aspiration, or vice versa.

1.2. English perception of laryngeal contrasts

The literature on aspirating and voicing languages mostly focuses on production rather than perception, but this thesis describes a perceptual study. Speech synthesis is often used to control for different acoustic variables and test their effects on perception (e.g. Abramson & Lisker, 1985; Haggard et al., 1970; Haggard, Summerfield, & Roberts, 1981). Such work has investigated the relationship between VOT and onset F0, but other cues to the English laryngeal stop contrast have barely been studied. The literature on non-native speech perception does include work on the perception of laryngeal stop categories by English listeners however, and offers some predictions for the present study. This section will review different approaches to the perception of laryngeal

[4] These are sometimes referred to as “breathy voiced stops” due to the breathiness of their release. I will refer to them as “voiced aspirates” to emphasize their phonological properties, but both appellations have their merits.

contrasts by English listeners and offer three frameworks for understanding the perception of non-native speech.

1.2.1. Using speech synthesis

Studies of the perception of the English laryngeal contrast by native listeners generally focus on VOT and onset F0 using synthesized speech. As previously mentioned, differences in onset F0 can be sufficient to cue a listener's perception of a stop as fortis or lenis, if VOT is ambiguous (Haggard et al., 1970). This naturally leads one to question the relationship between VOT and onset F0. This line of research has shown a trading relation between the two, meaning that a more extreme onset F0 can compensate for a more ambiguous VOT (Abramson & Lisker, 1985; Haggard et al., 1981). Nevertheless, some have suggested that onset F0 is actually a secondary cue to VOT. Abramson and Lisker (1985), for example, point out that some VOT values are categorical and not subject to such influence. Haggard, et al. (1981) do not discuss this as explicitly, but their data do show the same pattern. Also, Haggard et al. (1970) note that while onset F0 differences cue the laryngeal contrast for most listeners, six of their twenty-one participants do not show evidence of incorporating onset F0 into their perception of the laryngeal contrast.

However, Francis et al. (2008) dispute that onset F0 is a secondary cue to VOT as they were able to train English listeners to prefer either cue over the other with comparable ease. They argue rather that the previously established preference for VOT is due to discrepancies in perceptual distance rather than an a priori psychological or linguistic preference for VOT itself as a cue. The primacy of VOT has also been undermined by work showing that under certain conditions, the effect of onset F0 on perception is increased, while the effect of VOT diminishes., Examples of such tasks include doing arithmetic operations while listening (Gordon, Eberhardt, & Rueckl, 1993) or attempting to make very fast judgments (Whalen, Abramson, Lisker, & Mody, 1993).

All the same, in the present study, differences in perceptual distance between VOT and onset F0 will not be controlled as the stimuli are natural speech, and listeners will be able to pay careful attention to the stimuli. So, although VOT may not be as universally dominant over onset F0 as was once thought, the design of the study at hand can be expected to show a much larger effect of VOT than onset F0 on perception.

While the effect of interactions between VOT and onset F0 on English-natives' perception of the fortis-lenis contrast has received some attention, the interaction between prevoicing and aspiration has not been investigated. Repp (1979) found a trading relation between VOT and the relative amplitude of aspiration to that of the following vowel, but did not synthesize stimuli with negative VOTs. So the speech synthesis literature indicates that VOT will likely prove a more important cue than onset F0, and it suggests that ambiguous VOTs may be swayed by such secondary cues as onset F0, but this literature offers little insight into how a conflict between prevoicing and aspiration would be perceptually resolved, which is the central question of this thesis.

1.2.2. Theories of non-native speech perception

Non-native speech perception has obvious implications for the acquisition of an additional language, and has therefore received much attention. It is also clearly relevant to the present study as it involves the perception of Marathi speech by non-native listeners. There is ample evidence that a person's first language (L1) influences their perception of non-native speech sounds (e.g., Flege & Eefting, 1986; Goto, 1971; Moon, Cooper, & Fifer, 1993; Ringbom, 1992; Strange, Levy, & Law, 2009), but it is somewhat controversial exactly what level of the L1 grammar (e.g., phonology, phonetics) is responsible for this influence. The Perceptual Assimilation Model (Best, 1995; Best & Tyler, 2007) suggests that non-native speech perception is affected by gestural, phonetic, and phonemic aspects of the L1. In contrast, the Speech Learning Model (Flege, 1987, 1995) states that non-native speech is analyzed at the phonetic, not phonemic, level, based on acoustic properties of the phones involved. Brown (1998) instead claims that interference in perception (and production) of non-native phones occurs at the featural level. In this subsection, I review these theories of non-native speech perception and discuss them in the context of English-native perception of Marathi laryngeal categories.

The Perceptual Assimilation Model (PAM[5]) predicts that non-native sounds are "perceptually assimilated" based on the articulatory similarity between L1 and non-native phones, whether at the gestural, phonetic, or phonemic level (Best, 1995; Best & Tyler, 2007). This model was initially conceived to explain patterns of discriminability between

[5] Best (2007) refers to the model incorporating second language acquisition as PAM-L2, but I will refer to all iterations of this model as PAM for the sake of simplicity.

different types of non-native phones (Best, 1995) and later expanded to predict patterns of learnability (Best & Tyler, 2007). According to PAM, when a listener hears non-native sounds, one of three things happens: the sound is *categorized*, meaning it is assimilated to an L1 speech category, whether as a good or deviant exemplar of that category; the sound is *uncategorized*, meaning it is not assimilated to any L1 speech category; or the sound is *non-assimilated*, meaning it is not understood to be speech. For example, Strange, Bohn, Nishi, and Trent (2005) found that North German /i/ was consistently *categorized* as a good exemplar of North American English /i/, while North German /ø/ was *uncategorized*. Best, McRoberts, and Sithole (1988) describe Zulu clicks as an example of sounds that are *not assimilated* as speech by English listeners. PAM makes several explicit predictions about how well pairs of sounds may be discriminated based on how they were (or were not) assimilated to L1 categories (Best, 1995; Best & Tyler, 2007), but I will not discuss these predictions in detail here as this thesis concerns the *identification* of non-native speech, while ability to discriminate between different non-native categories is not as relevant. Like the SLM, PAM predicts that uncategorized sounds will be learned faster than categorized ones, as uncategorized sounds simply require the formation of a new category, while categorized sounds also require existing L1 categories to be adjusted for the new language (Best & Tyler, 2007). It is likely that Marathi voiced aspirates will be uncategorized by English listeners due to the presence of both aspiration and voicing. Voiceless inaspirates will likely be categorized as English lenis stops due to their lack of prevoicing and of aspiration, and voiceless aspirates will likely be categorized as English fortis stops due to their similar voicelessness and aspiration. Voiced inaspirates will likely be categorized as lenis stops, though their prevoicing may make them poor exemplars of this category. These predictions are summarized, along with SLM-based predictions, below in Table 1.4.

The Speech Learning Model (SLM) posits that the learning of new speech sounds is based on acoustic similarities at the allophonic (not phonemic) level (Flege, 1987, 1995). While this model ”is concerned primarily with the ultimate attainment of L2 pronunciation,” (Flege, 1995, pp. 237–238) it does touch on the initial stages of non-native perception as well (Flege, 1987, 1995). Most relevant to this thesis, it predicts that listeners will identify non-native sounds as “new,” “similar,” or “identical,” depending on the phone’s relationship to the L1 (Flege, 1987). “New” sounds cannot be identified as being the counterpart of any native phone, such as French /y/ heard by English listeners,

"similar" ones are comparable to a native phone but with a systematic difference, such as French /t/ heard by English listeners, and "identical" sounds are acoustically identical between the two languages (Flege, 1987). The SLM predicts that, aside from "identical" sounds, "new" sounds will be learned the most quickly as it is relatively easy to identify the differences between them and native phones (Flege, 1987, 1995). Based on the SLM, it is reasonable to expect that Marathi voiced aspirates will be perceived as "new" sounds since they have both prevoicing and aspiration, which are normally complementary and redundant in English. Likewise, voiceless aspirates are likely "identical" to English fortis stops and voiceless inaspirates might be considered "identical" to English lenis stops. Voiced inaspirates are probably "similar" to English lenis stops, as they have no aspiration but do have prevoicing, while English lenis stops only sometimes have prevoicing. These patterns, and predictions based on PAM, are shown in Table 1.4.

Table 1.4. Likely English perception of Marathi sounds based on SLM and PAM

Sound	SLM	PAM
Voiced aspirate	New	Uncategorized
Voiced inaspirate	Similar	Categorized (possibly deviant)
Voiceless inaspirate	Identical	Categorized (good exemplar)
Voiceless aspirate	Identical	Categorized (good exemplar)

Brown (1998) proposed that the acquisition of non-native speech (in perception and production) is mediated by the features present in the L1. Where PAM and the SLM focused more on identifying patterns of perception and production of non-native phones, Brown's (1998) model aims to provide a clear mechanism for such interference. So, for example, in light of Japanese listeners' well-known difficulties in discriminating English /l/ from /r/ (Goto, 1971, among others), Brown (1998) posits that this is due to the fact that the Japanese liquid does not have a [coronal] feature, which in English distinguishes /l/ from /r/. Further, Brown (1998) shows that Mandarin Chinese listeners, who do have a [coronal] feature specified for their liquids, can identify the two sounds successfully. From this model, and keeping in mind the consensus that English uses the feature [spread glottis] to distinguish its laryngeal stop categories, English listeners can be expected to identify Marathi stops without [sg] as lenis, and those that do have it as fortis. In other words, Marathi aspirates would be identified as similar to English fortis

stops and Marathi inaspirates would be identified as similar to English lenis stops, as shown in Table 1.5.

Table 1.5. Predicted English perception of Marathi stops, according to Brown's (1998) model

Marathi category	Features*	Likely English category
Voiced aspirate	[voice], [sg]	Fortis ([sg])
Voiced inaspirate	[voice]	Lenis (no [sg])
Voiceless inaspirate	[Ø]	Lenis (no [sg])
Voiceless aspirate	[sg]	Fortis ([sg])

*Feature specifications shown for Marathi identical to those in Table 1.2, based on Beckman et al. (2013)

This thesis will use PAM's terminology as it was designed explicitly to explain naïve perception of non-native speech (whereas the SLM is generally more concerned with advanced learners). However, while PAM does not endorse analyses based on acoustic similarity due to its foundations in direct realism (Best, 1995; Best & Tyler, 2007), I will nonetheless incorporate acoustic analyses as there is evidence that acoustic information is highly relevant to perception generally (e.g. Ohala, 1996), laryngeal contrasts in particular (Abramson & Whalen, 2017; Berkson, 2012; T. Cho & Ladefoged, 1999; Dmitrieva & Dutta, 2019, among others) and, even more specifically, to the perception of non-native laryngeal contrasts (Guion & Pederson, 2007; Jackson, 2009). I also take acoustic information such as VOT and measurements of aspiration as implicating phonological features [voice] and [spread glottis], in order to make use of Brown's (1998) predictions.

1.2.3. Studies of English perception of non-native laryngeal contrasts

To my knowledge, no study has specifically investigated English listeners' perception of Marathi stops. However, the laryngeal stop contrast in Marathi and Hindi is fairly similar in terms of VOT (Lisker & Abramson, 1964) and is also similar in that both languages contrast voicing and aspiration (Dmitrieva & Dutta, 2019). Some studies have tested English listeners' perception of Hindi stops, and those that have tested the discrimination of Hindi laryngeal categories by English listeners are particularly relevant to this thesis.

Jackson (2009) tested English and French listeners' ability to discriminate between different Hindi laryngeal stop categories at four places of articulation: bilabial,

dental, retroflex, and velar. English listeners were found to perform better at contrasts involving the feature [spread glottis], such as /k-kʰ/, /g-gʱ/, /g-kʰ/, and /k-gʱ/, than those that did not, such as /g-k/, while French listeners showed inverted results (Jackson, 2009). Still, English listeners were able to discriminate even [voice]-based contrasts at a higher-than-chance level, though Jackson (2009) attributes this to low-level acoustic properties of the sounds in this contrast rather than to any phonological distinction. Jackson (2009) did not investigate which specific English categories listeners may have mapped the Hindi sounds to, but based on the pattern of results, it seems likely that the [spread glottis] feature would have heavily influenced, if not determined, English listeners' identification of these stops, which is in line with Brown's (1998) predictions.

Other studies have investigated effects of training. For example, initially, the Hindi /tʰ-dʱ/ contrast is difficult for English-native adult listeners (Tees & Werker, 1984; Werker, Gilbert, Humphrey, & Tees, 1981), but they are able to improve with a short training period (Tees & Werker, 1984). This improvement was shown to last 30-40 days after the initial training (Tees & Werker, 1984). This is convergent with Guion and Pederson's (2007) findings, where English listeners were trained on the discrimination of three Hindi contrasts, two of which were laryngeal contrasts: /k-g/ and /b-bʱ/. Listeners showed a surprisingly high ability to discriminate both contrasts at pre-test, scoring 89% and 87% correct responses, respectively (Guion & Pederson, 2007). This is particularly striking for the /k-g/ contrast, as the VOT values of both of these phones fall within the range of an English /g/ (Guion & Pederson, 2007). While initial discrimination of /k-g/ and /b-bʱ/ was comparable, listeners did improve more on the /b-bʱ/ contrast than /k-g/, which Guion and Pederson (2007) attribute to (Hindi) /b/ being better suited to membership in English /b/ than (Hindi) /bʱ/. Along the lines of Brown's (1998) model, I infer this is because /b/ and /bʱ/ differ based on the [spread glottis] feature, which is active in English. It is unclear, though, why Guion and Pederson's (2007) participants discriminated /k-g/ so well (89% correct) relative to Jackson's (2009), who scored an average of 63.7% correct. Comparing participants' reported linguistic backgrounds does little to illuminate this: Jackson's (2009, p. 62) English-speaking participants were "functional monolinguals… in that they were not actively using an L2 or in the process of learning an L2," and were all native English speakers residing in Calgary. Guion and Pederson's (2007) were English monolinguals with less than three years of formal non-

native language learning, who had not spent more than six months in a non-English speaking region – the location of the study is not reported.

While it is striking that Guion and Pederson's (2007) participants achieved such high pre-test discrimination accuracy, there is pre-existing evidence that English listeners can learn to identify prevoiced and voiceless unaspirated stops as separate sounds (Pisoni, Aslin, Perey, & Hennessy, 1982). Pisoni et al. (1982) were able to train English listeners to identify synthesized stops with negative, short-lag, and long-lag VOT as members of one of three categories, based on their VOT, after a 1-hour training session. In other words, English listeners were able to quickly learn to identify negative and short-lag VOT stops as different from one another, despite lacking this contrast in their native language (Pisoni et al., 1982). This contrasts with previous work (e.g. Abramson & Lisker, 1967), which suggested that linguistic experience significantly limits the perception of laryngeal contrasts. Pisoni et al. (1982) assert that in fact English listeners (for example) can very quickly learn to identify categories based on VOTs that are *not* contrastive in their native language.

Polka (1991) investigated English-native listeners' perception of Hindi place contrasts using stops of different laryngeal categories. Stops from different categories were not directly compared with one another as Polka (1991) was more interested in the place contrast, but the study included an identification task in which listeners heard a Hindi sound and wrote down an orthographic equivalent. Unfortunately this section of the report (Polka, 1991, pp. 2267–2268) contains several inconsistencies between the text and table presented, but I have taken the text as authoritative. Polka's (1991) participants all identified Hindi voiceless aspirates as "t" (*n* = 18) and sometimes also "th" (*n* = 2); voiceless inaspirates were transcribed as "d" (*n* = 18), and sometimes also "th" (*n* = 9) or "t" (*n* = 2); voiced inaspirates were transcribed as "d" (*n* = 18) and sometimes also "th" (*n* = 3); and voiced aspirates were transcribed as "t" or "d" by all participants[6], and sometimes both (*n* = 13) or "dh" (*n* = 2) or "th" (*n* = 2). This suggests that Hindi voiceless aspirates were categorized as good exemplars of English fortis stops, Hindi voiceless inaspirates were assimilated as perhaps slightly deviant English lenis stops, Hindi voiced inaspirates were assimilated as English lenis stops, and Hindi voiced

[6] It is unclear, out of participants who wrote only either "t" or "d," how many wrote "t" versus "d," due to the aforementioned discrepancies between the table presenting these data and the text describing them.

aspirates were uncategorized by English listeners. Given the similarity between Hindi and Marathi stops, we may expect a similar pattern to emerge in the rating study described by this thesis.

Taking the results of Hindi perception studies together, it is possible to make more general claims about English listeners' perception of a laryngeal contrast involving both [voice] and [spread glottis]. English listeners perceive contrasts based on the [sg] feature better than others (Jackson, 2009) and can learn contrasts involving this feature better than those based on [voice] (Guion & Pederson, 2007), though they are also able to learn [voice]-based contrasts under some training paradigms (Guion & Pederson, 2007; Pisoni et al., 1982; Tees & Werker, 1984). Yet despite being more capable of discrimination based on the [sg] feature, English listeners still show better-than-chance discrimination of [voice]-based contrasts (Jackson, 2009), sometimes much better than chance (Guion & Pederson, 2007), even without any training. So while [spread glottis] certainly plays an important role in discrimination, it has been suggested that some low-level acoustic property or properties correlated with [voice] also have an effect on English listeners' perception (Guion & Pederson, 2007; Jackson, 2009). As for identification, Polka's (1991) orthography-based identification task indicates that English listeners likely categorized Marathi sounds based primarily on their aspiration (i.e., phonetic realizations of [spread glottis]), though voiced aspirates seemed to pose a particular challenge and were likely uncategorized. This may be due to the prevoicing of voiced aspirates, or to the fact that their release is different than English aspiration (Dmitrieva & Dutta, 2019; Lisker & Abramson, 1964); these acoustic qualities will be further discussed in the following section.

1.3. Acoustic relationships between Marathi and English laryngeal contrasts

Having already discussed the acoustic cues to the English laryngeal contrast in the opening of this chapter, I will now introduce how these and other acoustic cues are used in Marathi and compare the phonetics of the laryngeal categories in the two languages. Recall that Marathi, unlike English, has a four-way contrast between voicing and aspiration, including voiced aspirates, plain (i.e., unaspirated) voiced plosives, voiceless aspirates, and plain voiceless plosives (Berkson, 2012; Dmitrieva & Dutta, 2019; Lisker & Abramson, 1964). In English, VOT is the most common way of measuring

its two-way laryngeal contrast, but VOT does not work as well for Marathi, as it fails to distinguish voiced aspirates from voiced inaspirates: both categories are prevoiced and therefore have negative VOT (Berkson, 2012; Dmitrieva & Dutta, 2019; Lisker & Abramson, 1964).

Another difference between the phonetics of laryngeal categories in Marathi and English is what constitutes "aspiration." The quality of aspiration in Marathi voiced aspirates is different than voiceless aspirates in English or Marathi. Lisker & Abramson (1964) note that the vowel following voiced aspirates takes on a breathy quality, and some acoustic evidence has been found to support this (Berkson, 2012; Dmitrieva & Dutta, 2019). This is different than aspiration in English, where breathy voicing is generally perceived as similar to modal voicing (Hillenbrand, J; Cleveland, R; Erickson, 1994; Kane & Gobl, 2011), and aspiration is phonetically realized as noisy, turbulent airflow (Klatt, 1973; Lisker & Abramson, 1964).

The relative ineffectiveness of VOT in measuring a four-way laryngeal contrast (like Marathi's) has motivated several proposals for how to measure the acoustics of such a contrast. Davis (1994) found a contrastive difference in Hindi between stop categories using a measure called "noise offset," the temporal difference between the onset of the stop release and the onset of a visible second formant band. Mikuteit & Reetz (2007) achieved similar findings for East Bengali voiced aspirates using "after closure time," the difference between the onset of the stop release and the onset of regular glottal pulsing. However, Berkson (2012) found that these measures cannot be directly applied to Marathi. Noise offset time could not distinguish all four laryngeal categories for Marathi, as Davis (1994) had shown it did for Hindi, besides which the visibility of the onset of the second formant is inconsistent depending on Marathi speakers' production and the researcher's spectrogram settings (Berkson, 2012; Mikuteit & Reetz, 2007). Additionally, not all voiced aspirates had clear landmarks to indicate after closure time (Berkson, 2012). Thus, she proposed "pre-vocalic interval" (PVI) as a durational measurement of aspiration in Marathi, defined as the difference between the onset of the stop release and the offset of "that portion of the vowel which is heavily flavored by the breathy release," which is marked by a jump in amplitude (Berkson, 2012, p. 43).

In contrast, Dmitrieva & Dutta (2019) reject approaches to the phonetics of Marathi laryngeal categories that are based on a single acoustic measure and instead prefer a multi-dimensional approach using these acoustic cues: the duration of voicing lead, the length of the stop release (similar to PVI or noise offset time), the percentage of voicing in the stop release, the f0 at the onset of voicing (whether during or after release), and the breathiness of the first 30% of the following vowel (measured as H1-H2 and H1-A1). These measures are summarized in Table 1.6, with duration of voicing lead collapsed into VOT, and breathiness of the following vowel not shown (the first 30% of the vowel is breathier after aspirated stops).

Table 1.6. Acoustic cues to Marathi laryngeal categories

Category	VOT	Onset F0	Release duration	% voiced release
Voiced aspirate	Negative	Low	Long	High
Plain voiced	Negative	Low	Short	High
Plain voiceless	Positive, short	High	Short	Low
Voiceless aspirate	Positive, long	High	Long	Low

Not shown: Measures of breathiness in the first 30% of the following vowel. Measures associated with breathiness are higher in the first 30% of vowels that follow aspirated stops (Dmitrieva & Dutta, 2019).

These acoustic cues are used differently, if at all, in English. In Marathi, prevoicing, lower onset F0, and a higher proportion of voicing during release is associated with voiced stops. In English, prevoicing and proportion of voicing during release[7] are weakly, if at all, associated with lenis stops (Beckman et al., 2013; T. Cho & Ladefoged, 1999; Lisker & Abramson, 1964). Lower onset F0 is consistently associated with lenis stops however (Abramson & Lisker, 1985; Haggard et al., 1970; Ohde, 1984). Conversely, long release duration and breathiness in the first 30% of the vowel are associated with Marathi aspirated stops. Release duration is virtually equivalent to VOT for non-prevoiced stops and thus very strongly associated with English fortis stops (T. Cho & Ladefoged, 1999; Lisker & Abramson, 1964). So onset F0 aligns the English fortis-lenis contrast with Marathi's voiceless-voiced contrast, while release duration aligns the English fortis-lenis contrast with Marathi's aspirated-unaspirated contrast, although the quality of aspiration in Marathi voiced aspirates is not the same as that of English fortis stops. Table 1.7 shows which English laryngeal category is acoustically

[7] While proportion of voicing during release has not been specifically investigated in English, it is likely high for lenis stops as Abramson and Lisker (1964) reported mean VOTs of less than 10ms for English lenis stops.

similar to each Marathi laryngeal category according to three cues. For example, the first cell indicates that Marathi voiced aspirates have similar VOT to English lenis stops.

Table 1.7. Acoustic similarity of Marathi laryngeal categories to English laryngeal categories by each of three cues

Marathi category	VOT is similar to English	Release duration is similar to English	Onset F0 is similar to English
Voiced aspirate	Lenis stops	Fortis stops*	Lenis stops
Plain voiced	Lenis stops	Lenis stops	Lenis stops
Plain voiceless	Lenis stops	Lenis stops	Fortis stops
Voiceless aspirate	Fortis stops	Fortis stops	Fortis stops

* But the spectral quality of the release is not always like English fortis stops

Given the consensus that English bases its laryngeal contrast on aspiration, which is best reflected here by release duration, we may expect the English fortis-lenis contrast to map onto the Marathi aspirated-unaspirated contrast. However, as Marathi voiced aspirates have more voicing during their release and are followed by breathier vowels (Dmitrieva & Dutta, 2019), this aspiration may not be perceived as similar to English aspiration. In Brown's (1998) model of L1 interference, phonetic information from non-native speech is translated into L1 phonological features. But if the aspiration of Marathi voiced aspirates is sufficiently distinct from English aspiration, then its phonetic properties may not be properly interpreted as demonstrating a [spread glottis] feature by English listeners. The highly variable identification of Hindi voiced aspirates by English listeners that Polka (1991) found supports that this may be the case. But this is an empirical question that the present study will address.

1.4. The present study

This thesis investigates the status of English as an aspirating language that relies on VOT by highlighting how [voice] and [spread glottis] features, mediated by acoustic cues such as VOT and onset F0, may be weighted against one another in perception. Specifically, I conducted a rating study of Marathi laryngeal categories by English listeners in Canada. Recall that Marathi voiceless aspirated and voiced unaspirated plosives use VOT, aspiration, and onset F0 similarly to English lenis and fortis categories, but voiceless unaspirated and voiced aspirated plosives do not. English listeners' perception of these latter two categories has implications for the conventional view of English (and by extension, Canadian English) as an aspirating language (and not

a true voicing language). If Canadian English is an aspirating and not a voicing language, Marathi voiceless inaspirates should be categorized as lenis stops based on their VOT. In Marathi voiced aspirates, VOT is similar to English lenis plosives, while their aspiration is more like English fortis plosives. If English listeners categorically perceive these stops as either lenis or fortis, that would provide strong evidence for English as a voicing or aspirating language (respectively). If they are judged as not resembling either category more closely than the other, that would support an analysis like Hunnicutt & Morris' (2016),that is, that Canadian English laryngeal categories use both [voice] and [sg]. If listeners in this study assimilate Marathi stops similarly to Polka's (1991) Hindi stops, this study will confirm that English is an aspirating (and not voicing) language.

Chapter 2 describes the design and collection of stimuli for the rating study and outlines the acoustics of the Marathi tokens I collected. Chapter 3 details the rating study itself and its results, which indicate that English listeners used both [voice] and [spread glottis] features to make their judgments. Chapter 4 discusses these results in the context of the phonetics and phonology of English, suggesting an analysis of English as a voicing and aspirating language and predicting the perception and acquisition of four-way laryngeal contrasts based on the laryngeal contrast of one's L1.

Chapter 2.

Stimuli: collection and acoustics

In this chapter, I outline the design and collection of stimuli for the rating study presented in Chapter 3. I also discuss the process of annotating the stimuli for acoustic analysis and explain how acoustic measures associated with laryngeal contrasts were derived from those annotations. Chapter 3 will investigate the relationship between these measures and English listeners' judgments of laryngeal category membership.

2.1. Stimuli design and collection

2.1.1. Design

The stimuli for this thesis consist of disyllabic Marathi nonce words. Each word begins with a denti-alveolar or velar plosive, followed by one of the three corner vowels of Marathi. Only long vowels were used as short vowels are often reduced. The second syllable begins with [s] followed by the same vowel. This structure was intended to ensure that the plosive was the most salient part of the nonce word for English listeners, while the rest of the word would sound relatively familiar. This stimulus structure also allowed me to investigate effects of vowel context and place of articulation on perception.

Table 2.1. Stimuli set.

	ई /i:/	**आ /a:/**	**ऊ /u:/**
क /k/	कीसी /ki:si:/	कासा /ka:sa:/	कूसू /ku:su:/
ख /kʰ/	खीसी /kʰi:si:/	खासा /kʰa:sa:/	खूसू /kʰu:su:/
ग /g/	गीसी /gi:si:/	गासा /ga:sa:/	गूसू /gu:su:/
घ /gʱ/	घीसी /gʱi:si:/	घासा /gʱa:sa:/	घूसू /gʱu:su:/

Dental plosives were also used; velars shown for convenience.

2.1.2. Talkers

Five talkers were recruited from the greater Vancouver area. A technical error during recording prevented Talker 1's tokens from being usable, so they were excluded

from the study. All talkers were born in the Maharashtra state of India (where Marathi is the official state language and local language) and immigrated to Canada as adults, except Talker 2, who normally resides in India, not Canada. All talkers reported Marathi as their native, most dominant language. They also knew Hindi and English. Talkers 3-5 informally reported that their proficiency in Hindi had declined due to lack of use, though they indicated that they still frequently used Marathi, especially with friends and family. Other aspects of talkers' linguistic backgrounds are shown in Table 2.2.

Table 2.2. Talker background information.

Talker	Age	Sex	Age of arrival in Canada	Languages known, in order of dominance
Talker 2	58	F	N/A – non-resident	Marathi, Hindi, English
Talker 3	39	M	37	Marathi, English, Hindi
Talker 4	39	F	37	Marathi, English, Hindi, Spanish
Talker 5	34	F	27	Marathi, English, Hindi

2.1.3. Materials

The talkers resided in different parts of the Vancouver area and could not all access the same location with comparable ease, so they were recorded at three different locations: a sound-attenuated room at Simon Fraser University, a sound-attenuated booth in the Vancouver Public Library, and a quiet room in a private residence. Talkers were recorded using a Yeti Blue USB microphone with a sampling frequency of 44.1 kHz and 16-bit quantization. Audacity 2.2.0 (Audacity Team, 2017) was used to reduce noise in the recordings (12 dB noise reduction, sensitivity of 6.00 with 3-band frequency smoothing), extract tokens to use as stimuli, and normalize the amplitude of the extracted tokens.

2.1.4. Procedure

Each talker received a list of short sentences, which followed the format shown in Table 2.3. The target word was preceded by a common name and followed by the word म्हणाली "said." Different names were used to make the task slightly less repetitive. Talkers were instructed to speak clearly and at a moderate pace, then they were recorded reading the list.

Table 2.3. Carrier sentence structure.

Marathi	इरा / जाई / माया	target word	म्हणाली.
IPA	/ira/ /dʑai:/ /maja/		/ mʱənʱali/
Gloss	Ira / Jai / Maya (names)	target word	said
Translation	(name) said “target word.”		

Tokens were selected from the recording based on similarity in prosody and duration. From each talker’s recording, 3 tokens of each nonce word were extracted, yielding 72 tokens per talker. Across all tokens and talkers, this yielded a total of 288 tokens.

2.2. Acoustic measurements

For each token, I annotated the first syllable using Praat TextGrids (Boersma & Weenink, 2020) for prevoicing, release, and the voiced portion of the vowel, following Dmitrieva and Dutta’s (2019) annotation scheme. I also annotated which portion of the release showed periodic glottal pulses (i.e., voicing), as Dmitrieva & Dutta (2019) found a relationship between laryngeal category membership and the percentage of voicing during release. Annotation guidelines are summarized in Table 2.4.

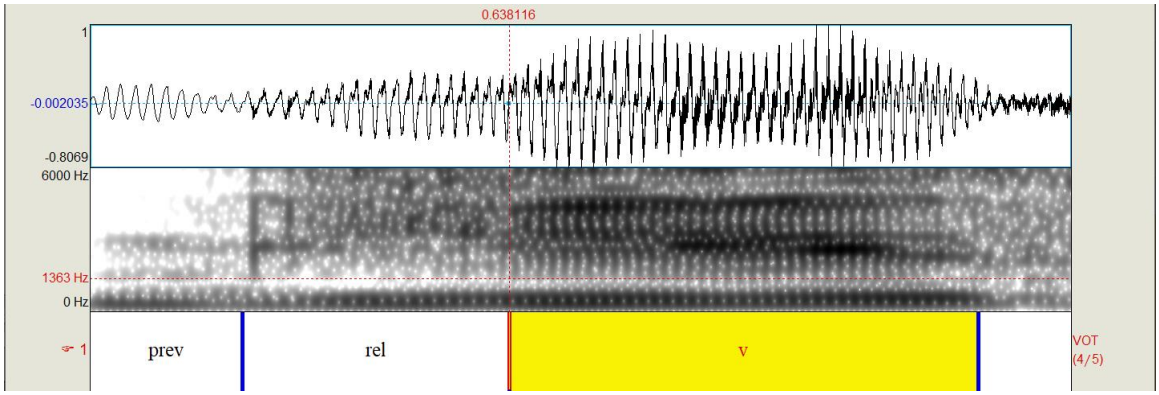
Table 2.4. Summary of annotation guidelines.

Annotation	**Basis for annotation**
Prevoicing	Low-frequency glottal pulses visible in the spectrogram and a periodic waveform
Release (voiceless)	Noisy waveform and spectrogram
Release (voiced)	Somewhat periodic waveform, low-frequency glottal pulses visible on spectrogram.
Modal voicing (vowel)	Clearer formant visibility, spike in intensity, or complex periodic waveform

Although Berkson (2012) suggested measuring release duration (PVI in her terminology) based on a spike in intensity, some tokens showed multiple spikes in intensity, or a gradual transition that made a specific “spike” difficult to pinpoint. While the onset of modal voicing was relatively easy to find for most laryngeal categories, it was more difficult to pinpoint for voiced aspirates, which Dmitrieva and Dutta (2019) also found. For example, Figure 2.1 shows a clear change in formant visibility that was taken to indicate the offset of release, concurrent with an increase in amplitude, as described by Berkson (2012). Figure 2.2 shows a more ambiguous release interval – in this token,

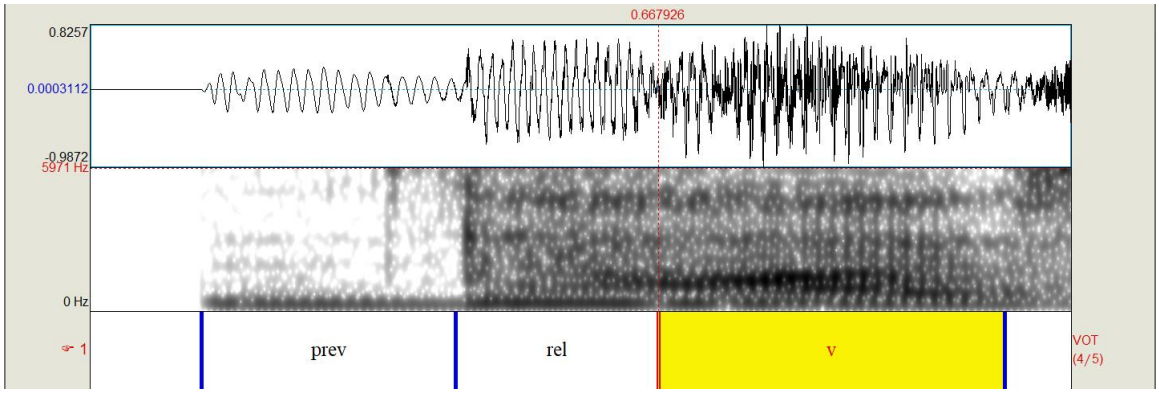
the formant structure becomes somewhat clearer after the indicated release, and the waveform shows a clear change, but this transition occurs at a *decrease* in amplitude. Further complicating Berkson's (2012) intensity peak guideline is a token shown in Figure 2.3, which has several peaks in intensity. However, the transition in formant structure and increased complexity in the waveform make the boundary between release and vowel somewhat clearer here.

Figure 2.1. Release offset marked by amplitude and formants



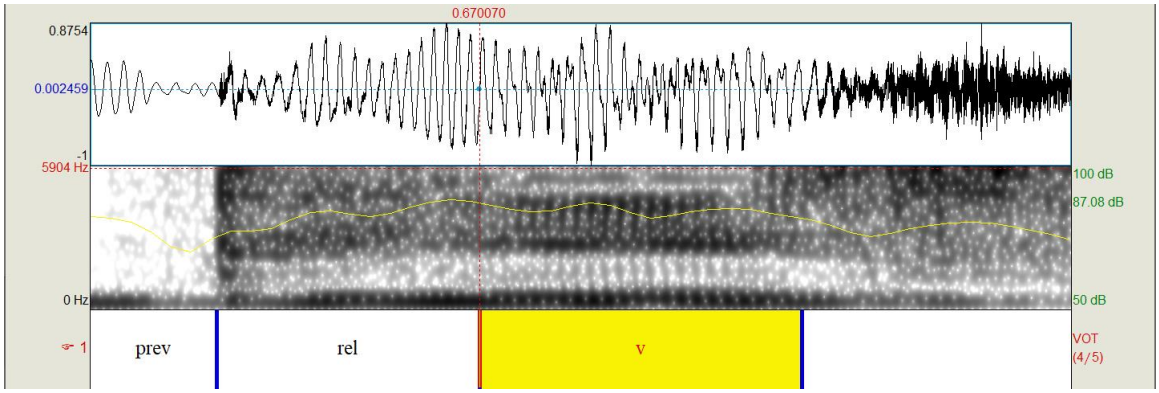
Shown: a token of /g^{h}i:si:/ from Talker 3. "prev" indicates the prevoicing, "rel" indicates the release, and "v" indicates the vowel.

Figure 2.2. A more ambiguous release



Shown: a token of /d^{h}a:sa:/ from Talker 5. "prev" indicates the prevoicing, "rel" indicates the release, and "v" indicates the vowel.

Figure 2.3. A release/vowel with multiple intensity peaks



Shown: a token of /d^{h}i:si:/ from Talker 5. “prev” indicates the prevoicing, “rel” indicates the release, and “v” indicates the vowel. Intensity is plotted in yellow (see green text on left for dB).

Figure 2.4 and Figure 2.5 show the results of this annotation process. Due to the relatively low number of tokens per Marathi category, and because a detailed description of Marathi phonetics is beyond the scope of this thesis, I have only analyzed these tokens’ acoustics using descriptive statistical measures. Release durations are generally similar to those reported by Dmitrieva and Dutta (2019) (compare Table 2.6 with Table 2.7). However, prevoicing for the voiced tokens was much longer than in Dmitrieva and Dutta’s (2019) study, though these data corroborate the trend they reported of plain voiced stops having longer lead voicing than voiced aspirates. I suspect the difference in absolute timing is due to differences in the recording procedure; Dmitrieva and Dutta’s (2019) participants read individual words at their own pace without supervision, while mine read somewhat slowly and carefully from a sentence, under my supervision. Additionally, Dmitrieva and Dutta (2019) used real Marathi words, while my tokens were nonce words. Thus, participants were likely more careful in articulating the tokens for this thesis, leading to longer prevoicing.

Figure 2.4. Prevoicing in stimulus tokens

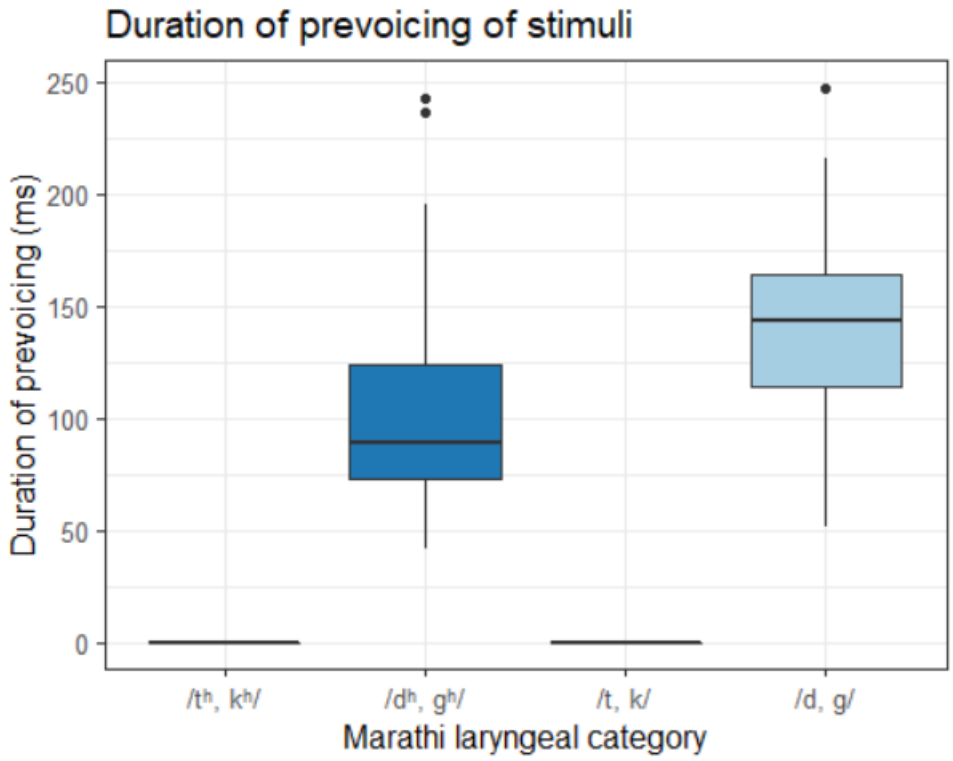
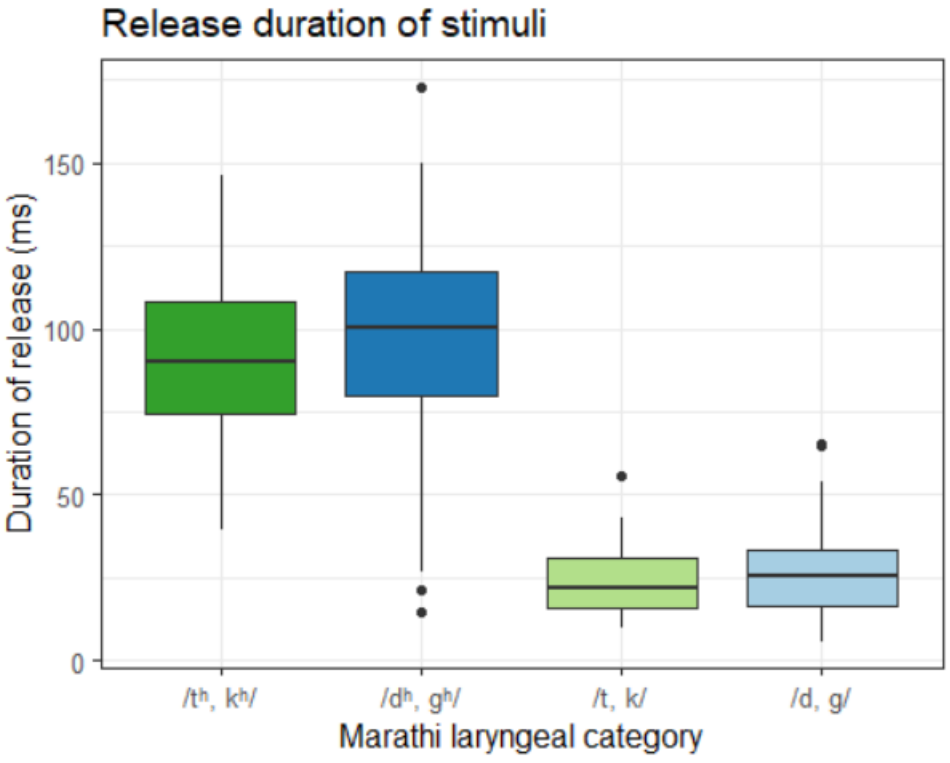


Figure 2.5. Release of stop closure in stimulus tokens



Acoustic measures were derived from the annotations described above to compare the acoustics of these tokens with those of similar tokens used in prior work. Specifically, VOT, onset F0, the percentage of voicelessness during release, and an adaptation of the noise-to-harmonics ratio of the release were measured. Table 2.5 summarizes how each acoustic measure was extracted from the stimuli.

Table 2.5. Implementation of acoustic measures

Measurement	Implementation
Voice onset-time (VOT)	Time difference between onset of stop release and onset of voicing
Onset F0	Earliest automatically detectable pitch after stop release
Percentage of voiceless release	Ratio of voiceless to voiced portion of release based on manual annotation (not Praat voice report)
Noise-to-harmonics ratio (NHR) of aspiration	Praat's mean NHR measurement over aspiration and breathy portions of the vowel, adjusted so (automatically found) voiceless frames had an NHR of 1.0

VOT was calculated as the negative value of the duration of prevoicing if present, or else the duration between the onset of release and the onset of either a voiced portion of the release or the vowel. Three stimuli were not prevoiced but had a fully voiced release, so their VOT was set to zero.

Onset F0 was measured, similarly to previous work (Dmitrieva & Dutta, 2019; Dmitrieva et al., 2015), as the earliest pitch in Hertz that Praat's (Boersma & Weenink, 2020) pitch detection algorithm could detect after the onset of the release. These values were manually checked for pitch halving or octave doubling errors. In dubious cases, the automatically generated value would have been compared against the reciprocal of the duration of the first regular period in the waveform, but no errors were found. The Hertz value of the pitch was then converted to the difference in semitones between the onset F0 value and the talker's mean pitch – in other words, a positive value if the onset F0 was higher than the talker's mean pitch and a negative value if lower.

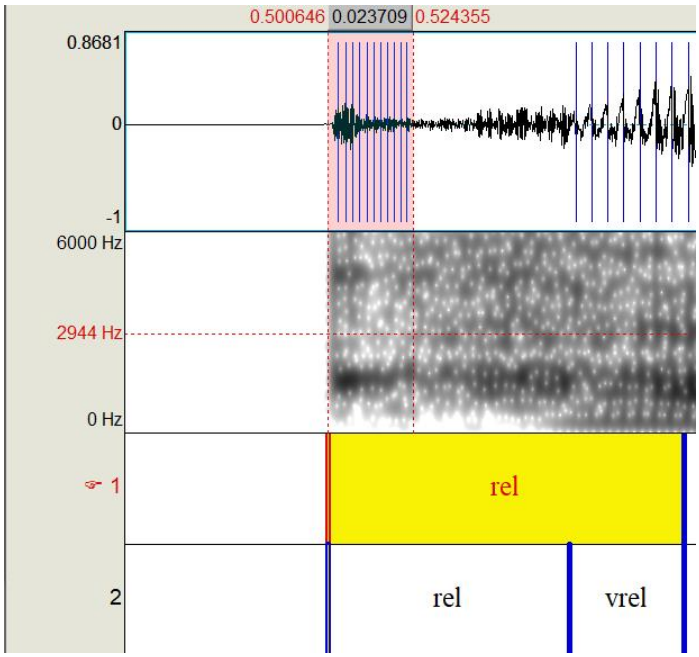
The percentage of voicelessness during release was calculated as a ratio of the duration of voiceless release to the total release duration, according to the annotations discussed in Section 2.2. This was done because Praat's (Boersma & Weenink, 2020) detection of glottal pulses (i.e., voiced frames) produced several false positives (see Figure 2.6) and false negatives (see Figure 2.7), and also varied according to the zoom level of the program (see Figure 2.8). Of perhaps greater concern, the percentage of voiceless frames given by Praat's voice report function does not have a clear

relationship with its detection of glottal pulses. For example, zooming in only on the release region of the waveform shown in Figure 2.8 yields 12 / 36 = 33.3% voiceless frames, while querying the same region zoomed fully out yields 18 / 35 = 51.2%. Curiously, another token of the same type shows no glottal pulses during its release, but its voice report indicates only 26 / 42 = 62% voiceless frames. However, the noise-to-harmonics function of the voice report appears to be more stable; a noise-to-harmonics ratio of 0.0 would indicate a perfectly harmonic signal with no noise, while a value greater than 1 indicates more energy in noise than harmonics (Boersma & Weenink, 2020). For the token shown in Figure 2.8, when zoomed in, the noise-to-harmonics ratio is 0.87 (reflecting the high noise during the misidentified glottal pulses) while zoomed out it is 0.23. Likewise, for the token where no glottal pulses are detected but the amount of unvoiced frames is not equal to zero, the noise-to-harmonics ratio is undefined (Praat does not calculate this ratio when it does not detect voicing).

Therefore, as an alternative to manual annotation, an adjusted measure of noise-to-harmonics ratio (NHR) was also used: Praat's (Boersma & Weenink, 2020) noise-to-harmonics value (generated using the voice report function) was recorded for voiced frames, then combined in a time-weighted average with the proportion of voiceless frames as if voiceless frames had an NHR of one. Releases without any defined NHR were given a value of 1. This was intended to correct some of the instability of the "percentage of locally unvoiced frames" measurement, while still using a gradient measure of the quality of a release. This measure was calculated according to equation (1), in which f represents the percentage of locally unvoiced frames and h indicates the NHR given by the voice report. $1 - f$ is taken to be the percentage of voiced frames.

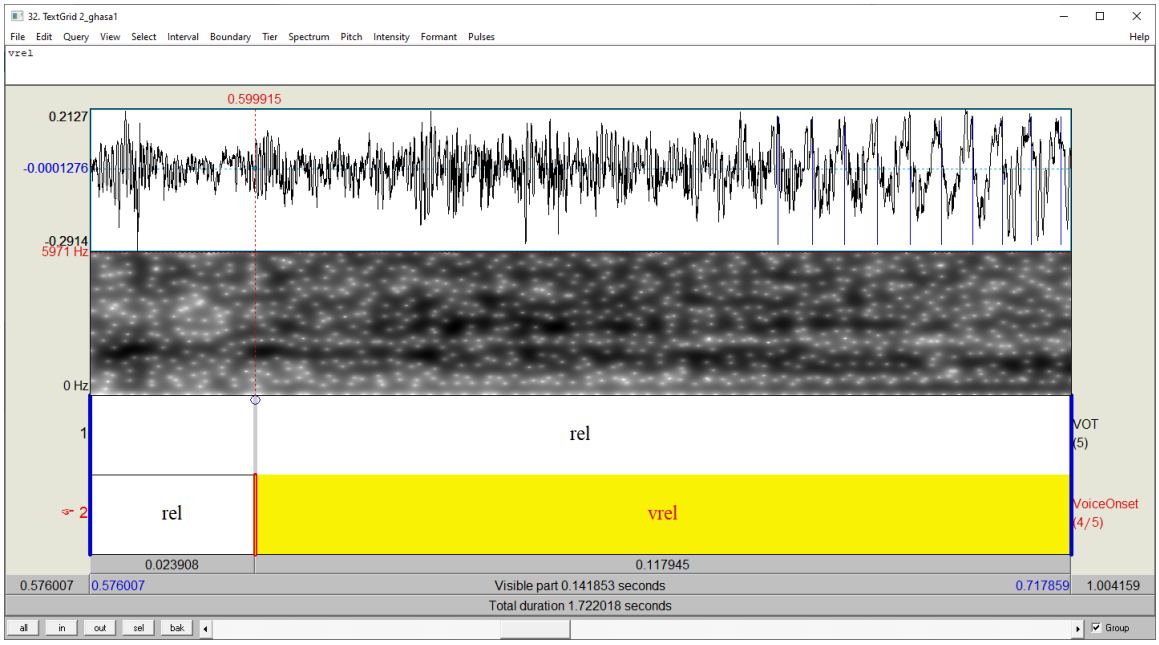
(1) $f + (1 - f) * h$

Figure 2.6. False positive glottal pulses



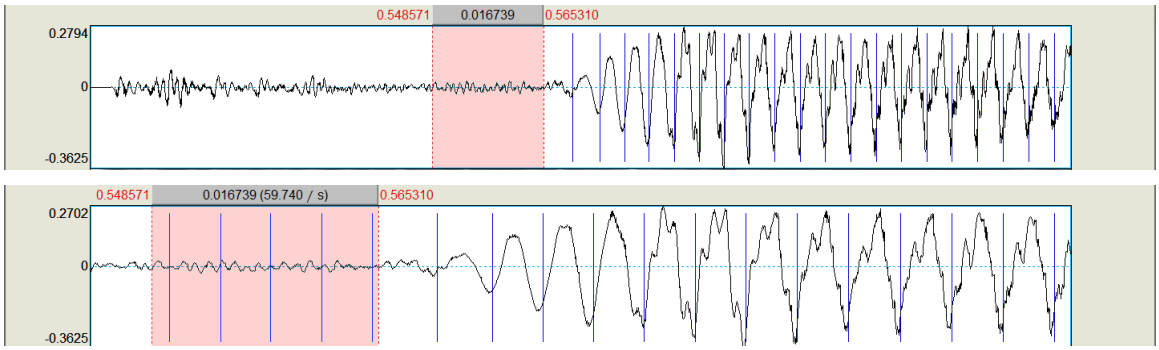
Shown: One of Talker 2's tokens of /k^{h}a:sa:/. Glottal pulses detected by Praat are shown as vertical blue lines in the waveform. The false positive region is highlighted in red. The label "vrel" indicates the actual voiced portion of the release.

Figure 2.7. False negative glottal pulses



Shown: One of talker 2's tokens of /g^{h}a:sa:/. Glottal pulses detected by Praat are displayed as blue vertical lines in the waveform. A red vertical dashed line indicates the actual onset of voicing. The label "vrel" indicates the actual voiced portion of the release.

Figure 2.8. Glottal pulses depend on zoom



The waveform of Talker 2's tokens of /k^{h}u:su:/. Above: a more zoomed-out view of part of the waveform. Below: a more zoomed-in view of the waveform. In both, glottal pulses detected by Praat are displayed as blue vertical lines in the waveform. The same region of the same waveform is highlighted in red in both images. Glottal pulses appear when zoomed in, but not when zoomed out. The highlighted region is in fact unvoiced.

2.3. Acoustic analysis

Table 2.6 summarizes the crucial acoustic measures extracted from the stimuli. Release durations in my data fall within similar ranges to Dmitrieva & Dutta's (2019) observations, and are slightly lower than Berkson's (2012) PVI values. The onset F0 of Marathi tokens follows the trend noted by Dmitrieva & Dutta (2019) that Marathi voiceless plosives have a higher onset F0 than voiced ones. Semitone values and percentage of voicing during release are also comparable to those reported by Dmitrieva & Dutta (2019), though my data show a somewhat higher proportion of voicing during voiced unaspirated stops and lower proportion of voicing during voiceless stops. These differences may be due to previously discussed errors caused by Praat's voice report measurements.

Table 2.6. Acoustic measures of stimuli

Laryngeal category	Phone	Voicing lead (ms)	Release duration (ms)	Onset F0 (st from mean)	Percent voiceless rel.	NHR of release
Voiceless aspirated	/k^{h}/	0.0 (0.0)	105.9 (22.9)	1.40 (1.80)	0.911 (0.098)	0.872 (0.143)
	/t^{h}/	0.0 (0.0)	78.6 (20.3)	1.37 (1.70)	0.857 (0.138)	0.887 (0.106)
Voiceless unaspirated	/k/	0.0 (0.0)	31.1 (8.0)	2.08 (2.28)	0.809 (0.219)	0.836 (0.202)
	/t/	0.0 (0.0)	16.7 (4.9)	3.10 (2.34)	0.645 (0.203)	0.825 (0.164)
Voiced unaspirated	/g/	130.5 (39.0)	33.5 (11.3)	-1.43 (1.28)	0.092 (0.282)	0.398 (0.373)
	/d/	144.6 (34.7)	18.6 (9.6)	-0.80 (1.21)	0.000 (0.000)	0.381 (0.300)
Voiced aspirated	/g^{h}/	88.9 (31.7)	104.8 (27.2)	-2.75 (1.32)	0.142 (0.285)	0.422 (0.296)
	/d^{h}/	114.9 (50.9)	84.4 (38.3)	-1.67 (1.68)	0.135 (0.238)	0.413 (0.281)

Mean values of each measure are shown, with standard deviations in parentheses.

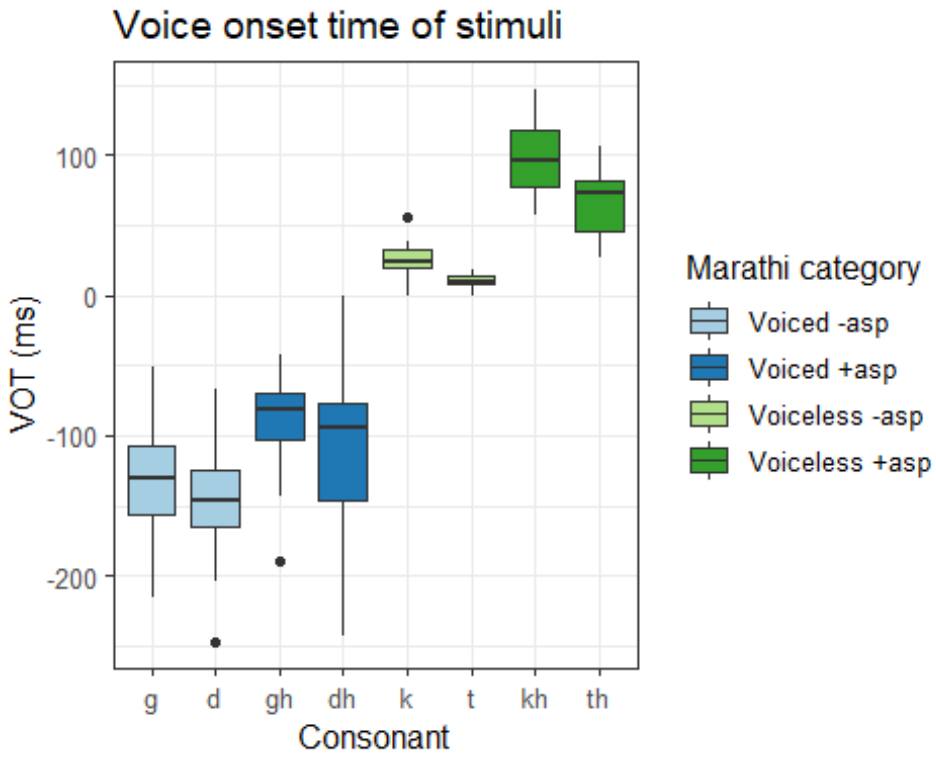
Table 2.7. Dmitrieva and Dutta's (2019, p. 13) acoustic measures, adapted from their Table 4

Laryngeal category	Voicing lead	Release duration (ms)	Percent voiceless rel.
Voiceless aspirated	0.0 (0.0)	76.8 (22.5)	0.793 (22.7)
Voiceless unaspirated	0.0 (0.0)	26.8 (10.7)	0.726 (29.4)
Voiced unaspirated	-87.9 (28.6)	29.8 (12.7)	0.112 (23.2)
Voiced aspirated	-69.5 (27.3)	92.4 (37.4)	0.913 (14.9

Mean values of each measure are shown, with standard deviations in parentheses.

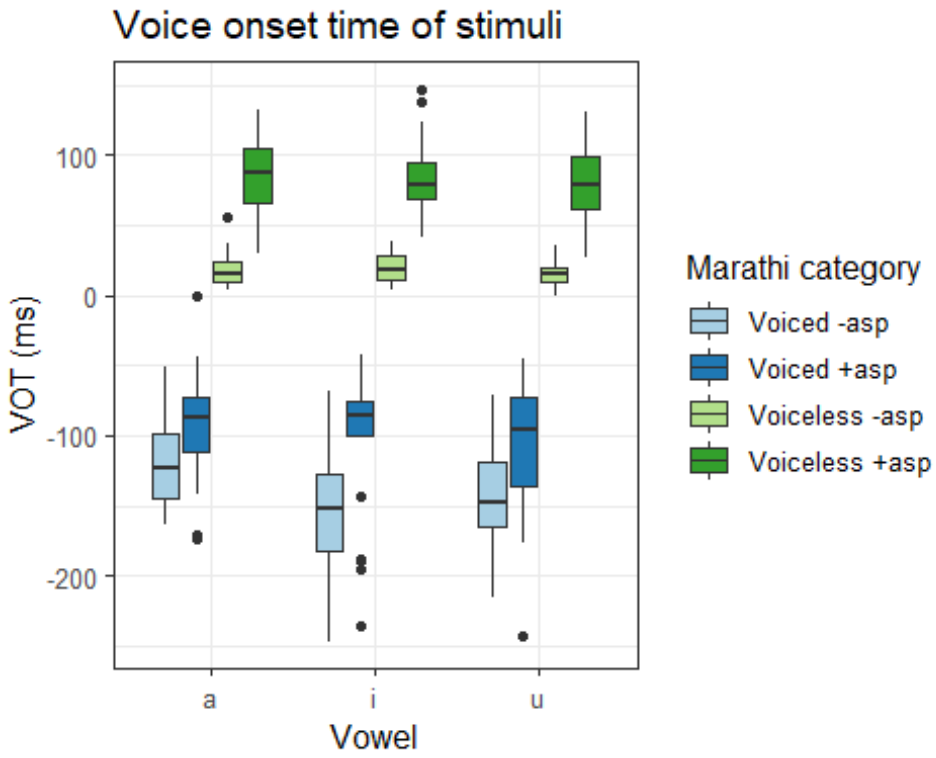
VOT is known to be correlated with certain contextual factors that are not strictly part of a laryngeal contrast, although this relationship has not been specifically confirmed in Marathi. In English, VOT varies slightly depending on the place of articulation of a stop; it is shorter for places of articulation towards the front of the oral cavity (e.g. bilabial stop /p/) than towards the back (e.g. velar stop /k/) (T. Cho & Ladefoged, 1999; T. Cho et al., 2019; Klatt, 1973). An English plosive's VOT also varies somewhat depending on the following vowel; before /i/ and /u/ it is longer than before /a/, but it is unclear whether and how other vowels are affected (T. Cho & Ladefoged, 1999; Klatt, 1973; Rotunno, 1979; Weismer, 1979). Both vowel and place of articulation effects on VOT have also been found in French (Nearey & Rochet, 1994). Figure 2.9 shows the relationship between VOT and the place of articulation of my stimuli for each of the four Marathi laryngeal stop categories. While the difference is more prominent for voiceless categories, the mean VOT of the velar plosive in each category is longer than the corresponding dental mean VOT. Figure 2.10 illustrates the relationship between a plosive's VOT and the following vowel. There does not seem to be a clear relationship between these factors in this dataset, though a more powerful study would be required to fully investigate the issue.

Figure 2.9. Effects of place of articulation on VOT.



Note: -asp is an abbreviation for “unaspirated,” and +asp for “aspirated.” These should not be taken as feature specifications.

Figure 2.10. Effect of following vowel on VOT.

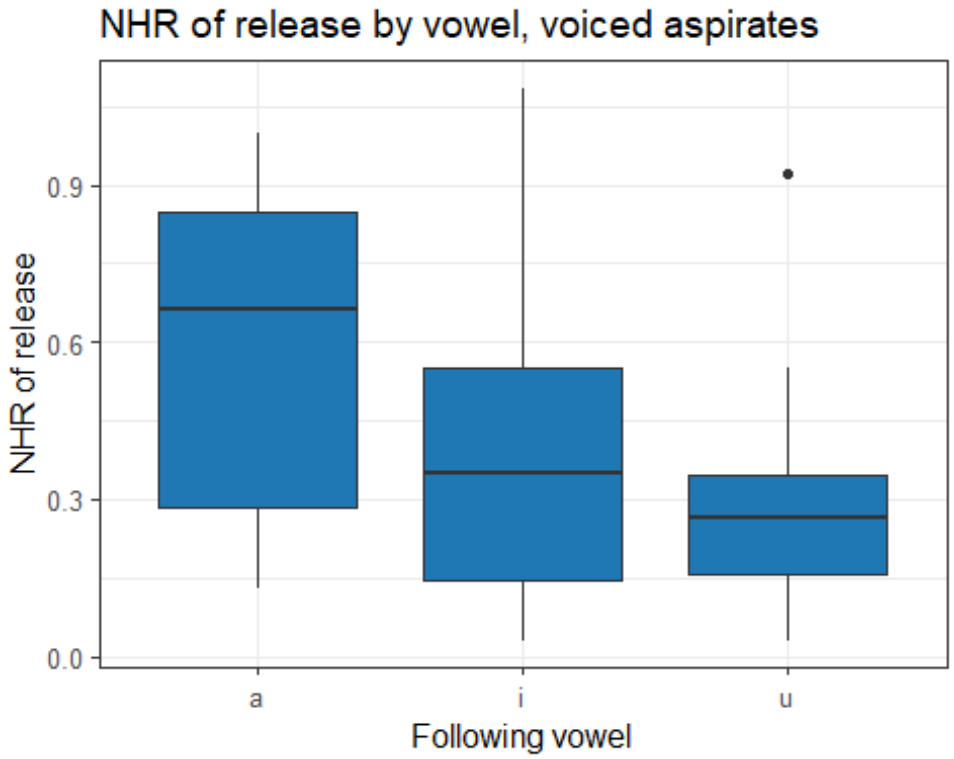


Note: -asp is an abbreviation for “unaspirated,” and +asp for “aspirated.” These should not be taken as feature specifications.

Whereas English aspiration involves basically turbulent airflow prior to the onset of the vowel (Iverson & Salmons, 1995), meaning that English fortis releases would be marked by a predominantly voiceless, noisy release (i.e., high percentage of voicelessness and high NHR), Marathi voiced aspirates score much lower on these measures than voiceless aspirates (see Table 2.6). However, the duration of the release of Marathi voiced and voiceless aspirates is comparable (see Figure 2.5) and falls within the range of English release durations (Lisker & Abramson, 1964). In other words, while the duration of these release intervals is similar, the quality of the aspiration of Marathi voiced aspirates is different. Therefore, Chapter 3 will include an analysis of measures of the quality of stop releases (i.e., NHR and percentage of voiceless release) as well as their durations. Interestingly, these measures seem to vary with the following vowel, but it is unclear whether this is due to some aspect of the recording procedure or part of some regular phonetic-phonological process in Marathi. Figure 2.11 shows the relationship between NHR of a stop release and the following vowel – the relationship between the percentage of voiceless release and the following vowel is similar, namely

that stop releases are somewhat noisier or more voiceless when the following vowel is /aː/. This means that any variation based on the vowel that English listeners perceive may actually be due to phonetic variation in Marathi rather than the perceptual system of the English listener.

Figure 2.11. Relationship between NHR and following vowel for voiced aspirates.



2.4. Summary

In this chapter, I defined the acoustic measures that I will analyze: release duration, VOT, onset F0, adjusted NHR of release, and percentage of voiceless release. Dmitrieva and Dutta (2019) showed that all of these acoustic cues correlate with the Marathi laryngeal contrast, and my data show similar acoustic patterns to theirs. Additionally, the first three of these measures (release duration, VOT, and onset F0) are all used in the English laryngeal contrast (although VOT and release duration are redundant in English). The latter two (adjusted NHR of release and percentage of voiceless release) have not been used to describe English phonetics, but may affect English listeners' perception of Marathi stops.

Chapter 3.

English perception of Marathi voiced aspirates

In this chapter, I describe a rating study which investigates English listeners' perception of voicing and aspiration cues, with special interest in these listeners' perception of Marathi voiced aspirates and voiceless inaspirates. English listeners' perception of these categories will indicate whether they pay greater attention to voicing or aspiration, which in turn has implications for their phonological representations of the English laryngeal contrast. I used linear mixed-effects models to analyze the influence of acoustic cues associated with the fortis-lenis contrast to determine which ones English listeners rely on. I conclude by briefly discussing what these phonetic results say about English listeners' phonology.

3.1. Research question and hypotheses

The central research question of this study is how English listeners will use the acoustic cues associated with the features [voice] and [spread glottis] in perceiving Marathi stops. Given the consensus that English is an aspirating language, English speakers are expected to rely more heavily on aspiration (or [sg]) than prevoicing (or [voice]), meaning that voiced aspirates should be judged roughly similarly to fortis stops (which also have aspiration) and voiceless inaspirates should be judged as lenis stops (which also lack aspiration). Since English lenis stops are sometimes realized with prevoicing, voiced aspirates may be perceived as more ambiguous between lenis and fortis, though voiceless inaspirates should still be perceived as lenis. If English listeners consider both categories to be ambiguous, that would indicate the use of both [voice] and [spread glottis], with neither weighed more heavily than the other.

Regardless of the exact phonological features at work, the Marathi categories of plain voiced and voiceless aspirated stops are expected to easily assimilate to English lenis and fortis categories, respectively, as their phonological features and acoustic properties resemble those of English laryngeal categories.

3.2. Methods

3.2.1. Participants

I used Figure Eight, a crowdsourcing platform, to recruit and run native English listeners. Figure Eight allows “customers” (e.g. the author) to create annotation or data collection “jobs” (e.g. different conditions of the study), which Figure Eight offers to its “contributors” (for this study, the participants) and facilitates payment from customer to contributor. For the present study, we ran two conditions: one for velar stops and one for dental stops.

Twenty-five people contributed to each job. Participants were excluded if they indicated a native language other than English (n = 6) or if they did not listen to every stimulus token (n = 12). Four participants both were not native English speakers and did not complete the job. After exclusions, there were 15 participants in the coronal condition and 14 in the velar condition. Because both conditions were posted at the same time, many people contributed to both conditions. There were 16 unique participants in this study: One participated in only the velar condition, two contributed only to the coronal condition, and the remaining 13 participated in both place conditions.

Figure Eight does not collect detailed background information on their participants, and technical limitations of the tool make a comprehensive background questionnaire burdensome (see Section 4.3 for further discussion), so data on participants’ linguistic background is limited. Participants were only included if they indicated their native language was English, and were excluded if they spoke any language with a four-way laryngeal contrast (e.g. Hindi, Marathi, Bengali) or voiced aspirates (e.g. Punjabi), though no participants reported speaking such a language. While participants were restricted to people in Canada automatically by Figure Eight, they were not asked for data such as their place of birth, arrival in Canada, or similar information. As well, there were no further restrictions on location than “Canada,” so participants may have been located anywhere in the country with an internet connection. So while I described these participants as speakers of Canadian English in Chapter 1, this should be taken in a very broad sense.

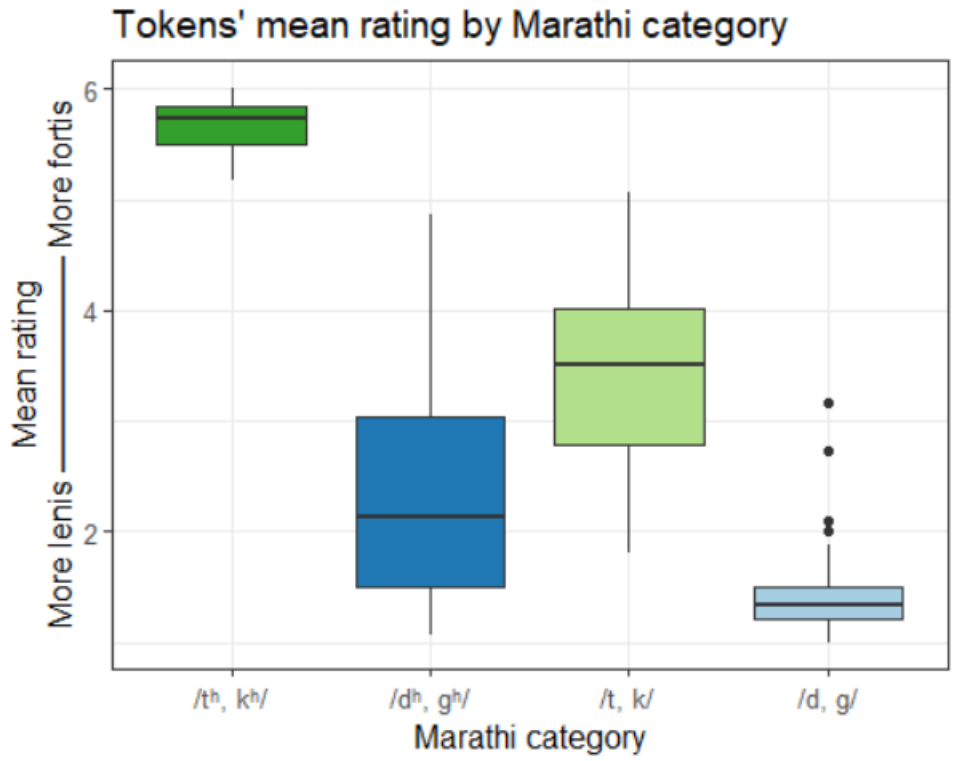
3.2.2. Procedure

Participants listened to each of the 288 stimuli (described in Chapter 2) in a random order. They were allowed to replay each token as many times as they liked. After hearing the token, they judged whether the initial consonant sounded more like a voiced consonant (i.e., “d” or “g”) or a voiceless consonant (i.e., “t” or “k”) using a six-point Likert scale. A rating of one indicated "a normal d" (or "g" for the velar condition) and six indicated "a normal t" (or "k").

3.3. Results

Figure 3.1[8] shows how the mean ratings of each unique stimulus token are distributed within each Marathi category.

Figure 3.1. Summary of rating results



[8] Figures and statistics were generated using R (R Core Team, 2020) via RStudio (RStudio Team, 2019); figures were created with the ggplot2 (Wickham, 2016) package.

As expected, English speakers overwhelmingly heard the voiceless aspirated and voiced unaspirated stops as fortis and lenis stops, respectively, while there was high variability in the ratings of voiced aspirates and plain voiceless plosives.

Voiced aspirated tokens tended to receive lenis mean ratings, and generally received more stable ratings (SD = 1.75) than plain voiceless tokens (SD = 2.00). Therefore, the effect of prevoicing seems to have generally outweighed that of aspiration for these tokens. This suggests, surprisingly, that English gives preferential weight to voicing over aspiration. In contrast, plain voiceless tokens received highly variable ratings that did not lean strongly towards voiced or voiceless. Indeed, the mean rating of all plain voiceless tokens was 3.47 (and a perfectly intermediate average rating would be 3.5). Voiceless inaspirates seem to have been nearly perfectly ambiguous, indicating that listeners preferred to rate tokens as strongly lenis only when specified for [voice], and weighed a lack of prevoicing as no less important than a lack of aspiration. So, an examination of the distribution of ratings by Marathi laryngeal category yields mixed results. The ratings of voiced aspirates suggest a preference for prevoicing over aspiration, but the ratings of voiceless aspirates suggest no preference. But before continuing to discuss this discrepancy, I will show that different Marathi categories actually received statistically significantly distinct ratings from one another.

To confirm that the patterns of English listeners' ratings vary according to the Marathi laryngeal category they heard, I ran a linear mixed effects model using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) for R (R Core Team, 2020). These models are robust to imbalanced designs (recall that some, but not all, participants rated sounds in both places of articulation), and thus ideally suited to this data set (Baayen, Davidson, & Bates, 2008). The dependent variable was the average rating for a given token, and independent variables included the Marathi laryngeal category (with four treatment coded levels: voiced aspirate [reference level], voiced inaspirate, voiceless inaspirate, and voiceless aspirate), following vowel (with three treatment coded levels: /a [reference level], i, u/), and place of articulation (velar or dental). The maximal random effects structure that would produce a converging model was used for this and all subsequent linear mixed-effects models, following Barr, Levy, Scheepers, and Tily (2013). However, due to the small dataset, in this case that consisted only of a random intercept for subject and no random slopes. Significance testing was then done using a chi-squared likelihood ratio test via the drop1() function (Bates et al., 2015). Both the

Marathi place of articulation and the following vowel were found to be significant predictors of participants' ratings. However, since there were differences in NHR and release duration between vowels, it is unclear whether the significance of the "vowel" factor is due to qualities of the vowel itself or the acoustic covariates found in these stimuli.

Table 3.1. Summary of LME results

Variable	Degrees of freedom	X^2	*p* value
Marathi laryngeal category	3	3955.9	< 0.0001
Following vowel	2	212.88	< 0.0001
Place of articulation	1	0.11	> 0.70

Non-significant results ($p < 0.05$) are shaded in gray.

The results of the model indicate that the Marathi laryngeal category of a sound has a statistically significant effect on how it is perceived by an English listener, but the model alone does not indicate whether each category was rated differently than each other category. In order to test this, Tukey's range test was run on the estimated marginal means of this model, implemented by the emmeans (Lenth, Singmann, Love, Buerkner, & Herve, 2020) package's emmeans() and pair() functions. The results of Tukey's range test are shown in Table 3.2: each Marathi category did receive significantly different ratings than each other category.

Table 3.2. Test of estimated marginal means

Contrast	Estimate	Standard err.	*t* ratio	*p* value
Voiced inaspirate – voiced aspirate	-1.03	0.066	-15.7	< 0.0001
Voiced inaspirate – voiceless aspirate	-2.07	0.066	-31.6	< 0.0001
Voiced inaspirate – voiceless aspirate	-4.32	0.072	-60.0	< 0.0001
Voiced aspirate – voiceless inaspirate	-1.05	0.062	-16.8	< 0.0001
Voiced aspirate – voiceless aspirate	-3.29	0.069	-47.8	< 0.0001
Voiceless inaspirate – voiceless aspirate	-2.25	0.069	-32.6	< 0.0001

Degrees of freedom = 3,644 for all pairs, calculated using the Kenward-Roger approximation.

As noted earlier, the ratings of voiced aspirates and voiced inaspirates seem to give rise to somewhat contradictory interpretations: the more-lenis-than-fortis rating of voiced aspirates suggests that prevoicing is weighted more heavily than aspiration, while the ambiguous rating of voiceless inaspirates suggests that both are considered roughly equally. To attempt to resolve this discrepancy, I used further linear mixed-effects models to investigate how English listeners used acoustic cues to laryngeal category membership in their perception of all Marathi stimuli.

3.3.1. Acoustical analysis of the perception of Marathi laryngeal categories

In this subsection, I investigate which acoustic cues English listeners attended overall, across all laryngeal categories. To do so, I generated another linear mixed-effects model (Bates et al., 2015; R Core Team, 2020). The dependent variable was again the rating given, and fixed effects were VOT, onset F0, and NHR of the token's aspirated/breathy-voiced portion, all of which were normalized to *z*-scores before being included in the model. The final model had random intercepts including the vowel following the consonant (a/i/u), the consonant place of articulation (velar/coronal), and the subject (and no random slopes). Due to convergence errors preventing the calculation of likelihood ratio tests, *p*-values were obtained instead using Type II Wald chi-square tests with the Anova() function from the cars package (Fox & Weisberg, 2019). Table 3.3 shows these results.

Table 3.3. Linear mixed-effects model of influence of acoustics on English perception of all Marathi categories

Factor	Estimate	Standard error	X^2	Approximate *p*
VOT	1.26	0.04	1049.9	< 0.0001
Onset F0	-0.24	0.03	51.9	< 0.0001
NHR	0.46	0.03	174.3	< 0.0001

The model shows statistically significant results for VOT, onset F0, and NHR. The relationship between each of these three acoustic cues and participants' ratings is plotted individually in the Appendix. Estimated coefficients in the model are indicative of the size of fixed effects, and based on these estimates, VOT was the strongest predictor of a token's rating, followed by NHR and then onset F0. The pre-eminence of VOT is not surprising, as it is typically considered the best measure of laryngeal category membership in English. However, as VOT implicates both [voice] and [spread glottis] in these data, it is difficult to make phonological claims using VOT in this model.

In this model, onset F0 shows a negative estimate, when it is expected to be positive (i.e., higher onset F0 correlates with perceiving the sound as more fortis). This is because the model is capturing the amount of variance explained by onset F0 while holding VOT and NHR constant. In fact, Marathi voiceless inaspirates had higher onset F0s than voiceless aspirates, but voiceless inaspirates were rated more ambiguously;

similarly, voiced aspirates had lower onset F0 than voiced inaspirates, but were also rated more ambiguously. So, controlling for VOT and NHR, participants tended to perceive lower-onset-F0 tokens as more voiced (contrary to expectation). This is most likely because the other acoustic cues overrode onset F0 in terms of their perception – in other words, when forced to choose between, for example, VOT and onset F0, participants rated according to VOT. This is in keeping with previous findings that VOT has a greater effect[9] on perception than onset F0 (Abramson & Lisker, 1985; Francis et al., 2008).

The NHR of a token's release showed a significant effect in this model, and one of greater magnitude (i.e., higher absolute value of the estimate) than onset F0. This indicates that aside from the VOT, the *quality* of a release also has an effect on listeners' perception of the corresponding plosive's laryngeal category. A more voiceless and turbulent release, which is more similar to English aspiration, correlated with a more fortis perception of the stop (hence the positive estimate). Few studies have previously investigated non-durational measures of aspiration (to my knowledge, only Repp, 1979 has), so it is notable that this variable has been shown to be significant, and in fact stronger than onset F0, which has been established as a noteworthy and distinct cue to laryngeal contrasts (T. Cho et al., 2019; Dmitrieva & Dutta, 2019; Dmitrieva et al., 2015, among others).

To sum up, English listeners were using the sorts of acoustic cues that they would be expected to, in addition to the quality of a token's aspiration. However, the source of ambiguity in ratings of voiced aspirates and voiceless inaspirates is still unclear, as is the differences between these categories that made voiced aspirates less ambiguous and more assimilable to English lenis stops, which are the motivating questions behind the following two analyses.

3.3.2. Marathi voiced aspirates and voiceless inaspirates

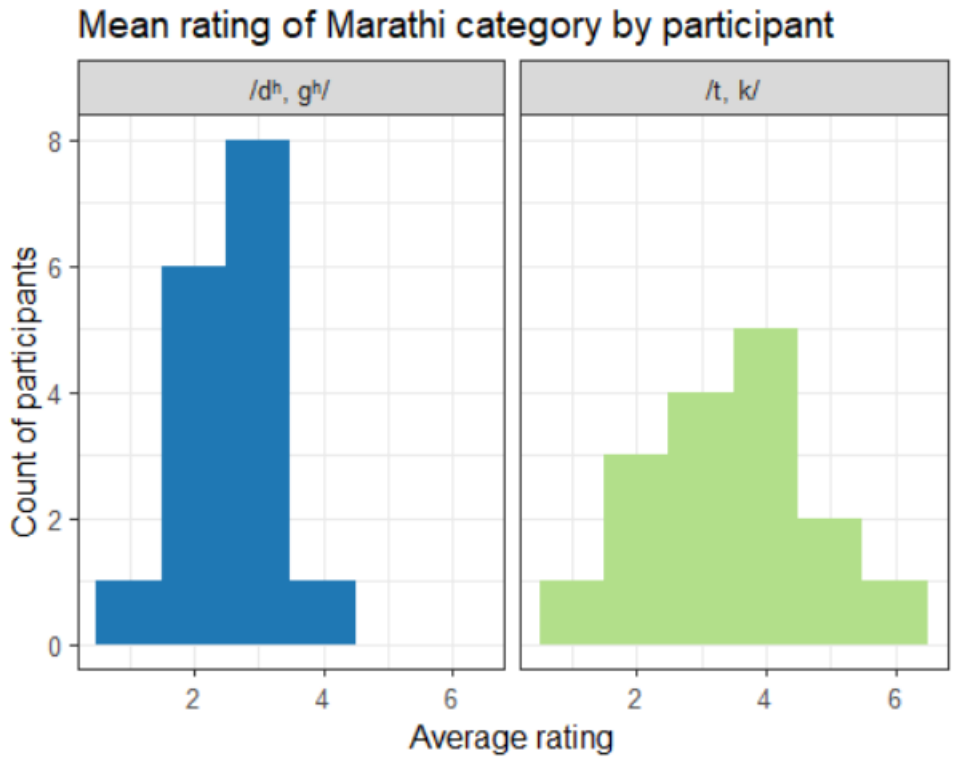
In order to determine why Marathi voiced aspirates were rated differently than voiceless inaspirates, I examined the effects of between-token and between-participant differences on the variability of listeners' ratings. If between-token differences

[9] Francis et al. (2008) would say that this discrepancy is actually due to the *perceptual distance* between differences in VOT and differences in onset F0, rather than a preference for VOT per se.

contributed high variability to ratings of a Marathi category, this would suggest that listeners are making linguistic distinctions between the *different acoustic properties* of tokens within a Marathi category. Conversely, if between-participant differences contributed high variability to ratings for a Marathi category, it indicates that listeners have *different interpretations* of the same phonetic information.

There is some inter-speaker variability in exactly where laryngeal category boundaries are (see, e.g., Shultz, Francis, & Llanos, 2012). Consequently, when tokens are somewhat ambiguous, as in this study, participants may vary considerably in their rating of different laryngeal categories. To assess whether this may have impacted participants' ratings of voiced aspirates and voiceless inaspirates, it is useful to consider the distribution of participants' mean rating of each category. If participants give very different mean ratings to each category, that suggests that participant-based factors are driving variability in the ratings, while if they give a consistent mean rating, then that means participant-based factors have little effect. Figure 3.2 shows the distribution of participants' mean rating of each Marathi category. The ratings for voiced aspirates show a nearly normal (though skewed) distribution (kurtosis = 3.15, skewness = -0.45), but the distribution of voiceless aspirated tokens shows a somewhat flatter distribution (kurtosis = 2.44, skewness = -0.20). This indicates that participants generally agreed on the rating of voiced aspirates, but not on the rating of plain voiceless plosives. So there was a higher degree of across-participant variability in ratings of plain voiceless tokens than voiced aspirates.

Figure 3.2. Participant variability in ratings



Voiced aspirates (left) show more consistent ratings across participants than voiceless inaspirates (right).

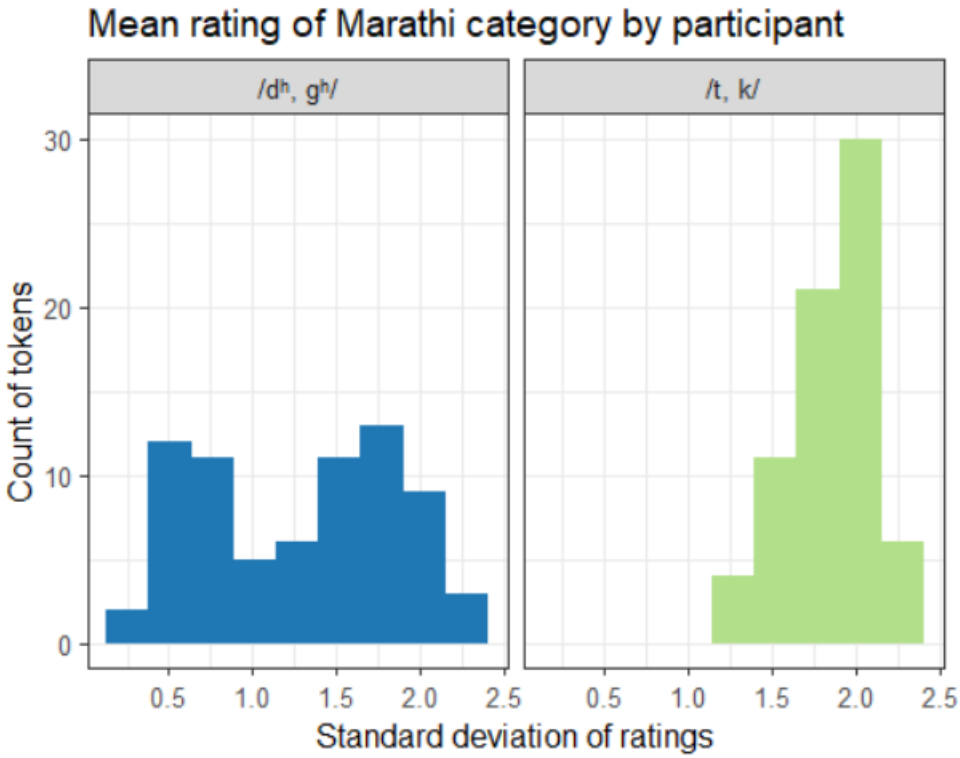
If variability in voiceless inaspirates is more participant-driven than that of voiced aspirates, this does not fully address the question of why voiceless inaspirates were rated more ambiguously than voiced aspirates. Unfortunately, little participant background data was collected for this study, so participant-based factors cannot be thoroughly investigated here, though the role of English participant-based factors on perception of laryngeal categories will be further contextualized in Chapter 4.

However, the relatively token-dependent ratings of voiced aspirates may be explained by the fact that that some voiced aspirates are less ambiguous than others. Less ambiguous tokens would receive less variable answers, which can be measured by the standard deviation of tokens' ratings, shown in Figure 3.3. The bimodal distribution of variability in voiced aspirates (kurtosis = 1.57) shows that there is a cluster of low-variability tokens and a cluster of high-variability tokens. In contrast, variability in plain voiceless tokens is in a relatively normal, unimodal distribution (kurtosis = 2.97), indicating that variation within these tokens did not contribute much variability to listeners' judgments. Furthermore, Figure 3.4 shows that in terms of mean ratings and

the variability of a token's ratings, the high-variability voiced aspirates are very similar to voiceless inaspirates, and Figure 3.5 shows that the low-variability voiced aspirates are likewise very similar to the voiced inaspirates. So English listeners seem to have perceived some voiced aspirates as being similarly ambiguous to voiceless inaspirates, while others were perceived as similar to voiced aspirates.

Taking these sources of variability together, it is evident that within-token variability contributed more to the ratings of voiced aspirated tokens than plain voiceless tokens. After looking more closely at this between-token variability in the voiced aspirates, it also seems that voiced aspirates are being perceived in two different ways. Assuming that token-based variability is rooted in different phonetic properties of voiced aspirate tokens, the acoustic differences between voiced aspirated tokens and their effects on English listeners' judgments warrant closer examination. The following subsection will show that the difference between more- and less-ambiguous voiced aspirates is in fact due to differences in the acoustics of aspiration between these tokens.

Figure 3.3. Token variability in ratings



There seems to be a cluster of low-variability voiced aspirates and a cluster of high-variability ones. Voiceless inaspirate tokens seem to have received comparably variable ratings.

Figure 3.4. Some voiced aspirates are like voiceless inaspirates

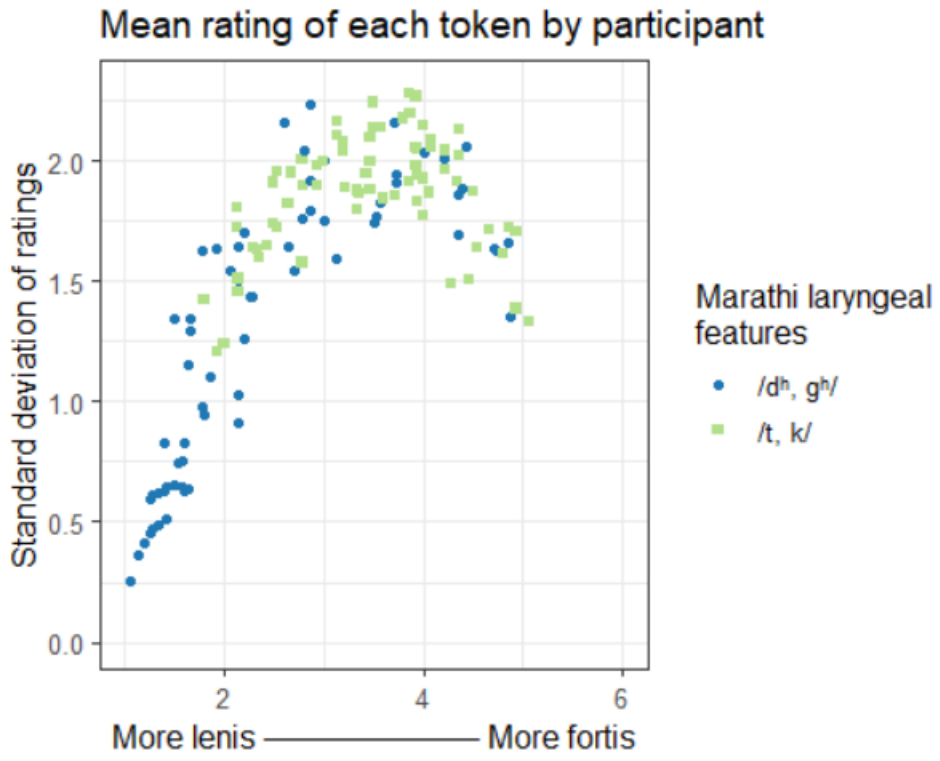
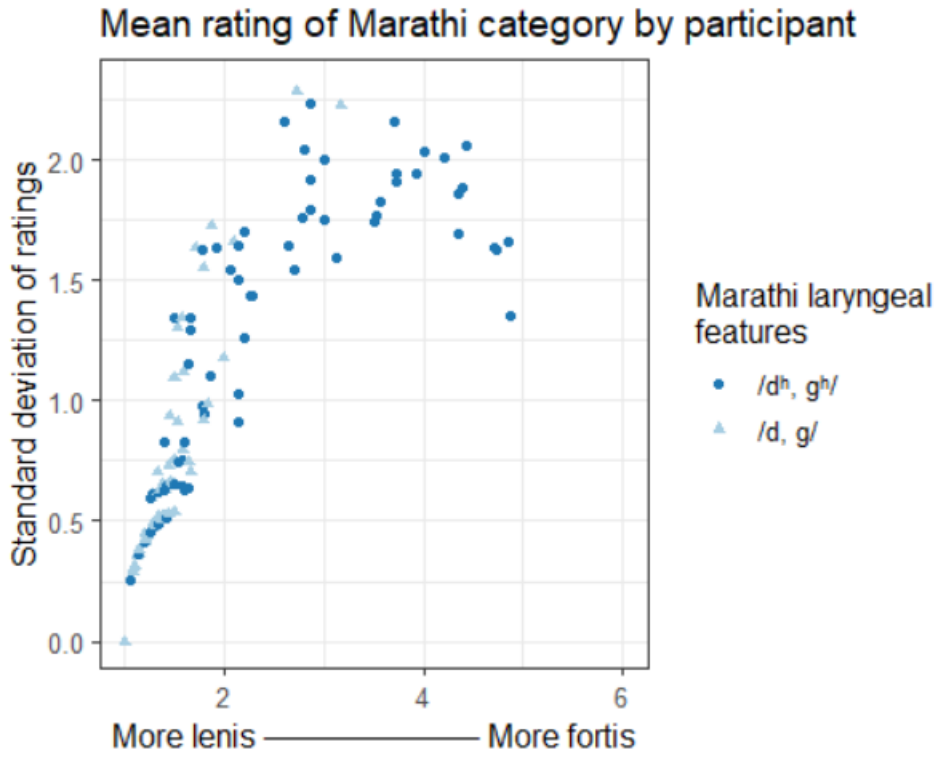


Figure 3.5. Other voiced aspirates are like voiced inaspirates



3.3.3. Acoustical analysis of the perception of voiced aspirates

A linear mixed-effects model was generated (Bates et al., 2015; R Core Team, 2020) to determine which cues English listeners made use of in rating Marathi voiced aspirates. Fixed effects were VOT, onset F0, release duration[10], and NHR of the token's aspirated/breathy-voiced portion. Random intercepts used the same contrast coding as in the previously-described model (see p. 39) and included the vowel following the consonant (a/i/u), the consonant place of articulation (velar/coronal), and the subject (and nothing else). Visual inspection of residual plots did not show any clear violations of assumptions of homoscedasticity or normality. Again, due to convergence errors in calculating likelihood ratio tests, *p*-values were obtained using Type II Wald chi square tests with the Anova() function from the cars package (Fox & Weisberg, 2019). Table 3.4 shows the results of the model.

Table 3.4. Linear mixed-effects model of acoustic cues and voice rating

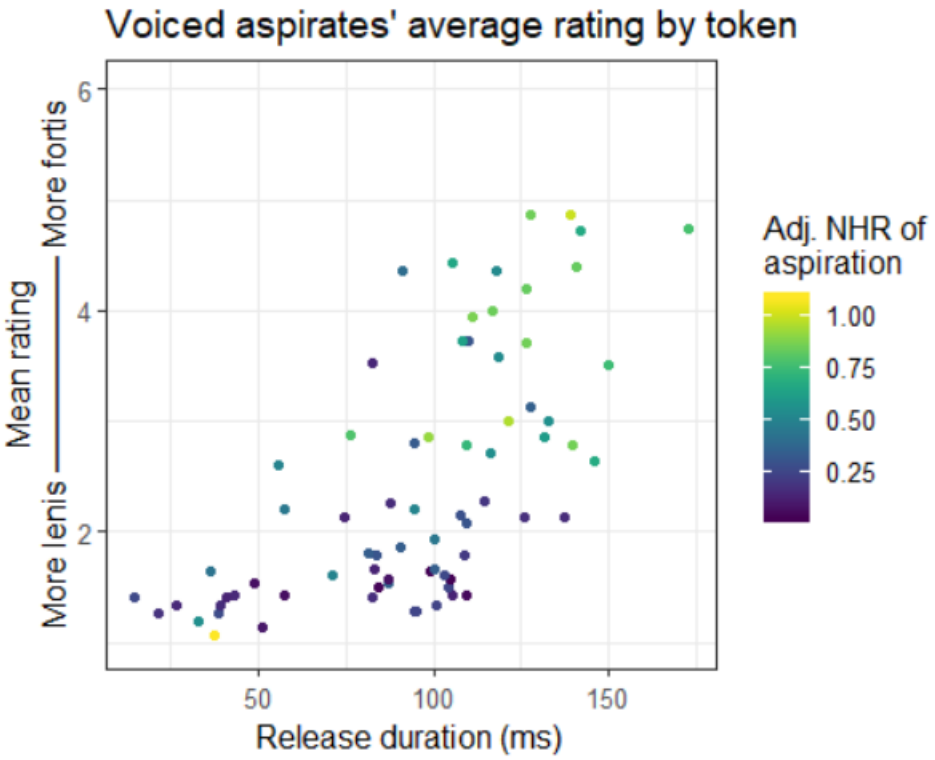
Factor	Estimate	Standard error	X^2	Approximate *p*
VOT	0.19	0.13	2.0	> 0.1
Onset F0	0.02	0.08	0.04	> 0.5
Release duration	0.40	0.09	20.5	< 0.0001
NHR	0.59	0.07	67.7	< 0.0001

Non-statically-significant factors (CI = 95%) are shaded in gray.

This model shows that variability in participant ratings of Marathi voiced aspirates was based on the duration and quality of their release. VOT and onset F0 were not found to be statistically significant predictors in this model. This is not very surprising, as VOT and onset F0 are within the same range as Marathi voiced inaspirates, which were overwhelmingly rated as very lenis-like. Figure 3.6 illustrates the relationship between release-based acoustic measures and a token's rating. Notably, the more ambiguous voiced aspirates have higher NHRs and longer release durations, while the less ambiguous (i.e., more lenis) voiced aspirates have shorter release durations and/or NHRs. Taken with the statistical analysis presented above, this confirms that the difference between the two kinds of voiced aspirates discussed in Section 3.3.2 is indeed based on the acoustics of the voiced aspirates of a given token.

[10] Release duration was not included in the previous model due to its overlap with VOT in most laryngeal categories. But in voiced aspirates, VOT reflects prevoicing and release duration reflects aspiration.

Figure 3.6. Average rating of voiced aspirates is related to both release quality and release duration.



Note that tokens with higher (more fortis) ratings show longer release duration (i.e., points are to the right) and higher NHR (i.e., lighter/yellower color).

3.4. Discussion

Examining the trends of ratings across all Marathi laryngeal categories reveals a surprising result: English listeners seem to base their ratings on both [voice] and [spread glottis]. Voiceless inaspirates were rated very ambiguously, which is unexpected as most prior reports of English perception had indicated that unaspirated stops without prevoicing should be perceived as lenis stops (e.g. Keating, Mikoś, & Ganong, 1981; Nearey & Rochet, 1994; Polka, 1991). Voiced aspirates were also rated ambiguously when their aspiration was English-like, while those tokens with less English-like aspiration were more easily assimilated to the lenis category. In fact, it is likely that English listeners did not perceive all Marathi voiced aspirate releases as phonetic realizations of [spread glottis]. In other words, tokens with English-like aspiration were perceived as being specified for [sg] and [voice], while those with less English-like aspiration were not perceived as specified for [sg], only [voice], hence the similarity in their ratings to voiced inaspirates.

Similar to Hunnicut & Morris's (2016) account of Southern American English, these participants seem to have both an active [voice] and [spread glottis] feature in their phonology. They perceived voiced aspirates, which are specified for both [voice] and [spread glottis] features, quite ambiguously when [sg] was realized as phonetically similar to English, and they perceived voiceless inaspirates, which have neither phonological feature, as ambiguous, suggesting that they were attending both features when judging a plosive's laryngeal category. Furthermore, only prevoiced tokens (voiced aspirates with non-English-like aspiration and voiced inaspirates) achieved average ratings below 1.5 (a rating of 1.0 would indicate that all participants perceived that token as lenis). This indicates that English listeners associate [sg] with fortis and [voice] with lenis stops.

Chapter 4.

General discussion

The results of the study described in Chapter 3 indicate that English listeners paid attention to cues to aspiration and voicing with roughly equal weight: Marathi voiceless inaspirates were perceived as ambiguous, and voiced aspirates were perceived as ambiguous, except for those tokens with less English-like aspiration, which were perceived as voiced. This contradicts the mainstream view that prevoicing is a redundant cue subordinate to aspiration – if that had been the case, Marathi voiceless inaspirates should have been perceived as English lenis stops.

This study has thus found evidence that both [voice] and [sg] features are active in English phonology, which has implications for the broader understanding of English phonetics and phonology as well as implications for English learners of additional languages. I discuss these implications in turn in the following two sections, then address the limitations of the present study and how future work might expand upon it. The final section comprises a summary of this thesis and its key findings.

4.1. English phonetics and phonology

As discussed in Section 1.1, most phonological accounts of the English laryngeal contrast analyze it as depending on a privative feature (Beckman et al., 2013; Y. Y. Cho, 1990; Honeybone, 2012; Iverson & Salmons, 1995; Keating, 1993), and generally identify that feature as [spread glottis] (Beckman et al., 2013; Honeybone, 2012; Iverson & Salmons, 1995). Likewise, phonetic accounts generally describe the English laryngeal contrast as being based on aspiration, distinguishing stops with long VOTs from those with short VOTs, as has been shown both in production (Abramson & Whalen, 2017; T. Cho & Ladefoged, 1999; Lisker & Abramson, 1964; Nearey & Rochet, 1994, among others) and perception (e.g. Benkí, 2005; Francis et al., 2008; Nearey & Rochet, 1994; Polka, 1991; Repp, 1979). The results of the present study contradict these accounts.

Though it was unclear exactly how to predict the ratings for Marathi voiced aspirates, it was expected that their aspiration would make them generally fortis-like.

However, voiced aspirates skewed toward being assimilated as lenis stops, and the highest average rating any such token received was 4.87 (out of 6, with 6 being like a typical fortis stop). Since several of these tokens have aspiration similar to English, prevoicing is likely what is causing the ambiguity in perception. This could support that participants are paying attention to the [voice] feature, or it could be that the prevoicing was only perceived as an acoustic-phonetic (but not phonological) dissimilarity. The ratings of Marathi voiceless inaspirates provide stronger evidence for the former.

English listeners were expected to rate Marathi voiceless inaspirates as lenis stops, but instead they gave highly variable ratings, resulting in a total average rating very near the center of the scale (3.47 out of 6). In contrast, voiced inaspirates were categorized overwhelmingly as typical lenis stops, and the ratings of voiced and voiceless inaspirates were found to be significantly different. This means that English listeners found it easier to rate stops with prevoicing as lenis stops, compared to those without prevoicing. Thus, short-lag VOT was not sufficient for these English listeners to identify these stops as lenis, despite the general consensus in the literature that short-lag VOT is the typical phonetic realization of English lenis stops.

This study contradicts the observation that the English laryngeal contrast is primarily based on short-lag versus long-lag VOT in plosives. Because this observation has been the basis for identifying English as an aspirating language that uses the [spread glottis] feature to make its laryngeal contrast, this study also calls that analysis into question. The difference in ratings between Marathi voiced inaspirates and voiceless inaspirates indicates that [voice] is part of the feature specification for lenis stops, at least for these English listeners because they only consistently rated stops specified for [voice] (namely, voiced inaspirates) as lenis. They rated stops without that feature as fortis (in the case of voiceless aspirates) or ambiguous (in the case of voiceless inaspirates). The difference between voiceless inaspirates and voiceless aspirates indicates that [spread glottis] is also part of the feature specification of English fortis stops, as expected. This means that these participants make a laryngeal contrast similar to that of Southern American English (Hunnicutt & Morris, 2016), Swedish (Helgason & Ringen, 2008), and Norwegian (Ringen & van Dommelen, 2013), involving both the features [voice] and [sg].

One might interpret these results as evidence in favor of accounts that the English laryngeal contrast is based on a binary [± voice] feature, where [– voice] is realized as aspiration (e.g. Keating et al., 1981; Kingston & Diehl, 1994; Wetzels & Mascaró, 2001), or a privative [voice] feature where a stop without laryngeal specification is realized with aspiration (e.g. Y. Y. Cho, 1990; Keating, 1993). In this type of account, the complication introduced by my study is purely phonetic. This type of theory would simply need to be modified to indicate that a [+ voice] or [voice] specification corresponds phonetically with prevoicing, while a [– voice] or null laryngeal specification is phonetically realized as aspiration. However, this study is not sufficient to indicate that such an analysis is preferable, as there is significant phonological evidence (such as patterns of voicing assimilation, diachronic change, and word-final devoicing) supporting the current mainstream view (Beckman et al., 2013; Y. Y. Cho, 1990; Iverson & Salmons, 1995; Jessen & Ringen, 2002). Rather, it is more likely that the participants in this study have laryngeal phonological specifications similar to speakers of Southern American English (Hunnicutt & Morris, 2016), Swedish (Helgason & Ringen, 2008), or Norwegian (Ringen & van Dommelen, 2013), which is to say that [voice] and [spread glottis] are both privative features, and exactly one is specified for each of the two English laryngeal categories.

Another explanation of these results is that participants are paying close attention to the phonetics of the Marathi stops (see Durlach & Braida, 1969), and because they can quickly learn to distinguish negative from positive VOT in tokens (Pisoni et al., 1982), they are deciding whether a token is more fortis- or lenis-like by dividing tokens into one of three groups: negative, short-lag, and long-lag VOT, much like participants after training in Pisoni et al.'s (1982) study. While voiced aspirates do not neatly fit into this tripartite categorization (having negative VOT but also some form of aspiration), they might likewise be recognized as generally phonetically distinct from plain voiced tokens. Under this interpretation, features such as [spread glottis] and [voice] play little or no role in ratings, as listeners are identifying these tokens on a purely phonetic basis. However, this account has a serious shortcoming. In Pisoni et al.'s (1982) study, when participants were asked to divide negative, short-lag, and long-lag VOT stops into two categories [ba] and [pa], they grouped both short-lag and negative VOT tokens as [ba], without any feedback from the experimenters (i.e., there was no training indicating which category short-lag tokens should belong to), and this happened regardless of whether the

participants had completed the three-response-category training on the previous day. If English listeners quickly adapt to a highly phonetic listening mode in the present study, it is reasonable to expect they would have done so in Pisoni et al.'s (1982) study as well, but this did not happen. So, since the phonology of English affected listener's judgments in Pisoni et al.'s (1982) study, and since this account supposes a similarity between that study and the present one, one must expect phonology to play a role in explaining the present results as well.

Aside from the phonological claims I have made, the results of this thesis also suggest additional ways of measuring aspiration in English. Most approaches to measuring aspiration have focused on duration – such as VOT (Abramson & Whalen, 2017; Lisker & Abramson, 1964), PVI (Berkson, 2012), Noise Offset Time (Davis, 1994), and After Closure Time (Mikuteit & Reetz, 2007). But this study has shown that English listeners also care about the quality of aspiration, namely its NHR and how much of it was voiceless. Repp (1979) also found evidence of non-durational aspects of aspiration on English listeners' perception of laryngeal categories – he found that English listeners take into account both the duration and intensity of aspiration when judging the laryngeal category of a stop. These kinds of measurements are useful in measuring laryngeal contrasts such as Marathi's (Dmitrieva & Dutta, 2019) and based on my results they are also useful in predicting and explaining the perception of stops with more- or less-breathy releases by listeners whose L1 involves stop aspiration.

4.2. Broader implications

If (some) English speakers have both [voice] and [spread glottis] laryngeal features, then one question is how this more sophisticated mental representation of laryngeal contrasts manifests in other domains. One implication of these results is that English speakers have an advantage in learning additional languages that have more complex relationships between voicing and aspiration such as Marathi or Hindi, which Swedish and Norwegian learners would also share, but which languages that use only [sg] (e.g., Cantonese, German) or only [voice] (e.g., French, Russian) would lack. Here I review how the proposed phonological specification for English could impact non-native language acquisition in terms of the theoretical frameworks I discussed in Section 1.2.2, and I also discuss what evidence from prior work supports or disputes this idea.

4.2.1. Predictions for acquiring additional languages

Following Brown's (1998) theoretical framework, since Swedish, Norwegian, and (some) English listeners have [voice] and [spread glottis] features in their L1, they could use these features to distinguish contrasts in other languages. In learning Marathi, a hypothetical French learner would already have access to contrasts based on the [voice] feature, but would have to learn to use the [sg] feature as well. A Cantonese learner would have access to the [sg] feature, though they might have to adjust their expectations of the phonetic realization of this feature to include the release of voiced aspirates, but would need to learn the [voice] feature to learn the entire four-way Marathi contrast. An English learner, however, might have both [voice] and [sg] features, which are the key features involved in the Marathi contrast. Such a learner would simply have to adjust their phonetic representation of [sg] to include breathy releases in voiced aspirates, and create a category for voiceless inaspirates (with no [voice] or [sg] specification) and voiced aspirates (with both a [voice] and [sg] specification). As summarized in Table 4.1, the Cantonese learner must make this adjustment *and* acquire the [voice] feature, and the French learner must acquire the [spread glottis] feature, but an English learner only needs to adjust their existing [spread glottis] feature's phonetic representation, and even before doing so could acquire the basics of the contrast.

Table 4.1. Acquiring a Marathi-like four-way laryngeal contrast, depending on L1 laryngeal features

Laryngeal feature(s)	Example language	Required adjustment
[spread glottis]	Cantonese, German	Adapt [spread glottis] to include the release of voiced aspirates **and** acquire the [voice] feature
[voice]	French, Russian	Acquire the [spread glottis] feature
[voice] and [spread glottis]	English (based on this study), Swedish, Norwegian	Adapt [spread glottis] to include the release of voiced aspirates

These predictions are supported by other models of second language acquisition as well. Namely, the SLM and PAM both state that when a learner of a language does not have a category for a certain sound in the new language, that sound will be learned relatively quickly (Best & Tyler, 2007; Flege, 1987, 1995). When two non-native phones are similar to one existing L1 category (each one a "similar" sound for SLM, or a "single-category assimilation" for PAM where both non-native sounds are "categorized" as one L1 category), a learner must divide their existing category into two, which is easier than

creating a new category for a previously-unmapped part of their phonological space (Best & Tyler, 2007; Flege, 1987, 1995). When acquiring the Marathi laryngeal categories, speakers of true voicing languages might more easily assimilate both voiceless aspirates and inaspirates as fortis, and voiced aspirates and inaspirates as lenis, and speakers of aspirating languages (e.g., Cantonese) might assimilate voiced aspirates and voiceless aspirates as fortis and voiceless inaspirates and voiced inaspirates as lenis. This would make different elements of the complete four-way contrast more difficult for them to distinguish and ultimately attain. In contrast, English speakers who already make use of both [voice] and [sg] would simply need to combine them in new ways.

4.2.2. Other evidence for these predictions

So far, little evidence has been collected that would confirm or deny the predictions I have just outlined. A study on Swedish, Norwegian, or Southern American English native speakers' ability to learn a four-way laryngeal contrast such as Marathi's would provide the most direct evidence for these predictions, as those language groups have already been shown to use both the [voice] and [spread glottis] features. However, I am unable to find such a study. The most relevant reports I have found are Jackson's (2009) and Guion and Pederson's (2007) studies.

Jackson (2009) has already shown that laryngeal feature specification has an impact on initial discrimination, though her English participants do not show evidence of a [voice] feature. They achieved better-than-chance discrimination on contrasts involving only the [voice] feature (i.e., where [spread glottis] was specified in both or in neither Hindi category), but Jackson (2009) attributes this to low-level acoustic processes rather than any phonological influence. Guion and Pederson's (2007) results, however, are more mixed. Their subjects achieved very high discrimination (89-91% correct) of the Hindi /k-g/ contrast, which is based on [voice] and not [spread glottis], before training. This, along with Pisoni et al.'s (1982) findings that English speakers can quickly learn to identify sounds with negative and short-lag VOT as different, could mean that participants were using the [voice] feature to contrast these sounds. Guion and Pederson (2007) also found that English learners improved on the /b-b^{h}/ contrast but not the /k-g/ contrast. Guion and Pederson (2007, p. 75) use PAM (Best & Tyler, 2007) to suggest that Hindi /k/, /g/, and /b/ were assimilated as good exemplars of the analogous

English lenis stop, while /bʰ/ was assimilated as a deviant lenis stop. However, participants' pre-test scores for /k-g/ were *better* than for /b-bʰ/, which is inconsistent with this explanation – if /k-g/ are good exemplars of one English category and /b-bʰ/ differ in their goodness of fit, PAM (Best, 1995; Best & Tyler, 2007) predicts that /k-g/ would be more difficult to discriminate than /b-bʰ/. It is unclear why learners were able to improve on /b-bʰ/ but not /k-g/, however. A fuller study including more contrasts might be able to reveal a broader pattern.

Guion and Pederson's (2007) pre-test results seem to support that English listeners were using the [voice] feature even before training, while Jackson's (2009) discrimination results do not. There may be several reasons for this discrepancy in results. First, English feature configuration with both [voice] and [sg] specifications has been most robustly shown for Southern American English (Hunnicutt & Morris, 2016; Jacewicz et al., 2009), but it is possible that greater regional trends are at work – although Jackson's (2009) study, like this one, was conducted in Canada (specifically, hers was conducted in Calgary, Alberta). As well, it has been some time since the classic studies on VOT (Flege, 1982; Lisker & Abramson, 1964; Repp, 1979, 1982; Rotunno, 1979), and while it is unlikely that a diachronic change has reached completion in that time, it is possible that one has begun, at least for certain groups. Further variationist work investigating how consistently English speakers really use voicing lead in lenis stops might give a valuable insight into the differences between the results of studies investigating English listeners' perception of four-way laryngeal contrasts.

It is also possible that something about the particular task used in this study is responsible for the evidence that English listeners use the [voice] feature. This study uses a different procedure (i.e., a rating scale) than previous work on perception of this contrast. Polka's (1991) study is the closest, as it is also an identification task of a four-way laryngeal contrast, though it did not use a rating scale and was based on Hindi rather than Marathi.

4.3. Limitations and future directions

Like any study, this thesis has certain limitations. For instance, natural speech was used instead of synthesized speech, which means that the phonetic variables that I analyzed were not carefully controlled for. This, in turn, prevents a clear analysis of

trading relations (Repp, 1982), for example, between prevoicing and release duration, or between NHR and duration of release. Still, the use of natural stimuli is sufficient to show that English listeners used measures associated with the [voice] and [spread glottis] features in their discrimination of Marathi stops, which has been sufficient to show the activity of these two features.

In addition, this study used Figure Eight as a platform for recruiting and running subjects, which was chosen as it frequently offers annotation and transcription jobs, which are broadly similar to this study. However, it is limited in its ability to collect linguistic background information, as typically it is designed to ask only one set of questions of each stimulus (whereas a background questionnaire is only filled out once, not for each sound presented). The background information requested in this study was intentionally limited to ensure a short job, as Figure Eight workers typically try to process jobs as quickly as possible. Any studies on this platform or similar crowdsourcing platforms must be designed with this in mind, especially as it is difficult to ensure that participants complete the full study, since workers are allowed to quit partway through. As a result of the shortened background questionnaire, it is difficult to determine any specific sources of between-participant variability, such as in the ratings of voiceless aspirates.

Future research may improve on these limitations and expand upon the insights presented in this thesis. A study that tests English listeners on synthesized stimuli with both prevoicing and aspiration could reveal interesting patterns of trading relations between the acoustic variables discussed in this thesis and expand our understanding of the acoustic cues to English laryngeal categories. As well, a training study comparing languages with different laryngeal feature specifications could test the hypothesis that speakers of languages with [voice] and [spread glottis] specifications, but which only have two laryngeal categories, would have an advantage in learning a four-way contrast over speakers of a language with only [voice] or [spread glottis] features. For example, based on the predictions I gave in Section 4.2.1, French or Cantonese native speakers would be expected to perform worse in a training task on Marathi laryngeal categories than Swedish or Norwegian native speakers.

The emerging findings of languages with two-way laryngeal contrasts involving both [voice] and [sg] features also pose a question for linguistic typology, as there may

be other languages with similar laryngeal feature specifications that are classically considered to only have one – for example, Helgason and Ringen (2008, p. 624) name Turkish, Swahili, and dialects of Western Armenian as likely candidates, and suggest more generally that this contrast may not be as rare as has been estimated. As well, more phonetically-minded researchers might be interested in whether the acoustic cues associated with voicing or aspiration are always weighted equally – it is theoretically possible that speakers use voicing lead more reliably than aspiration, for example. Most of Helgason and Ringen's (2008) Swedish speakers prevoiced all their lenis stops, but even for those that produced short-lag VOT lenis stops, short-lag VOTs were only found in part of the highest VOT quartile of their lenis tokens. In contrast, Norwegian lenis stops were more evenly split between having voicing lead and short-lag VOT (Ringen & van Dommelen, 2013). As a result, if speakers from different language groups were tested on stimuli that were synthesized to vary by VOT from negative to long-lag VOT, it might be the case that Norwegian speakers would have a higher or more positive VOT boundary between lenis and fortis stops than Swedish. And if they were presented with voiced aspirates, perhaps Norwegians would be able to resolve the contrasting [voice] and [spread glottis] specifications more easily than Swedish ones, since Norwegian lenis stops are less frequently prevoiced.

4.4. Summary and conclusions

In this study, I asked how English listeners identify the four laryngeal categories of Marathi in order to gain an insight into the mental representation of the English laryngeal contrast between fortis (i.e., /p, t, k/) and lenis (i.e., /b, d, g/) stops. I found that prevoicing, release duration, and the quality of a stop release all influence English listeners' perception of laryngeal categories, which suggests that both the [voice] and [spread glottis] feature are active in English. This contradicts the currently accepted view that only the [sg] feature is active – instead, the participants in this study have a laryngeal feature specification similar to speakers of Southern American English (Hunnicutt & Morris, 2016), Swedish (Helgason & Ringen, 2008), and Norwegian (Ringen & van Dommelen, 2013). This also suggests that English learners of other languages may have a greater capacity for learning fortis-lenis contrasts involving the features [voice] (e.g., Russian), [spread glottis] (e.g., Cantonese), or both (e.g., Hindi) than previously thought.

References

Abramson, A. S., & Lisker, L. (1967). Discriminability along the voicing continuum: Cross language tests. *Proceedings of the 6th International Congress of Phonetic Sciences*. The Hague: Mouton de Gruyter.

Abramson, A. S., & Lisker, L. (1985). Relative power of cues: F0 shift versus voice timing. In V. A. Fromkin (Ed.), *Phonetic Linguistics: Essays in honor of Peter Ladefoged* (pp. 25–33). Retrieved from http://www.haskins.yale.edu/Reprints/HL0507.pdf

Abramson, A. S., & Whalen, D. H. (2017). Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics*, *63*, 75–86. https://doi.org/10.1016/j.wocn.2017.05.002

Audacity Team. (2017). *Audacity*. Retrieved from https://www.audacityteam.org/

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

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

Beckman, J., Helgason, P., McMurray, B., & Ringen, C. (2011). Rate effects on Swedish VOT: Evidence for phonological overspecification. *Journal of Phonetics*, *39*(1), 39–49. https://doi.org/10.1016/j.wocn.2010.11.001

Beckman, J., Jessen, M., & Ringen, C. (2013). Empirical evidence for laryngeal features: Aspirating vs. true voice languages. *Journal of Linguistics*, *49*(2), 259–284. https://doi.org/10.1017/S0022226712000424

Benkí, J. (2005). Perception of VOT and first formant onset by Spanish and English speakers. *Proceedings of the Fourth International Symposium on Bilingualism*, 240–248. Retrieved from https://www.msu.edu/~benki/pubs/BenkiISB4.pdf

Bennett, W. G., & Rose, S. (2017). Moro voicelessness dissimilation and binary [voice]. *Phonology*, *34*(3), 473–505. https://doi.org/10.1017/S0952675717000252

Berkson, K. H. (2012). Capturing Breathy Voice: Durational Measures of Oral Stops in Marathi. *Kansas Working Papers in Linguistics*, *33*, 27–46.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience. Issues in cross-language research* (pp. 171–204). https://doi.org/10.1016/0378-4266(91)90103-S

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of Perceptual Reorganization for Nonnative Speech Contrasts: Zulu Click Discrimination by English-Speaking Adults and Infants. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 345–360. https://doi.org/10.1037/0096-1523.14.3.345

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 1–47). Amsterdam: John Benjamins.

Boersma, P., & Weenink, D. (2020). *Praat: doing phonetics by computer*. Retrieved from http://www.praat.org/

Brown, C. A. (1998). The role of the L1 grammar in the L2 acquisition of segmental structure. *Second Language Research*, *14*(2), 136–193. https://doi.org/10.1191/026765898669508401

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics*, *27*(2), 207–229. https://doi.org/10.1006/jpho.1999.0094

Cho, T., Whalen, D. H., & Docherty, G. (2019). Voice onset time and beyond: Exploring laryngeal contrast in 19 languages. *Journal of Phonetics*, *72*, 52–65. https://doi.org/10.1016/j.wocn.2018.11.002

Cho, Y. Y. (1990). Is Voicing a Privative Feature? *Arizona Phonology Conference*, *3*, 34–48.

Davis, K. (1994). Stop Voicing in Hindi. *Journal of Phonetics*, *22*(2), 177–193. https://doi.org/10.1016/S0095-4470(19)30192-5

Dmitrieva, O., & Dutta, I. (2017). Onset f0 as a correlate of voicing in Marathi. *The Journal of the Acoustical Society of America*, *142*(4), 2585–2585.

Dmitrieva, O., & Dutta, I. (2019). Acoustic Correlates of the Four-Way Laryngeal Contrast in Marathi. *Phonetica*, *47906*. https://doi.org/10.1159/000501673

Dmitrieva, O., Llanos, F., Shultz, A. A., & Francis, A. L. (2015). Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics*, *49*, 77–95. https://doi.org/10.1016/j.wocn.2014.12.005

Durlach, N. I., & Braida, L. D. (1969). Intensity Perception. I. Preliminary Theory of Intensity Resolution. *The Journal of the Acoustical Society of America*, *46*(2B), 372–383. https://doi.org/10.1121/1.1911699

Dutta, I. (2007). *Four-Way Stop Contrasts in Hindi: An Acoustic Study of Voicing, Fundamental Frequency and Spectral Tilt*.

Flege, J. E. (1982). Laryngeal timing and phonation onset in utterance-initial English stops. *Journal of Phonetics*, *10*(2), 177–192. https://doi.org/10.1016/S0095-4470(19)30956-8

Flege, J. E. (1987). The production of “new” and “similar” phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*(1), 47–65. https://doi.org/10.1016/s0095-4470(19)30537-6

Flege, J. E. (1995). Second Language Speech Learning: Theory, Findings, and Problems. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 233–277). https://doi.org/10.1111/j.1600-0404.1995.tb01710.x

Flege, J. E., & Eefting, W. (1986). Linguistic and developmental effects on the production and perception of stop consonants. *Phonetica*, *43*(4), 155–171. https://doi.org/10.1159/000261768

Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression* (Third). Retrieved from https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Francis, A. L., Kaganovich, N., & Driscoll-Huber, C. (2008). Cue-specific effects of categorization training on the relative weighting of acoustic cues to consonant voicing in English. *The Journal of the Acoustical Society of America*, *124*(2), 1234–1251. https://doi.org/10.1121/1.2945161

Gordon, P. C., Eberhardt, J. L., & Rueckl, J. G. (1993). Attentional Modulation of the Phonetic Significance of Acoustic Cues. *Cognitive Psychology*, *25*(1), 1–42.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds “L” and “R.” *Neuropsychologia*, *9*(3), 317–323. https://doi.org/10.1016/0028-3932(71)90027-3

Guion, S. G., & Pederson, E. (2007). Investigating the role of attention in phonetic learning. In O.-S. Bohn & M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning: In honor of James Emil Flege* (pp. 57–77). https://doi.org/10.1075/lllt.17.09gui

Haggard, M., Ambler, S., & Callow, M. (1970). Pitch as a Voicing Cue. *The Journal of the Acoustical Society of America*, *47*(2B), 613–617. https://doi.org/10.1121/1.1911936

Haggard, M., Summerfield, Q., & Roberts, M. (1981). Psychoacoustical and cultural determinants of phoneme boundaries: evidence from trading F0 cues in the voiced-voiceless distinction. *Journal of Phonetics*, *9*(1), 49–62. https://doi.org/10.1016/s0095-4470(19)30926-x

Helgason, P., & Ringen, C. (2008). Voicing and aspiration in Swedish stops. *Journal of Phonetics*, *36*, 607–628. https://doi.org/10.1016/j.wocn.2008.02.003

Hillenbrand, J; Cleveland, R; Erickson, R. (1994). Acoustic Correlates of Breathy Vocal Quality. *Journal of Speech and Hearing Research*, Vol. 37, pp. 298–306. https://doi.org/10.1044/jshr.3302.298

Honeybone, P. (2012). Diachronic evidence in segmental phonology: the case of obstruent laryngeal specifications. *The Internal Organization of Phonological Segments*, (2001), 317–352. https://doi.org/10.1515/9783110890402.317

Hunnicutt, L., & Morris, P. (2016). Pre-voicing and aspiration in Southern American English. *University of Pennsylvania Working Papers in Linguistics*, *22*(1), 215–224. https://doi.org/10.1017/CBO9781107415324.004

Iverson, G. K., & Salmons, J. C. (1995). Aspiration and Laryngeal Representation in Germanic. *Phonology*, *12*(3), 369–396. https://doi.org/10.1017/S0952675700002566

Jacewicz, E., Fox, R. A., & Lyle, S. (2009). Variation in stop consonant voicing in two regional varieties of American English. *Journal of the International Phonetic Association*, *39*(3), 313–334.

Jackson, S. (2009). *Non-Native Perception of Laryngeal Features*. University of Calgary.

Jessen, M., & Ringen, C. (2002). Laryngeal features in German. *Phonology*, *19*(2), 189–218. https://doi.org/10.1017/S0952675702004311

Kane, J., & Gobl, C. (2011). Identifying regions of non-modal phonation using features of the wavelet transform. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, (August), 177–180.

Keating, P. A. (1984). Phonetic and Phonological Representation of Stop Consonant Voicing. *Language*, *60*(2), 286–319.

Keating, P. A. (1993). Comments on privative versus binary features. *UCLA Working Papers in Phonetics*, 1–5.

Keating, P. A., Mikoś, M. J., & Ganong, W. F. I. (1981). A cross-language study of range of voice onset time in the perception of initial stop voicing. *Journal of the Acoustical Society of America*, *70*(5), 1261–1271. https://doi.org/10.1121/1.387139

Kingston, J., & Diehl, R. L. (1994). Phonetic Knowledge. *Language*, *70*(3), 419–454.

Klatt, D. H. (1973). Aspiration and Voice Onset Time in Word-Initial Consonant Clusters in English. *The Journal of the Acoustical Society of America*, *54*(1), 319–319. https://doi.org/10.1121/1.1978269

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2020). *emmeans: Estimated Marginal Means, aka Least-Squared Means*. Retrieved from https://cran.r-project.org/web/packages/emmeans/index.html

Lesho, M. (2018). Philippine English (Metro Manila acrolect). *Journal of the International Phonetic Association*, *48*(3), 357–370. https://doi.org/10.1017/S0025100317000548

Lisker, L. (1986). “Voicing” in English: a catalogue of acoustic features signaling/b/versus/p/in trochees. *Language and Speech*, *29*(1), 3–11. https://doi.org/10.1177/002383098602900102

Lisker, L., & Abramson, A. S. (1964). A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. *WORD*, *20*(3), 384–422. https://doi.org/10.1080/00437956.1964.11659830

Mikuteit, S., & Reetz, H. (2007). Caught in the ACT: The timing of aspiration and voicing in East Bengali. *Language and Speech*, *50*(2), 247–277. https://doi.org/10.1177/00238309070500020401

Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-Day-Olds Prefer Their Native Language. *Infant Behavior and Development*, *16*(4), 495–500. https://doi.org/10.1016/0163-6383(93)80007-U

Nearey, T. M., & Rochet, B. L. (1994). Effects of place of articulation and vowel context on VOT production and perception for French and English stops. *Journal of the International Phonetic Association*, *24*(1), 1–18.

Ohala, J. J. (1996). Speech perception is hearing sounds, not tongues. *The Journal of the Acoustical Society of America*, *99*(3), 1718–1725. https://doi.org/10.1121/1.414696

Ohde, R. N. (1984). Fundamental frequency as an acoustic correlate of stop consonant voicing. *Journal of the Acoustical Society of America*, *75*(1), 224–230. https://doi.org/10.1121/1.390399

Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance*, *8*(2), 297–314. https://doi.org/10.1037/0096-1523.8.2.297

Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *The Journal of the Acoustical Society of America*, *89*(6), 2961–2977. https://doi.org/10.1121/1.400734

R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. Retrieved from https://www.r-project.org/

Repp, B. H. (1979). Relative Amplitude of Aspiration Noise as a Voicing Cue For Syllable-initial Stop Consonants. *Language and Speech*, *22*(2), 173–189.

Repp, B. H. (1982). Phonetic Trading Relations and Context Effects: New Experimental Evidence for a Speech Mode of Perception. *Psychological Bulletin*, *92*(I), 81–110.

Ringbom, H. (1992). On L1 Transfer in L2 Comprehension and L2 Production. *Language Learning*, *42*(1), 85–112. https://doi.org/10.1111/j.1467-1770.1992.tb00701.x

Ringen, C., & van Dommelen, W. A. (2013). Quantity and laryngeal contrasts in Norwegian. *Journal of Phonetics*, *41*(6), 479–490. https://doi.org/10.1016/j.wocn.2013.09.001

Rotunno, R. (1979). Relation between voice-onset time and vowel duration. *Journal of the Acoustical Society of America*, *66*(3), 654–662. https://doi.org/10.1121/1.383692

RStudio Team. (2019). *RStudio: Integrated Development for R*. Retrieved from http://www.rstudio.com/

Shultz, A. A., Francis, A. L., & 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. https://doi.org/10.1121/1.4736711

Strange, W., Bohn, O.-S., Nishi, K., & Trent, S. a. (2005). Contextual variation in the acoustic and perceptual similarity of North German and American English vowels. *Journal of the Acoustical Society of America*, *118*(3), 1751–1762. https://doi.org/10.1121/1.1992688

Strange, W., Levy, E. S., & Law, F. F. (2009). Cross-language categorization of French and German vowels by naïve American listeners. *Journal of the Acoustical Society of America*, *126*(3), 1461–1476. https://doi.org/10.1121/1.3179666

Tees, R. C., & Werker, J. F. (1984). Perceptual flexibility: maintenance or recovery of the ability to discriminate non-native speech sounds. *Canadian Journal of Psychology*, *38*(4), 579–590. https://doi.org/10.1037/h0080868

Walker, A. (2020). Voiced stops in the command performance of Southern US English. *The Journal of the Acoustical Society of America*, *147*(1), 606–615. https://doi.org/10.1121/10.0000552

Weismer, G. (1979). Sensitivity of voice-onset time (VOT) measures to certain segmental features in speech production. *Journal of Phonetics*, *7*(2), 197–204. https://doi.org/10.1016/S0095-4470(19)31041-1

Werker, J. F., Gilbert, J. H. V., Humphrey, K., & Tees, R. C. (1981). Developmental Aspects of Cross-Language Speech Perception. *Child Development*, *52*(1), 349–355.

Wetzels, W. L., & Mascaró, J. (2001). The Typology of Voicing and Devoicing. *Language*, *77*(2), 207–244.

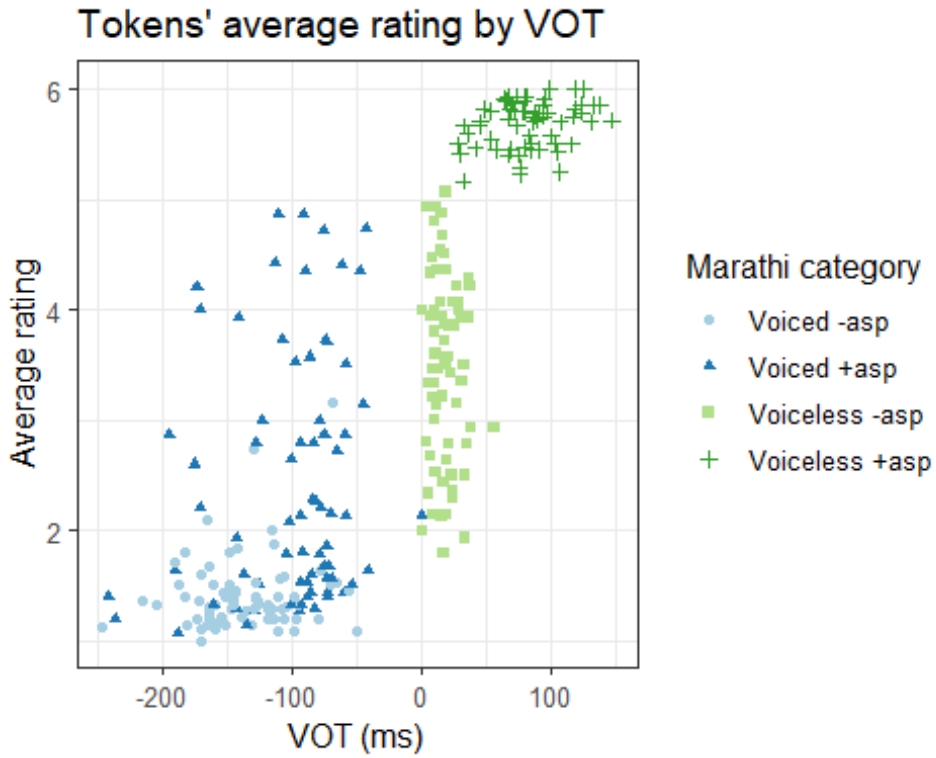
Whalen, D. H., Abramson, A. S., Lisker, L., & Mody, M. (1993). F0 gives voicing information even with unambiguous voice onset times. *Journal of the Acoustical Society of America*, *93*(4), 2152–2159. https://doi.org/10.1121/1.406678

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Retrieved from http://ggplot2.org

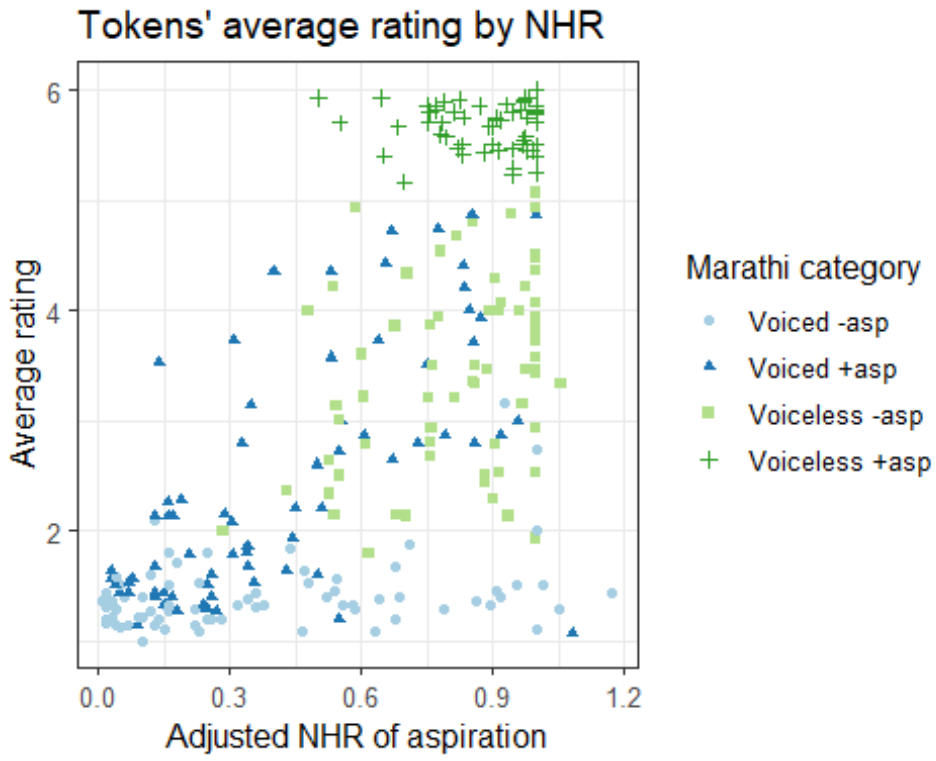
Appendix.

Plots of acoustic factors and ratings of all Marathi laryngeal categories

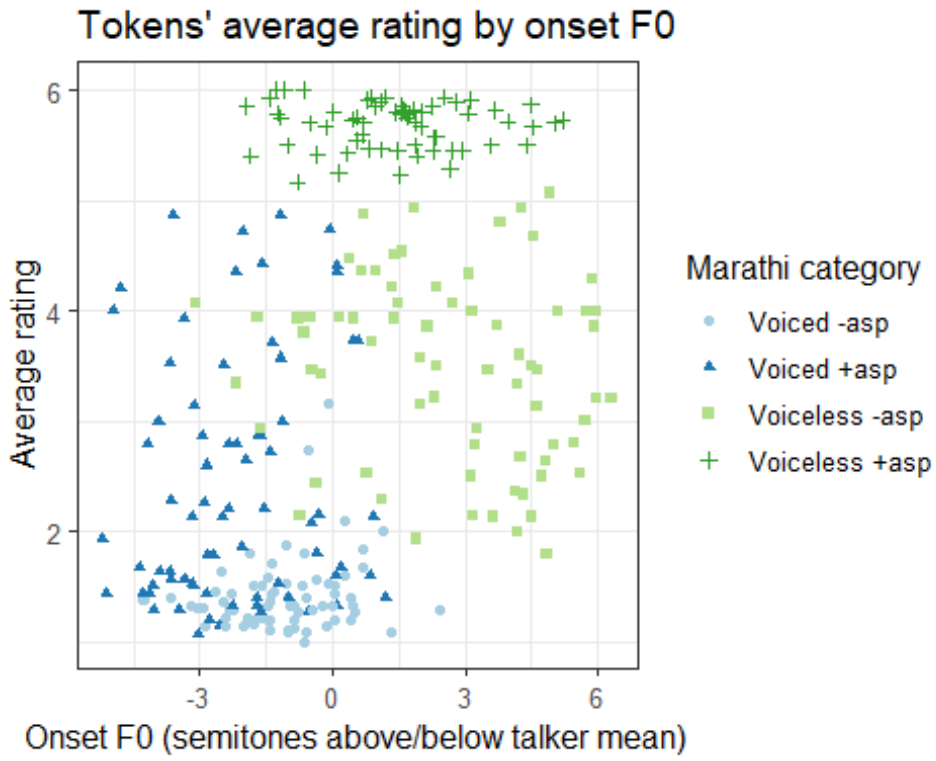
This appendix visualizes the relationship between the independent variable and fixed effects in the model described in Section 3.3.1.



Higher ratings indicate greater similarity to English fortis stops. VOT was the strongest predictor of participants' ratings.



NHR was the second-strongest predictor of participants' ratings.



Onset F0 was the third-strongest predictor of participants' ratings.