Title

Realization and representation of Nepali laryngeal contrasts

Author

Schwarz, Martha Reading

Publication Date

2024

Peer reviewed|Thesis/dissertation

Realization and Representation of Nepali Laryngeal Contrasts

by

Martha Reading Schwarz

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Linguistics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Sharon Inkelas, Co-chair
Professor Darya Kavitskaya, Co-chair
Professor Keith Johnson,
Professor Gasper Begus

Realization and Representation of Nepali Laryngeal Contrasts

Abstract

Realization and Representation of Nepali Laryngeal Contrasts

by

Martha Reading Schwarz

Doctor of Philosophy in Linguistics

University of California, Berkeley

Professor Sharon Inkelas, Co-chair

Professor Darya Kavitskaya, Co-chair

This dissertation investigates the phonetics and phonology of laryngeal contrast in Nepali, an Indic (aka Indo-Aryan) language with a four-way laryngeal contrast in its stops and affricates between classes often described as voiceless, voiceless aspirated, voiced, and voiced aspirated. Specifically, it brings Nepali data to bear on three phenomena central to the study of laryngeal contrasts: representation of laryngeal contrast, the effects of laryngeal specification on preceding vowel duration, and patterns of neutralization and weakening of laryngeal contrasts. Despite the diversity in language-specific phonetic cues used to realize laryngeal contrasts (voicing, aspiration, etc), each of these phenomena is sometimes referred to as phenomena of "voicing": voicing contrasts, the voicing effect on pre-coda vowels, and final devoicing. There is now a body of literature that grapples with the connections between these phenomena and the language-specific cues used to realize contrast, and their theoretical implications. But despite an ever-increasing amount of phonetic and phonological work in languages with complex laryngeal contrasts, these theoretical conversations and proposals often still focus on languages with two-way contrasts, asking: do these phenomena behave the same in languages with voicing contrasts vs. languages with aspiration contrasts? Should the segments be specified for the same laryngeal features in both types of languages? This dissertation asks: how do these phenomena behave in a language with both a voicing contrast and and aspiration contrast, including a class with both voicing and aspiration?

Each chapter of this dissertation addresses this question in light of a different phenomenon typically described as "voicing". Chapter 2 presents a production study examining the phonetics of "voicing" contrasts in initial position and measures of speaker control proposed to diagnose feature representation in initial and intervocalic positions, finding evidence for representations using [voice] and [spread] features. Chapter 3 engages with the longstanding question of how and why laryngeal specification of a consonant affects the preceding vowel's duration. In order to assess a variety of explanations for the source of this effect, it examined the effects on vowel duration of not

just voicing specification but aspiration specification as well. The results are problematic only for exclusively articulatory explanations, but consistent with perceptual explanations, and with analyses in which multiple factors and forces affect vowel duration. Through both production and perception studies, Chapter 4 shows that the label of "final devoicing" is insufficient to capture the patterns of contrast maintenance and weakening that target Nepali's aspiration and voicing contrasts in word-final position. It finds that the voicing contrast remains robust, while the aspiration contrasts is weakened, confused, and shifted.

This dissertation proposes that Nepali motivates a tighter connection between phonetics and phonological representation than whether segments should be specified for the same laryngeal features in voicing vs. aspirating languages, arguing for laryngeal specification on a subsegmental level.

Contents

List of Figures

List of Tables

Acknowledgments

First and foremost, this project would not exist if not for the support, contributions, and insights of the Nepali-speaking communities I worked with. In particular, Mina Rana, Chandra Rana, and Sanjib Pokhrel who first introduced me to the sounds and structures of Nepali, and who modelled the clear, patient, and joyful language teaching I now strive to provide for my own language students. Shrijana Chhetri and Samar Sinha, who welcomed me into their homes and departments at Sikkim University in Gangtok. Nirmala and Raj Karmacharya, the Regmi family, and the Kumal family (Durga, Shivakumari, Reeya, and Reetu) for being my homes and families in Kathmandu, Dabhung Thanti, and Gorkha. To Durga Kumal especially, who was a collaborator, thought partner, and teacher in everything Kumal-related. To everyone who participated in the studies presented in this thesis, now matter how tedious. And to Pramish Thapa for helping design the perception study.

An obvious and enormous thank you goes to my committee, Gašper Beguš, Sharon Inkelas, Keith Johnson, and Darya Kavitskaya, for their patience and support in the dissertation writing stage of this program. Especially as my belief in the project and myself waxed and waned, I reliably found myself re-inspired after conversations with them. In particular, I owe huge thanks to my advisor Sharon. When I attended her talk on Q-Theory at NELS, long before I even applied to Berkeley, I was immediately struck by the clarity and intuitiveness of her proposals. I've been so grateful for this ability to see straight to the heart of the issue, linguistic and beyond ever since.

I'm grateful to my MA advisors at McGill, Morgan Sonderegger and Heather Goad. Heather shared her endless enthusiasm and careful eye for precision in phonological representation. This dissertation really began in Morgan's seminar on laryngeal contrasts. During and beyond this class he has always had a gift for identifying the exciting kernels in the results of our studies.

Even earlier, my love for linguistics and my belief in the value of linguistics education was born in Lotus Goldberg's intro linguistics class my first semester at Brandeis. In that class and the next 5 classes I took with Lotus, I learned new ways to think about language and witnessed a master class in how to bring wisdom, warmth, enthusiasm, and deep care to your students. I've been extremely grateful for Lotus's mentorship in key moments, even long after I've graduated. I have a lot of fondness for the small but mighty Brandeis linguistics program more broadly, with uniformly excellent mentors and teachers including Sophia Malamud, Keith Plaster, Carl Sharif El-Tobgui, and Janet McIntosh.

There are so many people at Berkeley who have supported me over the years. I owe many thanks to Belén Flores, Paula Floro, Martine Alexander, Johnny Morales Arellano, and Susan Luong, who have all kept the linguistics department running with expertise and kindness. I was drawn to Berkeley in large part because of the collaborative, caring, and invested graduate student community and I'm so grateful to have been a part of it. I'm particularly grateful for time spent, classes taught, Qs subsegmented, and offices shared with Edwin Ko, Tyler Lemon, Ginny Dawson, Noah Hermalin, Zach O'Hagan, Myriam Lapierre, Karee Garvin, Julia Peck, and Dakota Robinson. To Schuyler Laparle for somehow always being on the same page; to Allegra Robertson for silliness, bike rides, pandemic runs, fancy pool swims, and a shared love for laryngeal contrasts when we remember to talk about them; and to my "mentor" Julia Nee whose friendship and joy has spanned the spheres of work, home, and big adventures. To all my undergrad students at Cal and SF State who kept me inspired, gave me (sometimes a little too much) to do besides working on my research, and who remind me why any of it is important.

The linguistics department brought me to Berkeley, and I'm lucky to have found deep community here. For the love, exploration, and consistency they've brought to my life in homes, on bikes, on trails, and at shul, I'd like to thank Sarah Smith, Ari Eisen, Hanah Goldov, Anna Luberoff, Elaina Marshalek, Emily McAfee, Frankie Sandmel, Max Walker, Leni Hirsch, Arielle Tonkin, Amy Van Scoyoc, Tarek Fouda, Lauren Cohen Fisher, Lizzie Gill, Jesse Gruber, Ted Maruyama, Jonah Sternthal, Milo Johnson, Katha Hajek, Julia Nee (again), Calliope Desenberg, Guadalupe Verta, Vivek Kembaiyan, and Uriel Wolfe-Blank. For friends who are farther away but equally dear: Ellen Hamel, Sam Italiano, Yael Rhodes, Mira Smith, Tali Smookler, and so many more.

And to my family. To my grandmother Trudy Schwarz, who I believe is responsible for my love of patterns and language, and who, at 105, continues to wow me with her Spelling Bee prowess. To my brother, Theo Schwarz for impromptu visits that always left my house and car better than he found them. To my parents Sarah Luria and Tom Schwarz, who are two of my favorite people to be around. Whose love comes through in their eagerness to share ambitious bike trips and dedication to understanding laryngeal contrasts and split ergativity.

I am incredibly grateful.

Chapter 1

Introduction

This dissertation investigates the phonetics and phonology of laryngeal contrast in Nepali, an Indic (aka Indo-Aryan) language with a four-way laryngeal contrast in its stops and affricates between classes often described as voiceless, voiceless aspirated, voiced, and voiced aspirated. Specifically, it brings Nepali data to bear on three phenomena central to the study of laryngeal contrasts: representation of laryngeal contrast, the effects of laryngeal specification on preceding vowel duration, and patterns of neutralization and weakening of laryngeal contrasts. Despite the diversity in language-specific phonetic cues used to realize laryngeal contrasts (voicing, aspiration, etc), each of these phenomena is sometimes referred to as phenomena of "voicing": voicing contrasts, the voicing effect on vowels, and final devoicing. There is now a body of literature that grapples with the connections between these phenomena and the language-specific cues used to realize contrast, and their theoretical implications. But despite an ever-increasing amount of phonetic and phonological work in languages with complex laryngeal contrasts, these theoretical conversations and proposals often still focus on languages with two-way contrasts, asking: do these phenomena behave the same in languages with voicing contrasts vs. languages with aspiration contrasts? Should the segments be specified for the same laryngeal features in both types of languages? This dissertation asks: how do these phenomena behave in a language with both a voicing contrast and and aspiration contrast, including a class with both voicing and aspiration? I propose that Nepali motivates a tighter connection between phonetics and phonological representation than whether segments should be specified for the same laryngeal features in voicing vs. aspirating languages, arguing for laryngeal specification on a subsegmental level.

1.1 Background: Laryngeal contrasts

In this section, I present relevant background on laryngeal contrasts and three of the core areas of conversation concerning them: realization and representation of the contrasts, their effect on vowel duration, and neutralization of the contrasts. Each of these areas will be expanded upon further in their dedicated chapter.

Realization and representation

Realization

Many languages have two-way contrasts, traditionally called 'voicing', between two categories of obstruents which we transcribe as *p, t, k* and *b, d, g*. It is well established that different languages realize this 'voicing' contrast with different phonetic cues (e.g. Lisker and Abramson 1964; Abramson and Whalen 2017). It is also well established that even within a single language the contrast can be realized with different phonetic cues depending on a segment's position in a word, utterance, or syllable. Attention has focused on phrase-initial position, as this is the position where cues are maximally contrastive. Some languages (like French) do employ phonetic voicing (i.e. vocal fold vibration) in initial position and contrast a negative VOT[1] (lead time) with a positive VOT (lag time). Other languages (like English and German) do not reliably show phonetic voicing in either of the stop series, but instead contrast a short positive VOT (short lag) with a long positive VOT (long lag).

Representation

A large body of literature debates the relationship between phonological feature representation and phonetic realization. Many agree that there is some link between the two (e.g. Jakobson et al. 1952; Clements 1985; Flemming 2005; Mielke 2008; but see Hale and Reiss 2008; Iosad 2012; Reiss 2017 for substance-free views of phonology), but it is debated how direct this link should be. This relationship pertains to features of all kinds, though the laryngeal features ([voice], [spread glottis], [constricted glottis]) and the contrasts they induce provide a particularly good test case: languages with equivalently complex laryngeal contrasts realize those contrasts in phonetically distinct ways, raising the question of whether features should reflect the particular realizations in the language, or just the complexity of the contrast. Within the domain of laryngeal contrasts (which includes voicing, aspiration, glottalization, etc.), the literature has focused primarily on two-way contrasts traditionally described as 'voicing', and the [voice] and [spread] features; in particular, on whether a 'voicing' contrast in a given language is better represented with either a [voice] or a [spread] feature. Should this two-way contrast be represented the same way in both English and French? Or should it be represented differently, reflecting the different phonetic realizations? Honeybone (2005) summarizes work which supports each of these views.

In the first approach, [voice] is used to capture the contrast between *b, d, g* and *p, t, k* in both types of languages (Chomsky and Halle 1968; P. A. Keating 1984; Rubach 1990; Lombardi 1991; Wiese 1996; Lombardi 1999), capturing the observation that on the phonological level there is a two-way laryngeal contrast between these stop classes. The phonological representation of the contrast is an abstraction from the phonetic realization of the classes, such that any two-way contrast on the VOT spectrum will be marked with the same feature no matter where on the spectrum

[1] In initial position, VOT (Voice Onset Time) is defined as the duration before the onset of voicing of the following vowel. For prevoiced segments in which voicing begins during the stop closure, the duration of voicing before the release of the stop is measured as negative VOT. In segments without prevoicing, VOT is the aspiration duration: the positive measure of how long it takes for voicing to begin after the release of the stop.

each class falls. Thus the same phonological primitive could be phonetically realized with prevoicing as in French, or with a short-lag VOT as in English and German. We will refer to this as the 'abstract' approach.[2]

The abstract approach contrasts with theories of 'laryngeal realism' (e.g. Harris 1994; Iverson and Salmons 1995; Avery and Idsardi 2001; Jessen and Ringen 2002; Iverson and Salmons 2003; Honeybone 2005; Vaux and Samuels 2005; Beckman, Jessen et al. 2013; Brown 2016) which propose that the difference in phonetic realization across these types of languages should be underlyingly represented by different phonological features. Languages like French (henceforth 'voicing languages') have a [voice] feature corresponding to the voicing cue that distinguishes the two classes, while languages like English (henceforth 'aspirating languages') have a [spread glottis] or [spread] feature corresponding to the aspiration duration that distinguishes the two classes.[3]

Table 1.1: Abstract and realist representations of voicing and aspirating languages

Abstract	English		French	
	b, d, g p, t, k	[voice] contrast	b, d, g p, t, k	[voice] contrast
Realist	English		French	
	b, d, g p, t, k	[spread] contrast	b, d, g p, t, k	[voice] contrast

In addition to whether the features should be represented as abstract [voice], or realist [voice] and [spread], there is further debate over specification—whether the features are binary or privative. In a binary system, the stop classes are specified as [+voice] and [-voice] in a voicing language and [+spread] and [-spread] in an aspirating language, with each class being specified for opposing values of that feature[4]. In a privative system, the presence of a feature on one class (represented with [voice] or [spread]) contrasts with the lack of specification in another class. Asymmetry in feature specification predicts asymmetric phonological behavior of the segments. The crucial

[2]This corresponds to what Honeybone (2005) refers to as the 'traditional' approach.

[3]While we call this a 'realist' approach and contrast it with the 'abstract' approach, it should be noted that even realist representations abstract away from the phonetics to some degree. English *p,t,k*, for example, are sometimes realized with glottal constriction rather than long-lag aspiration, but are still represented with a [spread] feature in laryngeal realism. We call the first approach 'abstract' to highlight that it is *more* abstracted from the phonetics than the realist approach, not because it is the *only* system to abstract away from phonetics at all.

[4]This does not preclude the possibility that only one value of a feature is specified underlyingly and throughout early parts of the derivation, as in, for example, the theory of Radical Underspecification e.g. Archangeli 1988. However, if features are inherently bivalent, then there should exist languages where, for any given feature, one value is specified underlyingly in some languages while the other value is specified underlyingly in others (see e.g. Archangeli and Pulleyblank 1989 on [atr] in Yoruba vs. Masaai and Abaglo and Archangeli 1989 on [high] in Yoruba vs. Gengbe). This does not hold for privative systems, as discussed below in the text.

difference between binary and privative representations is that with privative features, lack of specification and the asymmetry that it entails is always in one direction.

The representations proposed by laryngeal realism are both realist and privative. Table 1.2 shows the laryngeal realist representations for a voicing language (French), an aspirating language (English), and a language with both voicing and aspirating contrasts (Thai). Segments unspecified for any laryngeal feature are represented in this dissertation with empty brackets [ ]. By using privative features, laryngeal realism predicts asymmetries between stop classes in speaker intentionality behind phonetic realization, in variability of realization, and in phonological processes, as discussed in Chapter 2.

Table 1.2: Two- and three-way contrasting languages and their laryngeal realist representations

	[b, d, g]	[p, t, k]	[p^h, t^h, k^h]
French	[voice]	[ ]	——
English	——	[ ]	[spread]
Thai	[voice]	[ ]	[spread]

The voicing effect

The second phenomenon I consider is the voicing effect. A long history of work has reported cross-linguistic evidence for a 'voicing effect' in which vowels tend to be longer before voiced consonants than before voiceless consonants (e.g. House and Fairbanks 1953; Chen 1970; Coretta 2019), though this effect is not universal (P. Keating 1984; Mitleb 1984). Furthermore, it is not clear what drives this effect. Various explanations have been proposed, grounded in articulatory, timing, or perceptual reasons. In articulatory-based explanations, vowel length is intrinsically or phonetically linked to some aspect of the articulation of the stop's voicing. Timing-based explanations derive from the posited drive for speech units to be consistently timed, even when the sounds that make up those units vary: closures of voiced stops tend to be shorter than voiceless stops in a vowel-coda VC sequence, vowels preceding voiced stops will be longer to compensate. In perception-based explanations, vowel duration is manipulated by speakers to enhance or fill in for weaker cues to voicing specification. Some accounts of the voicing effect appeal to more than one of these sources. English, which displays a larger voicing effect than most other languages, is proposed to have taken a potentially intrinsic difference in vowel length and phonologized it, such that vowel length is now one of the primary cues to stop voicing in post-vocalic positions (e.g. Kluender et al. 1988).

Related to these proposals for the source of the voicing effect is another question: to what degree is 'voicing' used in its abstract interpretation here? The effect exists both in aspirating languages like English and German (House and Fairbanks 1953; Braunschweiler 1997) and in voicing languages like Italian and Polish (Coretta 2019), but laryngeal realist theories thus far have tended not to focus on the voicing effect – perhaps for this very reason. Even realist representations are abstracted to some degree from the phonetics of the contrast, and particularly in word-final positions, in which positive VOT cues to aspiration contrast are less robust, aspirating languages may rely on other cues to realize the contrast (English's reliance on vowel length, for example). Languages with a two-way contrast could conceivably rely on any cues while still maintaining their contrast. A language with a four-way contrast, however, may need to use cues specific to voicing and aspiration contrasts in order to maintain the four-way contrast.

Laryngeal contrast weakening and neuralization

The third phenomenon often abstracted simply to 'voicing' is the weakening of laryngeal contrasts in word- or syllable-final position. This phenomenon is sometimes referred to as simply "final devoicing", but doing so again abstracts over the diversity of the types of laryngeal contrasts that are neutralized and the results of their neutralization. German, for example, is commonly cited as a case of final devoicing: stops represented orthographically as <b,d,g> and <p,t,k> neutralize to the <p,t,k> series in final position. In an abstract representation, <b,d,g> stops have a [voice] feature and <p,t,k> stops do not, so neutralization to the <p,t,k> series would indeed be devoicing. In a realist representation, however, German <p,t,k> stops would be represented with a [spread] feature because. If we take this representation seriously, neutralization to the <p,t,k> series involves addition of a [spread] feature rather than loss of a [voice] feature.

In addition to languages like German that can only be analyzed as devoicing within a particular (abstract) theoretical framework, Vaux and Samuels (2005) and Iverson and Salmons (2011) assemble a sizeable set of languages whose empirical patterns of final laryngeal contrast reduction challenge classifying it uniformly as devoicing in any theoretical framework. Some such languages include Kashmiri, in which plain voiceless and voiceless aspirated stops reportedly neutralize to the aspirated class in syllable-final position (Vaux and Samuels 2005), and which may be analyzed as addition of a [spread] feature. Similarly, the three-way stop contrast of Klamath, /k, k, k'/ are all realized as [k] word-finally. The neutralization pattern in Sanskrit also challenges the notion that all laryngeal neutralization is simply devoicing because it unambiguously involves loss of a [spread] feature in addition to loss of the [voice] feature: its four-way contrast of T, Th, D, Dh all neutralize to T in final position (I. Dutta 2007; Iverson and Salmons 2011). Iverson and Salmons (2011) also cite the often discussed but controversial case of final voicing in Lezgian (Haspelmath 1993; Yu 2004; Blevins 2006) as addition of a [voice] feature.

If languages potentially reduce laryngeal contrasts in final position by losing a [voice] feature (e.g. Dutch), adding a [spread] feature (e.g. German, Klamath), losing a [spread] feature (e.g. Sanskrit), or adding a [voice] feature (e.g. Lezgian), it further challenges either the notion that contrasts universally neutralize to the unmarked class, or that the unmarked class is universally T, the voiceless unaspirated class. Vaux and Samuels (2005) take up the latter, maintaining that languages neutralize to the unmarked realization but arguing both that the unmarked class is not the voiceless unaspirated but a stop unspecified for any laryngeal features (or gestural targets), and that the identity of the unmarked class may differ crosslinguistically. The difference between an unspecified stop and a voiceless aspirated stop, they argue, is that an unspecified stop might be expected to vary over phonetic parameters that cue laryngeal contrasts, displaying a wider range of prevoicing and VOT durations than stops belonging to a particular specified class (see also Steriade (1997), Wilson (2001) and Hauser (2019)). In this view, loss of a [spread] feature may actually be realized with a wide and variable range of VOT durations rather than with consistently short VOT durations that would be expected of the unaspirated member of a contrast in an aspirating language. Vaux and Samuels (2005) also argue that Th is in fact the unmarked member of the T/Th contrast in aspirating languages: markedness cannot be defined in a single way crosslinguistically but rather depends on language-specific properties, including the types of contrasts in a language's inventory (Rice 1999; Flemming 2005).

As in initial position, debates over the proper representation and neutralization process in a language with a single laryngeal contrast, like German, can be seen as simply theoretical. A two-way laryngeal contrast can, conceivably, be represented with either a [voice] or a [spread] distinction. In languages like Nepali, however, a single feature can never be sufficient to represent the four-way laryngeal contrast. Furthermore, languages with both voicing and aspiration contrasts provide a particularly fruitful test case for examining contrast reduction since they present the potential of maintaining or reducing both voicing and aspiration contrasts independent of each other. Indeed, Indic languages with underlyingly four-way laryngeal contrasts display a range of contrast reduction patterns, explored in the next section.

1.2 Background: Nepali

While this dissertation deals primarily with the acoustics and representations of particular sounds in Nepali, it is important to situate the language in the context of the people who speak it. This section first presents background on the history and role of Nepali in Nepal, followed by background on some of its pertinent linguistic traits.

Linguistic context of Nepal

A brief language-centred history of Nepal

The earliest evidence of humans in the area now known as Nepal dates back about 100,000 years. In the time since then numerous groups have migrated into the area. Speakers of primarily Indo-Aryan languages came up over the flatlands from the south and relatively more recently speakers of Tibeto-Burman languages came down over the Himalayas from the north (Gurung oral tradition places this around 500 CE). For thousands of years these groups (and the languages they spoke) interacted with each other, formed autonomous kingdoms, expanded the kingdoms, conquered other kingdoms, and moved around the area (Whelpton 2005).

In the mid-18th century Prithvi Narayan Shah, ruler of the kingdom of Gorkha, was particularly successful at expanding his kingdom. Motivated by the desire to unify the many autonomous groups and kingdoms, by 1770 he had conquered the powerful Malla kingdom of the Kathmandu valley (the area that at that point was known as "Nepal") and ruled most of the eastern half of present-day Nepal as well as what is currently the Indian state of Sikkim. Prithvi Narayan Shah spoke an Indo-Aryan language known as either Gorkhali or Khasa Kura, and this language became the language of government and the ruling class of his new kingdom. Before being conquered by Prithvi Narayan Shah, the Kathmandu valley had been ruled by the indigenous Newar caste/ethnic group. Their Tibeto-Burman language, known as Nepal Bhasa ("Nepal language"), had long been the most prestigious and dominant language of the area, though Maharjan (2018) writes that under Newari rule people were encouraged to speak their own languages.[5]

From 1775-1950, the newly unified kingdom was ruled first by descendants of Prithvi Narayan Shah and then by the Rana dynasty. During this time Gorkhali remained privileged as the language of government (now centred in Kathmandu), and began impinging on the domains of local indigenous languages. In the 1930s, in the name of common Nepalese identity, the name "Nepal" was adopted for the entire kingdom and the name "Nepali" was adopted for its governing language previously known as Gorkhali. Meanwhile, Nepal Bhasa, the language of many in the Kathmandu valley, was overtly banned during parts of the Rana rule and was increasingly referred to as "Newari" rather than Nepal Bhasa. The Rana dynasty was supported by the British, who were starting to lose hold in India, and the ban on Nepal Bhasa may have been a move to suppress potential threat to Rana power (Maharjan 2018). The ban forced the headquarters of Nepal Bhasa to move to India, whose climate was more welcoming to minority languages.

In 1951, power moved from the Ranas back to the Shah dynasty. For a brief time this meant more freedom for languages besides Nepali, as well as the rise of political parties that functioned freely alongside the monarchy (which still retained actual control of the government) (Whelpton 2005; 86). But in 1960, just a year after the first parliamentary elections, King Mahendra took back full control of the government and instituted the Panchayat system of government founded on the tenets of one country (Nepal), one ruler (the king), one religion (Hinduism), and one language (Nepali). In 1959, the first constitution of Nepal declared Nepali the national language and gave no status to any other language.

The next major change in official policy came in 1990 when King Birendra re-legalized political parties and drafted a new constitution. This constitution defined Nepal as "a multiethnic, multilingual, democratic, independent, indivisible, sovereign, Hindu and Constitutional Monarchical Kingdom" and included as fundamental the right of every community within Nepal to preserve its own language, script, and culture, and to operate (primary) schools in its own mother tongue (Nepal Constitution 1990: 20, 24).

In 1996, Nepal's Maoist party launched a "People's War" against the monarchy, beginning a 10-year long civil war. The movement was ideologically motivated by, in their words, the mission to 'empower the disadvantaged' (Sharma 2014: 20). Their demands included access to farmable land, a voice in government and, most relevant, recognition of ethnolinguistic diversity and minority rights. While actual progress on changing policies and instituting systems were stalled due to the instability caused by the civil war, it did lead to "more demands by and awareness of indigenous and ethnic peoples' rights for education, state politics, and employment based on their historically marginalized ethnolinguistic identities." (Sharma 2014: 21).

At the end of the ten year civil war in 2007, the monarchy officially ended and Nepal transitioned to a fully parliamentary government. After several attempts, a new constitution was ratified in

[5] *nepal* and *newar* likely come from a common form.

2015. Among other things, this constitution classifies all languages spoken in Nepal as "languages of the nation," while still designating Nepali (written in Devanagari script) as the only official language of the country. Like the 1990 constitution, the newest version guarantees the right to mother tongue education for all communities. In response to minority and non-dominant groups' demands for greater autonomy, self-determination, and more proportional representation in government, the 2015 constitution also divides Nepal into 7 federal provinces, partially along ethnic/linguistic lines.

Current linguistic situation of Nepal

SIL lists 109 languages currently spoken in Nepal by its population of approximately 30.2 million. Of these 109 languages, SIL classifies about half of them as "threatened" to "dormant" on their endangerment scale. Nepali is the most widely-spoken language of Nepal and its lingua franca, though not everyone speaks it. The 2011 census reported that at that time 46% of the population spoke Nepali as their L1, and 36% more spoke it as an additional language. This leaves 18% of the population who does not report speaking it. In addition to Nepal, Nepali is also spoken in Northern India (primarily in the states of Sikkim, West Bengal, and parts of Uttar Pradesh and Assam), Bhutan, and the diaspora.

Nepali is a Indo-Aryan language, as are about 30 of the other languages spoken in Nepal. Nearly all of the remaining 90 languages are Tibeto-Burman[6]. The Tibeto-Burman vs. Indo-Aryan divide is often cited as a major characteristic of Nepal. This split often gets linked to religion, with the often-perpetuated idea that Tibeto-Burman speakers are Buddhist, while Indo-Aryan speakers are Hindu. In reality, there is a huge amount of syncretism in the religious practices of Nepalis, and this popular classification does not tell the full story. In terms of geography, Tibeto-Burman language/ancestry is associated with the mountains along the northern border with Tibet, and Indo-Aryan with the flatlands along the Southern border with India as well as the hills in the middle of the country. In most of these domains, the qualities tied to Indo-Aryan languages are privileged – Nepali is the official language, most government positions are held by people of Indic descent, the national language is Hinduism, and the middle hills area (home to Prithvi Narayan Shah and many Indo-Aryan speaking groups) gets the most attention for development projects.

Nepali consonant phonemes

The focus of this dissertation is Nepali's four-way laryngeal contrast between voiceless, voiced, voiceless aspirated, and voiced aspirated stops. This contrast exists in both stops and affricates, though the studies in this dissertation consider only stops to enable comparison to work on laryngeal contrasts in other languages (see (Clements and Khatiwada 2007) for an acoustic study of Nepali affricates). The consonant inventory of Nepali is shown in Table 1.3.

[6] There is one isolate (Kusunda), and three Austro-Asiatic languages (Kharia, Mundari, and Santhali).

Table 1.3: Nepali phonemic consonant inventory (Khatiwada 2009)

	Bilabial	Dental	Alveolar	Retroflex	Palatal	Velar	Glottal
Plosive	p b p^h b^h	t d t^h d^h		ʈ ɖ ʈ^h ɖ^h		k g k^h g^h	
Affricate			ʦ ʥ ʦ^h ʥ^h				
Nasals	m	n				ŋ	
Tap or flap			r				
Fricative			s				ɦ
Lateral			l				
Approximant	(w)				(j)		

Throughout this dissertation I will label the [p, t, ʈ, k] series 'voiceless' or 'T', the [p^h, t^h, ʈ^h, k^h] series 'voiceless aspirated' or 'Th', the [b, d, ɖ, g] series 'voiced' or 'D', and the [b^h, d^h, ɖ^h, g^h] series 'voiced aspirated' or 'Dh'.

All four laryngeal classes of plosives and affricates contrast phonemically in all positions. The following sample of minimal and near-minimal pairs evidence these contrasts in word-initial (#__V), intervocalic (V__V), and word-final (V__#) positions, since these are the three contexts most relevant to the studies in this dissertation. The minimal pairs presented here come from Acharya (1991) and **rana2012**.

Word-initial position: (#__V)

(1) a. /parsi/ 'day after tomorrow'
  b. /p^{h}arsi/ 'pumpkin'
  c. /ba:ri/ 'a dry cultivated field'
  d. /b^{h}a:ri/ 'load'

(2) a. /ta:l/ 'lake'
  b. /t^{h}a:l/ 'plate'
  c. /da:n/ 'donation'
  d. /d^{h}a:n/ 'rice (unhusked)'

(3) a. /ʈok/ 'bite'
  b. /ʈ^{h}ok/ 'hit(imperative)'
  c. /ɖoka:/ 'bamboo basket'
  d. /ɖ^{h}oka:/ 'door'

(4) a. /ka:m/ 'work/job'
  b. /k^{h}am/ 'envelope'
  c. /gar/ 'do (imp)'
  d. /g^{h}ar/ 'house'

(5) a. /ʦuri/ 'bangle'
  b. /ʦ^{h}uri/ 'dagger'
  c. /ʥutta:/ 'shoes'
  d. /ʥ^{h}utta:/ 'bunches'

Intervocalic position: (V__V)

(6)
a. /ʈa:pi/ ‘liar’
b. /map^{h}i/ ‘amnesty’
c. /sabai/ ‘all’
d. /anub^{h}ab/ ‘experience’

(7)
a. /ta:to/ ‘hot’
b. /sat^{h}i/ ‘friend’
c. /madat/ ‘help’
d. /a:d^{h}a:/ ‘half’

(8)
a. /pa:ʈi/ ‘rest house’
b. /pa:ʈ^{h}i/ ‘young female goat’
c. /ʥa:ɖo:/ ‘cold’
d. /buɖ^{h}o/ ‘old’

(9)
a. /ka:ki/ ‘aunt’
b. /ka:k^{h}i/ ‘armpit’
c. /bagar/ ‘river bed’
d. /lag^{h}a:r/ ‘chase (imp.)’

(10)
a. /sã:ʦo/ ‘true, key’
b. /sa:ʦ^{h}i/ ‘witness’
c. /a:ʥa/ ‘today’
d. /bã:ʥ^{h}o:/ ‘barren’

Word-final position: (V__#)

(11)
a. /ã:p/ ‘mango’
b. /sa:^{h}/ ‘clean’
c. /sab/ ‘all’
d. /lob^h/ ‘greed’

(12)
a. /pa:t/ ‘leaf’
b. /sat^h/ ‘company’
c. /khed/ ‘chase (imp.)’
d. /bã:d^h:/ ‘barrage’

(13)
a. /ka:ʈ/ ‘cut (imp)’
b. /a:ʈ^h/ ‘eight’
c. /ga:ɖ/ ‘bury (imp)’
d. /bã:ɖ^h/ ‘flood’

(14)
a. /ʦa:k/ ‘buttock’
b. /ʦa:k^{h}i/ ‘taste’
c. /la:g/ ‘stick’
d. /ba:g^h/ ‘tiger’

(15)
a. /sã:ʦ/ ‘save(imp)’
b. /kaʦ^h/ ‘groin’
c. /ka:ʥ/ ‘deputation’
d. /ba:ʥ^h/ ‘quarrel’

Variability in realization of stop and affricate phonemes

While a prototypical or idealized realization of the stop and affricate phonemes is expressed by the transcriptions above, there is variability in how they are produced. First, some of the aspirated phonemes are realized by some speakers and in some positions as fricatives (Khatiwada 2009; Nakkeerar 2011). The most commonly spirantized phoneme is p^h, realized as ϕ or f. While no other consonants spirantize nearly as regularly as /p^h, the other bilabial and velar aspirated stops can also be realized as homorganic fricatives, as the below examples from Khatiwada (2009, p. 376) illustrate. Spirantization is sometimes reported to happen most reliably in syllable-final position, and more often in spontaneous speech. The three production studies that I will present in this dissertation all involved read speech, and even so speakers across all studies spirantized /p^h/ nearly all the time in every position. Rates of /b^h/ spirantization were drastically lower and primarily in word-final position, and velars were spirantized even more rarely.

(16) a. /sʌp^{h}a/ [sʌɸa] ‘clean’
b. /sʌb^{h}a/ [sʌβa] ‘meeting’
c. /sak^{h}a/ [saxa] ‘branch’
d. /ʌg^{h}aːdi/ [ʌɣaːdi] ‘before’

More sociolingustic work is needed to understand if there are demographic or regional factors that contribute to the rates of spirantization. It may be notable that Acharya’s 1991 grammar does not mention spirantization, even of /p^h/, though it does mention other variants of the consonant phonemes.

In non-initial positions (and particularly syllable-final position), aspiration is also variably lost on the voiced aspirated phonemes, weakening the contrast between the voiced unaspirated and voiced aspirated stops in this position. This phenomenon is the basis for Chapter 4, and will be discussed there in much more depth.

The retroflex stops are reportedly produced with less retroflexion than in other South Asian languages (Pokharel 1989; Acharya 1991; Khatiwada 2009). A palatographic study Khatiwada (2007) shows that their productions may range between retroflex and apical, depending on the speaker and the preceding vowel. Intervocalically, I found that retroflex stops were sometimes realized more like approximants, without a full closure.

There is also variation reported in the realization of the affricates. Clements and Khatiwada (2007), in their acoustic study of Nepali’s affricates, find that aspiration is variably lost intervocalically on the aspirated classes. Khatiwada (2009) describes that while palatograms (Pokharel 1989) and palatograms and linguagrams (Clements and Khatiwada 2007) show that the affricates are produced with alveolar contact, auditorily, they are often perceived as palato-alveolar sounds, particularly to English and French speakers. In this dissertation I do not report on affricates, but I did find the voiced unaspirated affricate to be variably realized as [z], and all four of the affricates to range from alveolar to alveopalatal productions.

Phonetic cues previously found to distinguish four-way contrast in Indic stops

The contrast between the different stop classes in Indic languages with a four-way contrast like Nepali’s is said to be achieved by both durational cues (e.g. voicing duration, closure duration, or aspiration duration) and f0 and spectral cues of the following vowel (e.g. breathy voice quality measures) (Davis 1994; Clements and Khatiwada 2007; I. Dutta 2007; Mikuteit and Reetz 2007; Berkson 2012). Since laryngeal realism bases its feature diagnostic criteria on durational cues, this study will focus on durational cues as well. This section summarizes various durational cues that have been used previously and how to measure them, providing the basis for the annotation and measurement criteria used in this study. We draw from previous work on Nepali (Poon and Mateer 1985; Clements and Khatiwada 2007) as well as on related languages including Hindi (Lisker and

Abramson 1964; Davis 1994; I. Dutta 2007), Marathi (Berkson 2013), and Bengali (Mikuteit and Reetz 2007).

Voice Onset Time

The classic durational measure used to distinguish stop classes from each other is voice onset time (VOT) (Lisker and Abramson 1964). It is defined as the time difference between the beginning of the release and the onset of voicing of the following vowel. For word-initial segments in which voicing begins during the stop closure, the duration of voicing before the release of the stop is measured as negative VOT. We refer to this as ‘prevoicing duration’. In segments without prevoicing, VOT is the positive measure of how long it takes for voicing to begin after the release of the stop, or ‘lag time’. VOT values distinguish stop classes in a two-way contrast language like either French or German, and a three-way contrast language like Thai. The four stop classes of an Indic language, however, cannot be distinguished from each other by VOT alone; that is, they do not have four discrete VOT ranges along a continuous VOT scale. Rather, the VOT of Dh in Hindi, Marathi, and Nepali overlaps with that of all the other stop classes, both negative and positive (Lisker and Abramson 1964; Poon and Mateer 1985), as schematized in Table 1.4.

Table 1.4: Schematized VOT durations for stop classes in 2, 3, 4-way contrasting languages

	Negative VOT	Short positive VOT	long positive VOT
French	b	p	
German		p	p^h
Thai	b	p	p^h
Hindi	b	p	p^h
	b^h		

Moreover, because VOT is negative if any prevoicing exists, a continuous VOT scale can only capture *either* voicing *or* aspiration on any given segment. This does not pose a problem for French, German, or Thai because in languages with a two- or three-way contrast these two cues rarely coexist on the same segment, but it masks the fact that Indic voiced aspirated stops often have both. Thus, in order to adequately capture both voicing lead and voicing lag on the same segment we need to divide VOT into two distinct measures—lead time (voicing duration) and lag time (post-release duration)—and consider these cues separately.

Voicing duration

‘Voicing duration’ is defined as vocal cord vibration during stop closure. In initial position we call this ‘prevoicing’, and it corresponds to negative VOT. Prevoicing is measured from the beginning of voicing to the release of closure (Figure 1.1). Predictably, this measure has been found to distinguish voiced classes from voiceless classes in Indic languages. The prevoicing duration of D has been found to be slightly longer than the prevoicing duration of Dh in Hindi (Lisker and Abramson 1964; Davis 1994; I. Dutta 2007), but not different enough to distinguish the two voiced stop classes from each other based on this cue alone. Voicing duration is measured differently in word-medial (intervocalic) stops, as the percentage of the closure duration that has voicing (Iverson and Salmons 1995; Beckman, Jessen et al. 2013), as will be discussed further in Section 2.4.

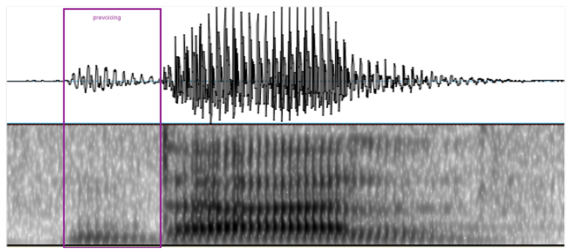


Figure 1.1: Example of prevoicing/negative VOT, for the Nepali word [dal] ‘lentils’.

Post-release Duration

Post-release duration is the combined burst and aspiration durations, from the release of closure to the beginning of the following vowel, and provides a way to distinguish the aspirated classes (both voiced and voiceless) from the unaspirated classes. In a voiceless aspirated stop, post-release duration begins at the release of the burst and ends at the onset of voicing, marked by periodicity in the waveform and voicing bar in the spectrogram (Figure 1.2).

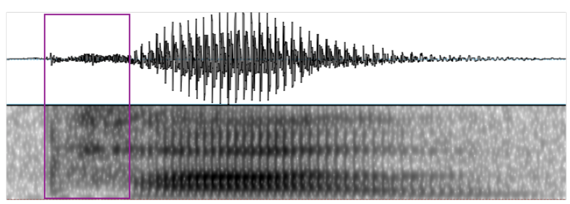


Figure 1.2: Example of post-release/positive VOT, for the Nepali word [t^hal] ‘plate’.

In voiced aspirates there is often no clear point of voicing onset after the burst since voicing may continue directly from the prevoicing during the closure through the burst, and into the voicing of the following vowel (Berkson 2012). This makes the end point of post-release duration more challenging to measure, and various studies have proposed slightly different guidelines, illustrated in Figure 1.3.

Davis (1994) uses Noise Offset Time (NOT) for Hindi, measured from the beginning of the release burst to the onset of F2 in the following vowel. Later studies found NOT difficult to replicate because F2 onset is often unclear, and propose revised measures (Mikuteit and Reetz 2007; Berkson 2012).

Mikuteit and Reetz (2007) divide the post-release interval into two distinct types. After Closure Time (ACT) is the aperiodic stretch from burst release to the first glottal pulse, and Superimposed Aspiration (SA) extends from the first glottal pulse to the end of high frequency frication noise, visible on the waveform as jaggedness on the vowel. This may correspond to breathy phonation of the vowel, but is measured as a duration and not as a spectral value. They find that the post-release duration of Th is generally ACT, while Dh is either entirely SA or a period of ACT followed by SA. Clements and Khatiwada (2007) use the same measurement scheme (ACT & SA) for Nepali affricates and find similar results.

The difference between ACT and SA can be very difficult to distinguish visually, so Berkson (2012) proposes a third measure: the Pre-Vocalic Interval (PVI). PVI is a single measurement similar to the concatenation of ACT and SA. It begins at the release of the burst, and ends at the end of breathiness on the vowel, marked by the clear onset of a dark F2 in the spectrogram, or an increase in amplitude in the waveform. PVI is essentially the Indic equivalent of positive VOT, generalized to apply to the voiced aspirated series as well as to the other three stop series. When applied to Marathi data, PVI was found to make a three-way distinction between T/D, Th, and Dh. The studies in this dissertation follow Berkson and use the PVI measure.

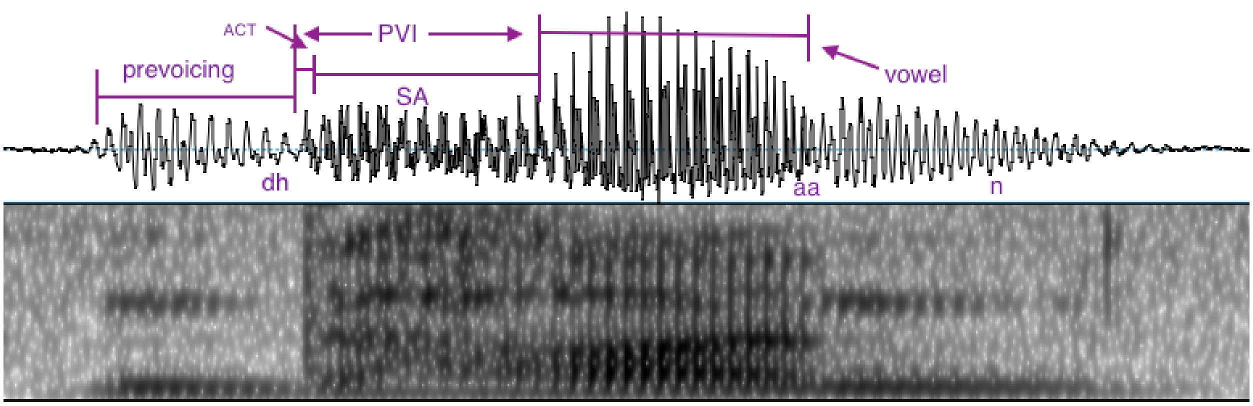


Figure 1.3: Example of post-release on a voiced aspirated stop, with various annotation schemes marked, for the Nepali word [d^{h}an] 'rice paddy'.

1.3 Central questions and aims

Each chapter of this dissertation reconsiders one of the core phenomenon related to laryngeal contrasts presented above, in light of data from Nepali. Chapter 2 deals with realization and representation of Nepali's laryngeal contrasts, asking: How is the four-way contrast realized in Nepali in initial and intervocalic positions, in terms of durational acoustic cues? How do proposals to diagnose feature representation based on phonetic properties hold up when applied to a language like Nepali, with a four-way contrast and voiced aspirated stops? Chapter 3 examines the effect of laryngeal specification on vowel length through a production study of post-vocalic stops: How does a laryngeal classification besides voicing affect the duration of preceding vowels, and how can this clarify the source of such effects? Chapter 4 presents the results of production and perception studies investigating contrast weakening in word-final position: Does Nepali show evidence of neutralization of any laryngeal contrasts in word-final position? How does this reflect on the representation of Nepali's stops as well as cross-linguistic typology of laryngeal neutralization?

Chapter 2

Laryngeal Realism: Realization and representation

2.1 Introduction

Chapter 1 presented two different approaches to representing laryngeal contrasts. In 'abstract' approaches, all two-way contrasts are realized with a [voice] feature, regardless of the language-specific phonetic cues used to realize the contrasts (Chomsky and Halle 1968; P. A. Keating 1984; Rubach 1990; Lombardi 1999). In 'realist' approaches, phonological representations more closely match the phonetic realization. Languages like French that rely on prevoicing to cue laryngeal contrast use a [voice] feature, while languages like English that contrast short-lag VOT with long-lag VOT use a [spread] feature. In a language like Nepali with a four-way laryngeal contrast, an abstract approach is less feasible. A single binary or privative feature cannot be used to capture four different laryngeal specifications. Rather, it is proposed to use minimally a [voice] and [spread] feature, fully crossed to capture all four classes. That Nepali uses the same two features that are central to the laryngeal realism literature makes it a particularly relevant language to study: we can apply the same diagnostics used for languages with fewer contrasts, and reflect back on the theory using data that are more complex than those which have been previously considered in this literature. Based on our analysis of acoustic data, our findings largely support the postulation of [voice] and [spread] for Nepali, but with caveats that there is an asymmetry between the two features best captured with subsegmental specification[1].

Types of evidence used by laryngeal realism

Laryngeal realist theories use three main lines of evidence to motivate the feature representation of a language's laryngeal contrast: phonetic realization of the segments, diagnostics of 'control',

[1]This chapter was originally published as Schwarz, Sonderegger et al. (2019) with co-authors Morgan Sonderegger and Heather Goad, though I've made some changes and some additions. I use 'we' in parts that are original to the co-authored paper, but I use 'I' for the parts that I have added to it for this dissertation, and which I do not want to presume that they endorse.

and phonological behavior. Phonetic realization, which has already been discussed, is the notion that the phonetic cues that distinguish stop classes in initial position should directly correlate to their feature representation. If presence of vocal fold vibration distinguishes the stops, [voice] is specified. If duration of the burst distinguishes them, [spread] is specified.

The second line of evidence hinges on the connection between feature representation and the status of phonetic cues as either intentionally controlled by the speaker, or a physiologically inevitable result of the articulation of a segment. Beckman, Helgason et al. (2011) and Beckman, Jessen et al. (2013) implement this connection in laryngeal-realist phonological theory as a pair of feature diagnostics, but the idea is very similar to the notions of 'automatic' and 'controlled' (a.k.a. 'mechanical') aspects of speech from the phonetic literature (reviewed by Solé 2007). Solé (2007) reports that controlled and mechanical gestures can be distinguished by their distinct behavior under several conditions, including their durations at slower speech rates. If the articulatory goal of a specified feature is available in the input to speech production, and if the segment-internal timing is specified at the same level, then the segment-internal proportional durations should remain consistent across different speech rates. In testing this, Solé finds that durations of controlled cues increase at slower speech rates (i.e. the segment-internal durations increase as the total duration of the segment increases), while durations of automatic cues remain fairly constant across speech rates, suggesting that their corresponding feature was not specified on the phonological level.

Beckman, Helgason et al. (2011) and Beckman, Jessen et al. (2013) turn similar ideas into a pair of feature specification diagnostics. Beckman, Helgason et al. (2011) review literature on the effects of speech rate on word-initial VOT durations in voicing and aspirating languages (e.g. Pind 1995; Kessinger and Blumstein 1997; Magloire and Green 1999), while providing new data on Swedish. In aspirating languages, the long-lag VOT of *p, t, k* is longer at slower speech rates, while the short-lag VOT of *b, d, g* is constant across speech rates. In voicing languages, the negative VOT of *b, d, g* is longer at slower speech rates, while the short-lag VOT of *p, t, k* remains constant. For Beckman et al. (and laryngeal realism more broadly), long-lag aspiration is the phonetic manifestation of a [spread] feature and prevoicing is the phonetic manifestation of a [voice] feature. Or, in Solé's terminology, they are phonologically specified controlled cues. The English *b, d, g* stops and French *p, t, k* stops remained constant across speech rates, behaving like a mechanical cue. Beckman et al. argue that this is because these classes are not phonologically specified for any laryngeal features. If the asymmetric speech rate results show that one stop class acts as if it is specified for a feature and the other class acts as if it is not, this supports the privative, realist feature representation in Table 1.1. Duration across speech rates thus becomes our first diagnostic of control: if a (durational) cue is enhanced at slower speech rates for a given class of sounds, it supports that class being specified for the feature corresponding to that cue.

Beckman, Jessen et al. (2013) propose that the amount of voicing in an intervocalic stop closure provides a second diagnostic of control, and by extension feature representation. They examine what proportion of the closures of *b, d, g* stops in voicing vs. aspirating languages are voiced. In Russian, a voicing language, 97% of *b, d, g* stops have fully voiced closures. In German, an aspirating language, 62% of *b, d, g* stops have fully voiced closures. They propose that voicing continues all the way through the stop closure of *b, d, g* in Russian because voicing is active and controlled by speakers, suggesting a [voice] feature. The inconsistent voicing in German is a passive or automatic unintended consequence of voicing in the preceding vowel, suggesting that German *b, d, g* stops are not specified for a [voice] feature, nor for any laryngeal feature (Beckman, Jessen et al. 2013, p. 259; Jessen and Ringen 2002; Jansen 2004). Meanwhile, Beckman, Jessen et al. (2013) argue that the [spread]-specified *p, t, k* stops in German and other aspirating languages block voicing intervocalically because the glottis is too wide, realized as voicing into only 20-30% of the closure (see Möbius (2004) and Pape and Jesus (2014) for similar results).

The third line of evidence looks to phonological patterning, arguing that asymmetrical phonological processes as well as principles of markedness and parsimony support realist representations. The output of neutralization processes, for example, tend to be the segment that in a realist representation lacks feature specification (e.g. Lombardi 1991; Iverson and Salmons 2011). Further, it always seems to be feature-specified segments in a realist representation that are active in assimilation: [spread]-languages show assimilation to [spread]-specified *p, t, k* obstruents, while [voice]-languages show assimilation to [voice]-specified *b, d, g* obstruents.

This chapter will focus on phonetic realization and diagnostics of control, while Chapter 4 focuses on phonoligical processes and markedness in its discussion of contrast weakening in final position.

Challenges in extending the theory: motivation for the current study

The predictions of phonetic realization, diagnostics of control, and phonological patterning hold up as expected—using privative, realist representations—in languages with two-way contrasts that are argued to employ either [voice] or [spread]. They have also been extended without issue to a language like Thai, which uses both laryngeal features in its three-way contrast between [b], [p], and [p^h], represented as [voice], [ ], and [spread] respectively (Pind 1995; Kessinger and Blumstein 1997; Beckman, Helgason et al. 2011), as shown in Table 1.2. These tests have also been used to argue that Swedish has a two-way contrast between a [voice]-specified class and a [spread]-specified class, despite the lack of economy (Beckman, Helgason et al. 2011). Beckman, Helgason et al. (2011) find that the negative VOT of Swedish *b, d, g* decreases (becomes more negative) as speech rate slows, as predicted of a class specified for a [voice] feature. At the same time, the long-lag VOT of Swedish *p, t, k* increases as speech rate slows, as predicted of a class specified for a [spread] feature, leading Beckman, Helgason et al. (2011) to propose overspecification.[2]

However, the same predictions pose potential conflicts (elaborated on in research question 2 below) for a language that includes segments specified for multiple laryngeal cues on the *same* segment, i.e. voicing *and* aspiration—a language such as Nepali, which contrasts four stop series traditionally described as voiceless, voiced, voiceless aspirated, and voiced aspirated.

This four-way contrast is typical of Indic languages, for which Iverson and Salmons (1995) propose the feature representation in Table 2.1. This representation uses the same [voice] and [spread] features as English, French, and Thai, but exploits every logical combination of them.

[2]See also Ramsammy and Strycharczuk (2016) for evidence of another hybrid (though not necessarily overspecified) laryngeal contrast system, in which phonetic realization suggests that European Portuguese stops are distinguished by a [voice] contrast while fricatives are distinguished by [spread].

Table 2.1: Feature representation of an Indic style four-way contrast, as proposed by Iverson and Salmons (1995)

[p, t, k]:	[p^h, t^h, k^h]:	[b, d, g]:	[b^h, d^h, g^h]:
[ ]	[spread]	[voice]	[spread], [voice]

This study analyzes Nepali's four-way stop contrast with the same types of evidence used to motivate the laryngeal realist representations for two- and three-way contrast languages. It aims to evaluate Iverson and Salmon's (1995) proposed feature representation, and laryngeal realism more generally, by examining whether the phonetic realization results of control diagnostics of Nepali's stop contrast is as predicted by laryngeal realism. A summary of the behavior of each of Nepali's stop classes as predicted by the laryngeal realist diagnostics is laid out in Table 2.2.

Table 2.2: Predicted behavior of a four-way stop contrast based on the diagnostics of laryngeal realism and Iverson and Salmon's (1995) proposed representation

Class	Rep.	Realization	Duration of cues at slow speech rates	Intervocalic voicing
t	[ ]	short-lag VOT	short-lag VOT does not increase	blocked (in voicing lang.) passive (in aspirating lang.)
t^h	[spread]	long-lag VOT	long-lag VOT increases	blocked
d	[voice]	prevoicing	prevoicing duration increases	full
d^h	[voice] [spread]	prevoicing long-lag VOT	prevoicing duration increases long-lag VOT increases	full blocked

The aim of this study is to evaluate the extent to which the stops in Nepali actually conform to these predictions. We note in particular the predicted conflict in the bottom right corner of the table concerning the amount of voicing in intervocalic voiced aspirated stops – a [spread] feature predicts very little voicing and a [voice] feature predicts full voicing. Each of the research questions that frame this study speaks to the predictions in in Table 2.2.

1. **How is the four-way stop contrast realized in Nepali in initial position, in terms of acoustic cues?** Addressing this question serves two goals of the study. First, it provides basic empirical data on Nepali stops, the acoustic realization of which is underdescribed in the literature. Second, it allows us to diagnose feature representation in terms of phonetic realization in word-initial position – the position used as the starting point for determining features in previous studies. We test the hypothesis predicted by laryngeal realism, that the voiced aspirates' feature specification with both [voice] and [spread] is appropriate if the segments display both prevoicing (like the [b, d, g] segments specified for [voice]) and long-lag VOT (like the [p^h, t^h, k^h] segments specified for [spread]). By examining cues in word-initial position we find support for employing [voice]

and [spread] primitives in the representation, provided that the phonetic diagnostic for [spread] is generalized somewhat.

2. **How well do speech rate and intervocalic voicing effects support the proposed feature specification of laryngeal classes in Nepali?** Beckman et al. (2011, 2013) propose speech rate in initial position and voicing in intervocalic position as two ways to diagnose feature specification. We apply these diagnostics and test whether they support Iverson and Salmon's (1995) privative feature representation in Table 2.1. Nepali's voiced aspirated stops pose a challenge for Beckman et al.'s (2013) intervocalic voicing diagnostic. Stops specified for [voice] are supposed to be voiced through the entire stop closure. Stops specified for [spread] are supposed to block voicing during the stop closure. Our findings from the first research question suggest that Nepali's voiced aspirated stops are specified for [voice] and [spread]. We find that voiced aspirated stops pattern as expected for a [voice]-specified stop rather than a [spread]-specified stop, suggesting an asymmetry between the [voice] and [spread] features, where the [voice] feature is 'stronger' than the [spread] feature.

The remainder of this chapter is organized as follows. Section 3.3 explains the methods of data collection and analysis. Sections 2.3–2.5 present the results of the study and their implications for feature representation. Section 2.3 addresses research question 1, examining the phonetic realization of stops in initial position. Section 2.4 addresses research question 2, applying the diagnostics of control. Section 2.5 concludes.

2.2 Methods

To address our research questions on realization of laryngeal contrasts, effects of speech rate, and intervocalic voicing, we collected production data that control for place of articulation and position of the segment in the word.

Participants

17 speakers (10 male, 7 female) participated in this study. All were graduate students in the Nepali department of Sikkim University in Gangtok, Sikkim (the Nepali-speaking region of northeast India). All participants were from Sikkim and had lived in the region their entires lives, grew up speaking Nepali at home, attended school in Nepali, and used Nepali as their main language for daily interactions. The ethnic background of Nepali speakers in Sikkim is very heterogenous (Nakkeerar 2011), so many participants additionally spoke a language associated with their ethnic group. They also all spoke Hindi (nearly all Sikkim residents are), and all spoke some amount of English. Many of the local languages of Sikkim which the participants spoke are Tibeto-Burman, so the variety of Nepali analyzed in this study is one that is in close contact with Tibeto-Burman languages (which do not have voiced aspirated sounds in their inventories). While all participants reported being dominant in Nepali, there were not enough speakers from each language background to analyze the potential effect of each language on the production of Nepali individually due to the small number of data points per person, so we leave examination of this issue to future work.

Stimuli

The stimuli consist of 32 Nepali words, corresponding to each of the 16 stops in Table 1.3, in word-initial and intervocalic word-medial position. These two positions are necessary to address our first and second research questions. The stimuli were all real words and, as much as possible, were controlled for the quality of the following vowel (a preference for [a]) and stress (a preference for word-initial segments to be onsets of stressed syllables and word-medial segments to be onsets of unstressed syllables) in order to optimize contrast in initial position and reduce it in intervocalic position.[3] Each word was written in Nepali's syllabic orthography (Devanagari script) on a separate cue card. The participants were shown each target word in an order randomized for place of articulation and position in word. They produced the word in the carrier sentence in (1), which is nearly identical to the carrier sentence used by Clements and Khatiwada (2007). We acknowledge that by recording read speech as opposed to spontaneous speech, it is possible that participants were influenced by the spelling and thus pronounce the sounds more 'correctly' than they would have in natural speech. Using the baseline of read-speech pronunciation established in the current study, future work could examine Nepali laryngeal contrasts in spontaneous speech.

(1) X_1 (pause). mʌ ʌbʌ X_2 b^hants^{h}u: (pause) X_3
'X_1 (pause) now I say X_2: (pause) X_3.'

The results presented here are from the X_1 and X_3 positions of all 17 speakers because both of these positions are preceded and followed by a pause, giving roughly comparable prosodic contexts. This yielded a total of 559 tokens with the segment in initial position.[4] After excluding 15 word-medial stops realized as approximants, we analyzed 418 tokens with the segment in intervocalic medial position, for a combined total of 977 tokens across both positions.

Acoustic analysis

The recordings were imported into Praat (Boersma and Weenink 2015), where they were hand-annotated for voicing, closure, and post-release duration measurements, as pictured in Figure 2.1. For word-initial segments, voicing duration was measured from the beginning of voicing (marked by the onset of periodicity in the waveform and voicing bar in the spectrogram) to the release of the stop (marked by a clear increase in amplitude and (often) the beginning of aperiodic noise in the waveform). For word-medial segments, voicing duration was measured from the beginning of

[3] We expect prosody to affect stop realization, but the factors governing word-level stress/prominence in Nepali are not very clear (Acharya 1991; Clements and Khatiwada 2007). Moreover, we found that stress was highly subject to sentence prosody, which placed strong prominence on the ultimate syllable of the sentence. Thus, the disyllabic words which had prominence on the first syllable in position X_1 (see (1) below in text), often had prominence on the final syllable in position X_3. One might expect that the medial stops would therefore have different profiles depending on whether they were in X_1 or X_3 position, but the intervocalic voicing effects to be reported in Section 5 do not seem to differ based on position in the carrier phrase.

[4] 17 speakers x 16 words x 2 tokens each = 544. There were a few additional words with word-initial stops, which are responsible for the 15 extra initial stop tokens.

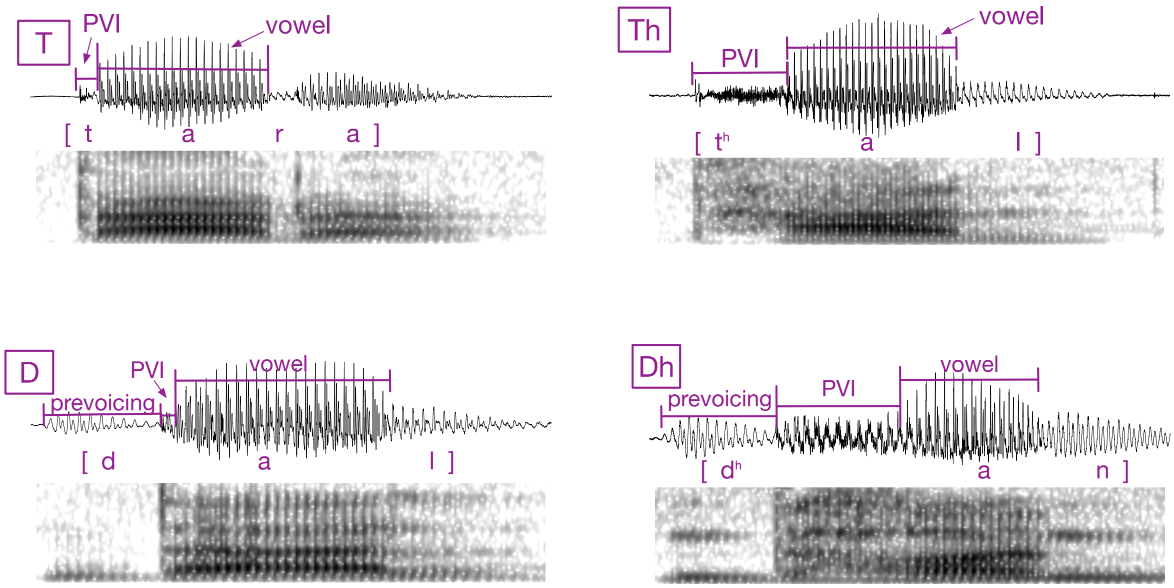


Figure 2.1: Examples of annotation. Top left: [tara] ‘star’. No prevoicing, end of stop closure marked by a burst with very short interval of aperiodic aspiration before voicing begins. Top right: [t^hal] ‘plate’. No prevoicing, ACT is much longer than for [tara], but still marked by aperiodic high frequency aspiration noise. Bottom left: [dal] ‘lentils’. Prevoicing interval is marked by periodicity in the waveform and voicing bar in the spectrogram. Bottom right: [d^han] ‘rice (paddy)’. Cues to prevoicing identical to [dal]. PVI begins at burst and ends at jump in amplitude and smoother waves; PVI is visibly different from [t^hal].

closure (marked by a sharp decrease in amplitude in the waveform signalling the end of the preceding vowel) to the end of visible periodicity in the waveform. Closure duration (annotated for word-medial segments only) ended at the release burst if present, or at the sharp increase of amplitude in the waveform signifying the beginning of the following vowel. Post-release duration was measured using Berkson’s (2012) PVI guidelines, from burst release to onset of a dark F2/increase in amplitude.

2.3 Realization of stops in word-initial position

This section addresses the first of our research questions: how is the four-way stop contrast realized in Nepali, in terms of acoustic cues? We first present our empirical data on the acoustic realization of stops in word-initial position, to add to the literature on the phonetics of Nepali stops, then discuss the implications for feature representation. Statistical models reported below (Section 2.4) confirm the (non-)significance of patterns discussed in the empirical data here.

Results: Acoustic data

Recall that previous studies found that unlike two- and three-way laryngeal contrasts, four-way contrasts cannot be distinguished along a single VOT axis (Lisker and Abramson 1964; Poon and Mateer 1985). In Figure 3.1 we see that the new Nepali data analyzed here shows the same result: the VOT durations of T, Th, and D have three non-overlapping distributions, while the VOT of Dh overlaps with that of the other classes. Considering prevoicing and PVI duration cues separately, however, we can capture a four-way distinction as four (nearly) distinct distributions. Table 2.3 summarizes the means and standards of deviation of these measures for each class.

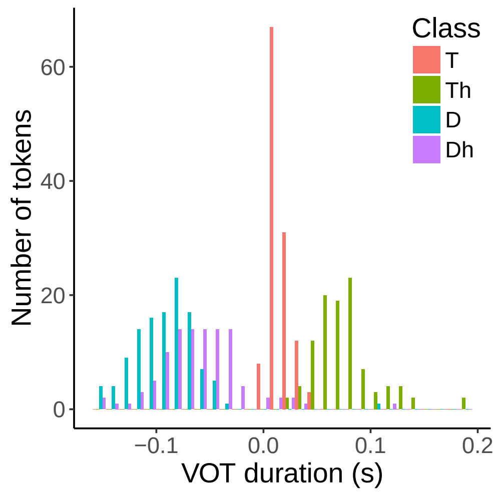


Figure 2.2: Distribution of VOT durations in initial position, for each stop class.

Table 2.3: Summary of prevoicing duration (VD) and post-vocalic interval duration (PVI) for each stop class in initial position (in milliseconds).

		VD		PVI	
Class	n	mean	sd	mean	sd
Voiceless (T)	120	0	0	16	10
Voiceless asp. (Th)	102	0	0	82	28
Voiced (D)	118	93	27	12	11
Voiced asp. (Dh)	104	63	33	56	42

Figure 3.2 (left) shows that (as expected) prevoicing duration yields a two way distinction between the voiced and voiceless classes. The voiceless classes never exhibit prevoicing while the

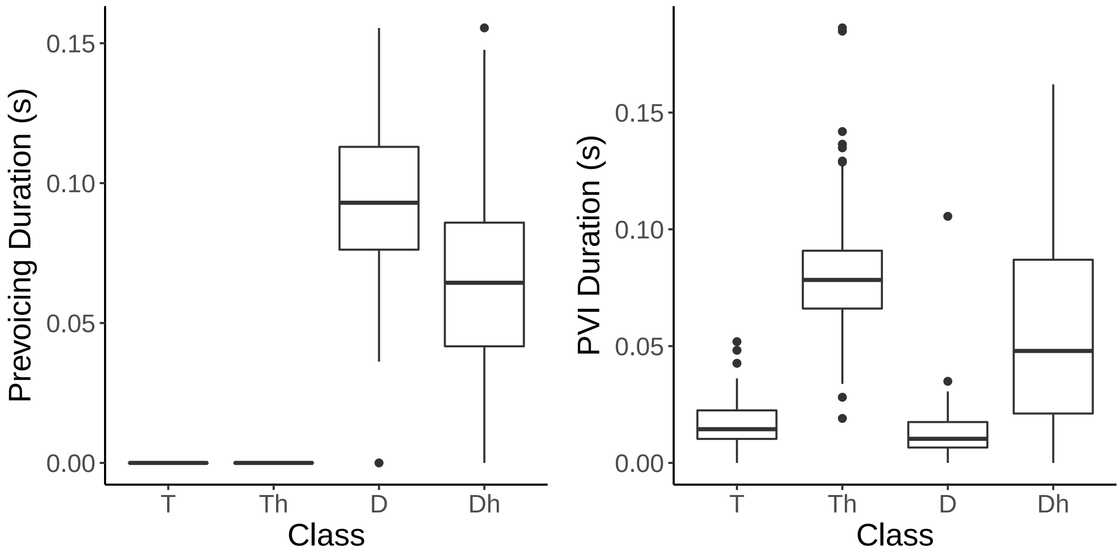


Figure 2.3: Prevoicing duration (left) and PVI duration (right) in initial position, across stop classes.

D and Dh classes have voicing duration means of 93ms and 63ms respectively. The shorter mean duration of Dh is consistent with what Dutta (2007) found for Hindi voiced aspirates.

Figure 3.2 (right) shows that PVI duration, the post-release-pre-vocalic interval used here as equivalent to positive VOT, can achieve a two-way contrast between the aspirated and unaspirated classes. The mean PVI of Dh segments is shorter than that of Th segments and the range is much larger, but the difference in PVI duration between the aspirated and unaspirated segments proved significant in a linear mixed effects model reported in Section 2.4.

Thus, voicing duration and PVI, corresponding to different parts of the traditional VOT measure, each differentiate two of the stop classes from the other two. Figure 2.4 shows that together, they separate the four classes fairly well. There is still overlap, especially between the D and Dh classes. It may be that the voiced aspirated stops with short burst durations are not perceptibly distinct from the plain voiced stops that surround them in the figure, in which case their overlapping cues show neutralization. It could alternatively be the case that they are perceptibly distinguishable, and that the difference comes from cues besides those which are considered here. Either way, the result is closer to a four-way distinction than was accomplished by a single VOT dimension alone in Figure 3.1.

Discussion: Implications for feature representation

One of the key principles of laryngeal realist representations is that the cues that distinguish stop classes from each other in word-initial position should correlate with the features that distinguish them in the representation. A [voice] feature is appropriate for segments that are consistently real-

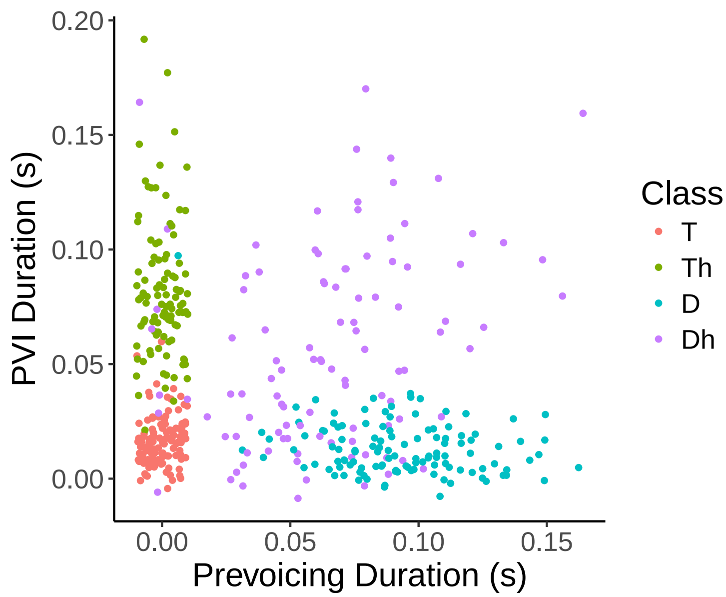


Figure 2.4: Prevoicing duration versus PVI duration for stops in initial position, for each stop class. Points are jittered for visibility.

ized with phonetic prevoicing (Honeybone 2005). Nepali's D and Dh classes are both consistently realized with prevoicing, and should therefore be specified with a [voice] feature.

Representation of a contrast with a [spread] feature depends on the length of the positive VOT duration (Honeybone 2005). When segments with long-lag VOT contrast with segments with short-lag VOT, the long-lag segments are specified with [spread]. The generally-accepted threshold for long-lag VOT is a duration longer than 30ms (Lisker and Abramson 1964). Since the PVI duration (the Indic correlate of positive VOT) of Th and Dh are each significantly longer than the PVI durations of T and D (confirmed in the statistical model below), and both have medians longer than 30ms (Figure 3.2), both Th and Dh are justified in being specified with a [spread] feature while T and D are not.

There has, however, been debate about whether both Th and Dh should be specified with the same [spread] feature, or whether Dh should instead be specified with its own feature, e.g. a [breathy] feature, given that the phonetic realization of the PVI of a Dh stop is often acoustically different from the realization of PVI in Th stops (e.g., Islam 2019). The difference can be seen in Figure 2.1, comparing [t^hal] to [d^han]. While the Th stop's PVI is aperiodic and clearly shows no voicing until the onset of the following vowel, the Dh stop's PVI has some periodicity throughout. In the annotation scheme of Mikuteit and Reetz (2007) these are classified as two different cues—ACT and SA, respectively. We offer three phonetically-based arguments that the [spread] feature is still appropriate for both.

First, one could argue that lacking a perfect one-to-one correlation between cues and features is precedented, and in fact never claimed as necessary by laryngeal realism. In English, for example, 'voicing' is signalled by long-lag vs. short-lag VOT word initially, but is primarily cued by vowel duration word finally. If the same feature is signalled by different cues in different positions, parallel logic could maintain that the same feature is signalled by slightly different cues in different stop classes.

A second argument maintains the cue-to-feature correspondence by arguing that the PVI of Th and Dh are not actually distinct cues. Berkson's (2012) PVI annotation intentionally ignores the difference between ACT and SA, considering them to be a single cue since in practice, they are too hard to distinguish reliably. Ridouane et al. (2010) offers other support, arguing that [spread] refers to a combination of articulatory gestures and acoustic measures that encompass the gestures and measures of PVI on both Th and Dh segments.

Asikin-Garmager (2017) offers additional arguments that Dh is specified for [spread], not [breathy voice]. He points to the variability in the realization of Dh segments, finding that in Hindi, the PVI is sometimes periodic, sometimes resembles the aperiodic PVI of Th segments, and sometimes includes both aperiodic and periodic intervals. We found the same variability in Nepali. Were these stops specified for a [breathy] feature, he argues, the target articulation would be periodic PVI, and it would be more reliably articulated thus. The variable realization is a product of the phonetic overlap of the [voice] and [spread] target articulations.

We have thus established the distinguishing phonetic cues of each stop class: Voiceless stops have short prevoicing duration and short PVI; voiceless aspirated stops have a short prevoicing duration but long PVI; voiced stops have a short PVI but long prevoicing duration; voiced aspirated sounds are the most variable, but may have both long prevoicing duration and long PVI. In Table 2.4, we see that by directly correlating these cues with features and proposing that prevoicing corresponds to a [voice] feature and PVI longer than 30ms corresponds to a [spread] feature, we arrive at the features in Iverson and Salmons's (1995) representation.

Table 2.4: Feature representations and their corresponding phonetic cue values.

	voiceless	voiceless aspirated	voiced	voiced aspirated
Representation	[ ]	[spread]	[voice]	[spread], [voice]
Voicing duration	short	short	long	long
Burst duration	short	long	short	long

We take this as a starting point for the representations ultimately proposed in this dissertation. Each of the laryngeal classes is specified with the features in the above table: T stops have no laryngeal specification, Th stops have a [spread] feature, D stops have a [voice] feature, and Dh stops have both [voice] and [spread] features. Based on analysis of intervocalic stops in the next sections as well as contrast weakening in Chapter 4, we will ultimately opt for a subsegmental representation, in which these features are distributed across the subsegments rather than features of the segment as a whole.

2.4 Diagnostics of control

Having established that the cues that distinguish the stops in initial position are (at least in part) a combination of voicing duration and PVI duration and that this supports a representation using both [voice] and [spread], we now turn to examine whether these features are privative or binary. We do so by addressing our second research question: What are the speech rate and intervocalic voicing effects as a function of laryngeal class in Nepali? Beckman, Helgason et al. (2011) and Beckman, Jessen et al. (2013) propose that speech rate effects and intervocalic voicing may each be used as diagnostics of laryngeal realist feature specification. We apply each diagnostic in turn, reporting the results and concluding that we find support for privative representation.

Speech rate effects in initial position

Recall that Beckman, Helgason et al.'s (2011) speech rate diagnostic proposes that phonetic cues corresponding to specified features increase in duration as speech rate slows, while cues that do not correspond to specified features remain fairly constant across speech rates. The theoretical basis of this claim is that specified features are realized physically by speakers as laryngeal gestures (Davis 1994; Beckman, Helgason et al. 2011): [voice] is manifested as prevoicing, [spread] gives rise to long-lag VOT. Specified features represent articulatory goals of the speaker, and at slower speech rates the speaker is able to achieve these goals more fully (Beckman, Helgason et al. 2011). If, for example, the short-lag VOT of English *b, d, g* and French *p, t, k* is merely an unintended mechanical consequence of transitioning from stop closure to vowel, there is no reason for it to increase at slower speech rates (Solé 2007). Supporting evidence for this view comes from studies showing that the long-lag VOT of [spread]-specified p^h, t^h, k^h stops of aspirating languages increase at slower speech rates, while the short-lag VOT of laryngeally-unspecified *b, d, g* stops does not, such as in Icelandic (Pind 1995), English (Kessinger and Blumstein 1997; Magloire and Green 1999) and Thai (Kessinger and Blumstein 1997). The prevoicing duration of [voice]-specified *b, d, g* of voicing languages (French, Thai, Spanish) also increases at slower speech rates while the short-lag VOT of *p, t, k* stops do not, such as in French (Kessinger and Blumstein 1997), Spanish (Magloire and Green 1999), and Thai (Kessinger and Blumstein 1997).

We now test the predictions of this diagnostic against Iverson and Salmons's (1995) feature representation of Nepali's four-way contrast. The diagnostic predicts that at slower speech rates prevoicing duration will increase on the [voice]-specified D stops, PVI duration will increase on the [spread]-specified Th stops, both prevoicing and PVI durations will increase on the [voice, spread]-specified Dh stops, while the PVI of unspecified T stops will not increase significantly. Figure 2.5 suggests that all of these predictions are borne out. Speech rate was calculated by dividing the duration of the carrier phrase by the number of syllables in the carrier phrase.[5] Figure

[5]Four points where speech rate was <1 syllable/second were discarded due to annotation errors. Note that speech rate was not explicitly controlled during the data collection; the participants were not instructed to speak more slowly or more quickly. The speech rate variation that emerged is based on unprompted fluctuation in speaking rate by the participants. The range of speech rates attested in this data is comparable to the rates reported in Beckman, Helgason et al. (2011), who did explicitly prompt fast speech. We use the term 'fast' to describe the faster rates in the data, but

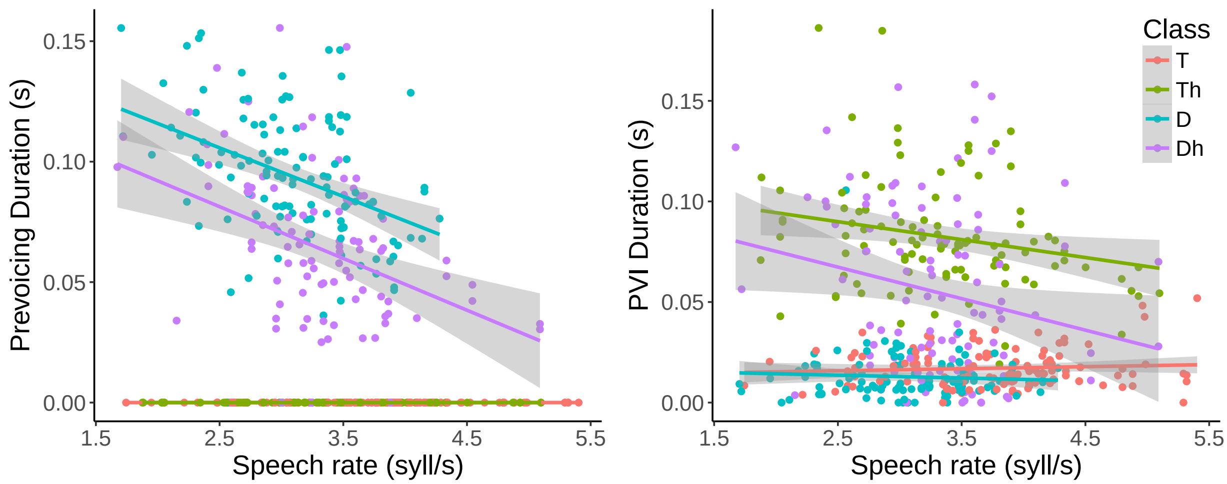


Figure 2.5: Effect of speech rate on prevoicing duration (left) and PVI duration (right) for stops in initial position.

2.5 (left) plots prevoicing durations across speech rates and shows negatively-sloped trend lines for the voiced and voiced aspirated classes, suggesting longer prevoicing durations at slower speech rates. Figure 2.5 (right) plots PVI duration across speech rates, showing that the trend lines of the aspirated classes appear to be more negatively sloped than those of the unaspirated classes.

Two linear mixed effects models (reported below in sections 2.4 and 2.4) were fitted in order to test the patterns observed in these empirical plots. The first model considers prevoicing duration, and the second considers PVI duration. These two dependent variables were log-transformed due to their highly right-skewed distributions, and to prevent the models from predicting negative durations. The models aim to determine whether speech rate affects the duration of cues that correspond to specified features significantly more than those that do not. Models were fitted using the `lme4` package in R Bates et al. (2015). Fixed-effect coefficients are shown with associated standard errors, test statistic (*t*), and significances, calculated with `lmerTest` Kuznetsova et al. (2015) using the Satterthwaite approximation. Random-effect terms are not shown. In both models, the continuous speech rate measure was standardized by centering and dividing by two standard deviations. In addition the coding system used for the predictors in each model means that the main effect coefficients can be interpreted at an average value of other predictors, for an average speaker and word.

Prevoicing duration

The only stop classes that ever display prevoicing are the voiced and voiced aspirated classes (see Figures 3.2 (left) and 2.5 (left)). The prevoicing duration model therefore checks for an effect of speech rate on both the voiced and voiced aspirated classes. To do so, the model was run on a subset

recognize that this is slower than the fast end of the continuum of spontaneous speech

Table 2.5: Linear mixed-effect model of log-transformed prevoicing duration (sec.) for D and Dh stops which showed prevoicing ($n = 212$).

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-2.666	0.0474	-56.276	< 0.001
Class (D)	0.172	0.032	5.334	< 0.001
Speech rate	-0.388	0.084	-4.619	< 0.001
POA (velar vs. bilabial)	-0.096	0.028	-3.430	< 0.01
POA (alveolar vs. bilabial/velar)	-0.037	0.016	2.303	0.031
POA (retroflex vs. bilabial/velar/alveolar)	0.017	0.013	1.343	0.187
Class:Speech rate (Dh)	0.049	0.075	0.653	0.530

of the data that included only the voiced classes, after excluding observations without prevoicing ($n = 9$: 3.6% of total). The model includes as predictors class, speech rate, the interaction of class and speech rate, and place of articulation (POA). The four-level POA factor is coded using Helmert contrasts designed to capture the bilabial versus velar distinction, since exploratory plots suggested this contrast would have the largest effect on prevoicing duration. The two-level class variable is coded using sum contrasts (-1 = Dh, 1 = D). All possible by-word and by-speaker random effects for terms of interest (speech rate and class) were included Barr et al. (2013), with correlations between random effects excluded to avoid an overparametrized model. The model's fixed effects are summarized in Table 3.4.

The significant main effect of Speech rate ($p < 0.001$), with a negative coefficient, establishes that VD increases at slower speech rates, as expected (averaging across stop classes and places of articulation). The effect of Class is also significant ($p < 0.001$): voiced stops have longer prevoicing than voiced aspirated stops, on average, as observed in Figure 3.2. The effect of speech rate on VD does not, however, interact significantly with class: voiced aspirated stops have a slightly steeper speech rate effect than plain voiced stops, but the difference is not significant (Class:Speech rate: $p = 0.53$). Lastly, velar stops have significantly shorter prevoicing durations than bilabial stops ($p < 0.01$), consistent with the difficulty of maintaining voicing for less anterior constrictions J. J. Ohala (1983). Alveolar stops also have a slightly shorter prevoicing duration than (the mean of) bilabial and velar stops ($p = 0.03$), but retroflex stops do not differ significantly from non-retroflex stops ($p = 0.19$).

PVI duration

Based on Figure 2.5, the PVI durations of both aspirated stop classes appear to increase at slower speech rates at least slightly more than the burst durations of the other two classes. Since even the unaspirated classes do have some positive VOT, the model of PVI duration seeks to establish whether the effect of speech rate on PVI duration is greater for the Th/Dh stop classes than the T/D classes. The four-way Class factor was therefore coded with a contrast that compares the as-

Table 2.6: Linear mixed-effects model of log-transformed PVI duration, for stops with PVI duration>0 ($n = 426$).

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-3.698	0.095	-38.588	< 0.001
Class (Th/Dh vs. T/D)	1.405	0.168	8.375	< 0.001
Class (T vs. D)	0.399	0.188	2.130	0.048
Class (Th vs. Dh)	0.825	0.256	3.223	0.0038
Speech Rate	-0.211	0.111	-1.885	0.061
POA (alveolar vs. bilabial)	0.120	0.098	1.227	0.24
POA (retroflex vs. bilabial/alveolar)	0.068	0.056	1.21	0.24
POA (velar vs. bilabial/alveolar/retroflex)	0.135	0.038	3.626	0.0025
Class: Speech Rate (Th/Dh vs. T/D)	-0.459	0.207	-2.214	0.028
Class: Speech Rate (T vs. D)	0.194	0.207	0.934	0.351
Class: Speech Rate (Th vs. Dh)	0.541	0.310	1.745	0.083

pirated to the unaspirated classes, as well as two contrasts coding the difference between the two unaspirated classes and between the two aspirated classes. Before running the model, observations without a burst (PVI=0) were excluded ($n = 13$, 3.0% of total). The model included fixed-effect terms for Class, Speech rate, and the interaction between the two, as well as Place of articulation (POA). The four-level POA factor is coded using Helmert contrasts comparing less-anterior with more-anterior places (alveolar vs. bilabial, retroflex vs. alveolar/bilabial, etc.). The model's random-effect structure was as 'maximal' as feasible Barr et al. (2013).[6] The model's fixed effects are summarized in Table 2.6.

The model confirms the hypothesis of primary interest: PVI duration increases as a function of speech rate significantly more for the aspirated (Th and Dh) classes than the unaspirated classes (T and D) (Class: Speech Rate (Th/Dh vs. T/D): $p = 0.028$), while the speech rate effect does not significantly differ between the voiced and voiceless aspirated classes ($p = 0.083$). The model also confirms the observation in section 2.3 that the PVI duration of Th/Dh is significantly longer than that of the T/D classes ($p < 0.001$), and that the PVI duration of Th is significantly longer than that of Dh ($p = 0.0038$). The PVI of T and D also differ significantly ($p = 0.048$), though the effect size is smaller and the p-value less significant than the other contrasts (Th/Dh vs. T/D, Th vs. DH). POA is also a significant predictor of PVI, with velar segments having a significantly longer PVI duration than the other places of articulation, consistent with VOT being greater for more posterior articulations Lisker and Abramson (1964). (Recall that PVI is closely related to positive VOT.)

[6] The model included by-speaker and by-word random intercepts as well as random slopes for Speech rate (by-speaker, by-word), Class (by-speaker), and for the Speech rate-by-Class interaction for the key Class contrast, capturing the difference between Th/Dh and T/D. Adding in the remaining two Speech rate-by-Class interaction terms led to an overparametrized model which did not converge. We also did not include correlations between random effects for the same reason.

Summary

The speech rate diagnostic provides evidence for the Iverson and Salmons (1995) representation of Nepali stops. Prevoicing duration of the D and Dh stops increases as speech rate slows, supporting the representation of those two classes with specified [voice] features. Burst duration of the aspirated stops also increases at slower speech rates, significantly more than the burst duration of the unaspirated stops, supporting the representation of Th and Dh with a [spread] feature and T and D without one.

Passive vs. active voicing in medial position

The intervocalic voicing diagnostic

Recall that Beckman et al. (2013) proposed examining voicing during the closure of intervocalic stops as another link between control and privative feature specification. They find that in voicing languages like Russian, *b, d, g* stops are voiced throughout the entire closure (operationalized as 90% of the closure) an average 97% of the time, with velar stops fully voiced the least often at 91%. In German, an aspirating language, the *b, d, g* series is voiced throughout the closure an average of only 62% of the time, with velar stops fully voiced least often at only 25%. In the remaining 38% of German *b, d, g* stops, voicing continues only partway into the closure. Beckman et al. propose that Russian's high percentage of fully-voiced stops requires that speakers actively maintain voicing during the closure just as they actively prevoice during the closure of word-initial stops. Actively maintaining voicing in intervocalic position is an action controlled by the speaker, a result of the stop being phonologically specified for [voice]. The low proportion of voicing during closure in German stops reveals a different type of voicing: passive voicing, an effect of the voicing of the preceding vowel bleeding into the closure of the stop Stevens (1998). Passive voicing is not controlled or intentional by speakers—it is an automatic phonetic consequence of being preceded by a voiced vowel—and thus follows from German's *b, d, g* stops being unspecified for voicing.

In addition to active and passive voicing, some stop classes actively block voicing during the closure. While Beckman et al. (2013) do not discuss data for this beyond an example of one instance of the word *papa* in Russian, they state that German's [spread]-specified *p, t, k* stops and Russian's laryngeally unspecified *p, t, k* stops both actively block voicing, a phonological status phonetically manifested as voicing approximately 20% of the closure. Möbius (2004) reports similar results for German *p, t, k*, finding that most have lost voicing 30% of the way into the closure. Pape and Jesus (2014) provide additional data, finding that European Portuguese *ptk* stops also have 20-30% of the closure voiced[7]. German *p, t, k* are said to block voicing because the active [spread] feature corresponds to a glottis that is too widely spread for voicing from the preceding vowel to continue into closure. It is harder to explain why the unspecified *p, t, k* stop in a voicing language should also block voicing, and this remains an open question in laryngeal

[7]One purpose of Pape and Jesus (2014)'s study is to evaluate whether European Portuguese behaves more like a voicing or aspirating language. The intervocalic voicing behavior of *p, t, k* supports Beckman, Jessen et al. (2013)'s arguments regardless of their conclusion, as unspecified *p, t, k* in a voicing language and [spread]-specified *p, t, k* in an aspirating language both actively block voicing, according to Beckman, Jessen et al. (2013).

Table 2.7: Intervocalic voicing in Russian and German based on data from Beckman, Jessen et al. (2013) and Möbius 2004.

Language	Representation	Segments	Voicing during closure	Type of voicing
Russian	[voice]	b, d, g	97% of stops are >90% voiced	Active voicing
	[ ]	p, t, k	mean voicing ∼20% of closure	Blocking voicing
German	[ ]	b, d, g	62% of stops are >90% voiced	Passive voicing
	[spread]	p, t, k	mean voicing ∼20-30% of closure	Blocking voicing

realism. Descriptively, however, it is the case that unspecified stops in languages with an active [voice] feature on another stop class block voicing. The various voicing profiles are summarized in Table 2.7.

Applying the diagnostic to Nepali

The link between intervocalic voicing and feature specification runs into two conflicts when extended to a language like Nepali, which exploits both a [spread] and a [voice] contrast. Under the assumption of laryngeal realist features taken here, the first conflict is that Nepali's plain voiceless stops are unspecified for both [spread] and [voice]. If these stops pattern like the unspecified class in a [voice] language, they should actively block voicing, like in Russian. If, however, they pattern like the unspecified class in a [spread] language, we expect them to permit passive voicing, like in German.

The voiced aspirated stops present a second conflict, since they are specified for both [voice] and [spread]. If they pattern like stops specified for [spread], we expect them to actively block voicing during closure. If they pattern like stops specified for [voice], however, we expect them to actively maintain voicing throughout the closure. The voicing proportions of intervocalic stops are shown in Figure 2.6, and summarized in Table 2.8.

Table 2.8: Intervocalic voicing in Nepali

Language	Representation	Segments	Voicing during closure mean (%)	median (%)	Type of voicing
Nepali	[voice]	b, d, g	88	100	Active voicing
	[spread]	p^h, t^h, k^h	9	0	Blocking voicing
	[ ]	p, t, k	10	0	Blocking voicing
	[voice, spread]	b^h, d^h, g^h	89	100	Active voicing

The voiced class has full voicing (again operationalized at 90% voicing) 88% of the time. The voiced aspirated stops have full voicing 89% of the time, as is expected of [voice]-specified stops

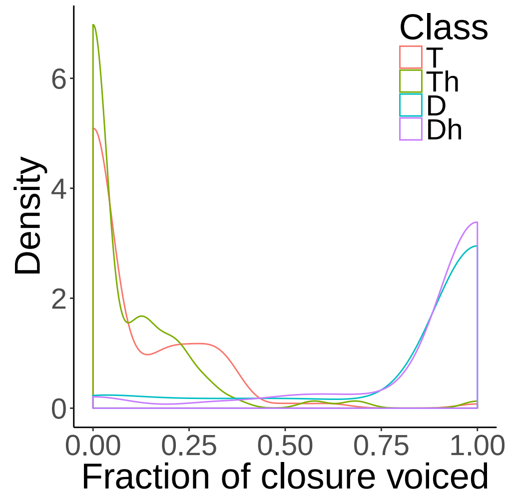


Figure 2.6: Distribution of the voicing fraction for stops in medial (intervocalic) position, for each stop class.

rather than [spread]-specified stops. We interpret the D and Dh class results as Russian-like active voicing and characteristic of stops specified for [voice], though voicing is not quite as consistent as the 92% found in the Russian study. At the same time, the mean proportion of voicing during closure in Th stops is 9%, which is on par with active blocking of voicing. The T stops have a similarly low voicing proportion (10%), displaying the active blocking characteristic of an unspecified stop in a [voice] language rather than a [spread] language. Recent unpublished work by Beckman on Hindi found similar amounts of voicing for intervocalic T stops, reaching a similar conclusion that they pattern like unspecified stops in a [voice] language Beckman (handout). Figure 2.6 shows the voicing fraction distribution for each class, displaying that the voiced aspirated stops have as much full voicing as plain voiced stops and plain voiceless stops have a similar degree of voicing to voiceless aspirated stops.

To confirm these observations, ideally we would run a statistical model evaluating the effect of stop class (i.e., voicing and aspiration specification) on voicing proportion. The structure of the voicing proportion data—which contains many 0's and 1's (no or full closure voicing) in addition to values between 0 and 1—makes familiar methods such as linear or logistic regression inappropriate. The dataset is not large enough ($n = 339$) to fit more complex models, such as zero-one inflated binomial regression, which respect its structure. Instead, we simply perform non-parametric hypothesis tests to establish the basic pattern in Figure 2.6. Wilcoxson rank-sum tests show that D and Dh classes have significantly larger voicing proportions than the T and Th classes ($W = 1347$, $p < 0.0001$), and that the T and Th classes do not significantly differ in VP ($W = 4625$, $p = 0.53$), nor do the D and Dh classes ($W = 2818$, $p = 0.92$). Although this method does not control for other factors affecting voicing proportion (such as place of articulation) or differences between speakers and words, the basic pattern in Figure 2.6 is clear: belonging to a class specified for [voice] is the main predictor of voicing proportion.[8]

Based on these results, we may conclude that both of Nepali's stop classes that pose conflicting demands in intervocalic position pattern like stops in a [voice] language rather than like stops in a [spread] language, suggesting that [voice] is the 'stronger' feature of the two. An explanation for this asymmetry is beyond the scope of this paper, but in closing we present two possible directions. Voicing could be stronger in intervocalic contexts due to its position between two (voiced) vowels. Alternatively, just as Beckman, Jessen et al. (2013) propose a phonological (i.e. representational) explanation for the intervocalic voicing effects of German vs. Russian stops, we could propose a phonological (i.e. representational) reason for the 'strength' of one feature over the other. Schwarz (2017) presents an explanation of the latter type, in which the [voice] and [spread] features are temporally ordered within a single stop.

2.5 Motivating a subsegmental representation

In this study we have examined the laryngeal contrasts of Nepali because they, like similar contrasts in other Indic languages, pose potential challenges for the types of evidence used and predictions made by laryngeal realist theories. The challenges stem largely from the existence of voiced aspirated stops, which were proposed to be specified for two seemingly oppositional features, [voice] and [spread]. We set out to extend the theory of laryngeal realism to an area where it had not yet been examined by asking the following question: do the types of evidence used to motivate feature representations in languages with two- and three-way contrasts without classes that are specified for two features work for Nepali, and do they support both the laryngeal realism view and the feature representation standardly used for this type of system? We investigated this question with two types of evidence so far: phonetic realization in initial position, and diagnostics of control in initial and medial position.

Acoustic analysis showed that a combination of prevoicing duration and PVI duration distinguishes the four stop classes from each other in word-initial position. These phonetic findings suggest that durational measures are indeed sufficient for distinguishing the Indic four-way contrast, as long as negative VOT and positive VOT are measured as distinct cues. This motivates the specification of the voiced stops as [voice] and the aspirated stops as [spread]. Speech rate effects showed that prevoicing duration and long-lag burst duration both appear to be controlled while short-lag burst duration does not, motivating the specification of the features as privative, as in Table 2.1 and repeated here.

[8] We experimented with more complex statistical models, including mixed-effects linear regression and mixed-effects zero-one-inflated binomial models, using models as complex as possible given the small sample size. All models came to the same qualitative conclusion: D/Dh shows more closure voicing than T/Th, while other stop classes do not reliably differ.

Table 2.9: Feature representation of an Indic style four-way contrast, as proposed by Iverson and Salmons (1995)

[p, t, k]:	[pʰ, tʰ, kʰ]:	[b, d, g]:	[bʰ, dʰ, gʰ]:
[ ]	[spread]	[voice]	[spread], [voice]

We next applied the intervocalic voicing diagnostic, which showed that T stops and Dh stops each patterned like the stops in a [voice] contrasting language rather than a [spread] contrasting language. This poses an issue if, like most phonological representations of Indo-Aryan voiced aspirated stops, we propose that Dh segments are simultaneously specified for the two laryngeal features [voice] and [spread] (Iverson and Salmons 1995).

I propose that the intervocalic behavior of voiced aspirated stops is expected in a Q Theoretic subsegmental representation that temporally orders the two laryngeal features within a segment[9].

Q Theory horizontally explodes a segment (Q) into a series of temporally ordered (q) subsegments (S. Shih and Inkelas 2014; Inkelas and S. S. Shih 2016; Inkelas and S. S. Shih 2017). The standard Q is comprised of three subsegments (q_1 q_2 q_3), as necessitated by the prenasalized affricate and HLH-toned contour segments in (2).

(2) a. [ⁿʧ]: (n t ʃ)
b. [ã]: (á à á)

Drawing from Like Articulatory Phonology (Browman and Goldstein 1986; Gafos 2002), which posits that segment-internal gestures are organized around the ‘landmarks’ of movement onset, target, and release, Q Theory offers a quantized representation of these landmarks more compatible with frameworks of phonoloigcal analysis. In Q Theory representations, q_1 corresponds roughly to the transition into the segment, q_2 to the target of the segment, and q_3 to the offset or transition out of the segment. While the default number of subsegments is three, this may be varied, provided there is the phonological motivation to do so (Inkelas and S. S. Shih 2016; Schwarz, Lapierre et al. 2019).

Each subsegment is a uniform feature bundle, and Qs may be comprised of identical qs, or of qs with distinct feature bundles. I propose the following Q Theoretic representation of Nepali stops, some of which are apparently simple and uniform, and some of which contain contour.

(3) T: (t t t)
D: (d d d)
Th: (t t h)
Dh: (d d h)

Laryngeal features are distributed across the progression of the segment as in (4).

[9]Aperture Theory (Steriade 1993; Steriade 1994) and Q Theory (S. Shih and Inkelas 2014; Inkelas and S. S. Shih 2017) make the same predictions. I model the Q Theoretic analysis here due to its broad utility in modeling a wide range of phenomena, but see Hussain and Nair (1995) for an Aperture Theory analysis of laryngeal contrasts in Urdu.

(4) With features:

T	(t	t	t)
	[ ]	[ ]	[ ]
D	(d	d	d)
	[voice]	[voice]	[voice]
Th	(t	t	h)
	[ ]	[spread]	[spread]
Dh	(d	d	h)
	[voice]	[voice]	[spread]

Dh segments (Qs) still contain both [voice] and [spread] features, but they are not both hosted by a single (subsegmental) unit. [voice] is associated with the onset and target (closure) subsegments, while the [spread] feature is associated with the offset (release) of the stop, matching the primary loci of the cues corresponding to each feature. Furthermore, the intervocalic voicing behavior is expected under this account. Intervocalic voicing is a diagnostic of the closure, and voiced aspirated stops have the same representation as voiced stops across the two subsegments that precede the release. In the representation proposed in (4), Th has a [spread] feature on both its q_2 and q_3 subsegments. This aims to capture the blocking of voicing during Th stop closure, proposing that the target of the stop involves the abduction during the closure (though this should be further investigated by articulatory studies of Nepali's stops)[10]. This representation also supports the asymmetric loss of aspiration on Dh but not Th in final position that will be the focus of Chapter 4.

[10]If a [spread] during the closure is used to explain the active blocking of voicing exhibited by Th stops, one might wonder if T stops should also be specified with a [spread] feature during the closure, since they also actively block voicing. I continue to consider this possibility, but have not included it in this representation for several reasons. First, specifiying a [spread] feature on an unaspirated stop would problematize the association of [spread] with long lag VOT that is central to phonetically-grounded laryngeal representations. Second, T did not pattern like a stop with a controlled aspiration cue in the speech rate diagnostic. One could address these two concerns by revising the diagnostics to specifically apply to stops with a [spread]-specified *release*, which would make some sense since the presence of long-lag VOT and its behavior at different speech rates is a cue of the release. However, in the absence of e.g. laryngopic evidence that the vocal folds are adducted during the closure of T stops, I hesitate to propose that T is specified for [spread], even just in its closure. Instead, I appeal to the explanation of Beckman, Jessen et al. (2013) and propose that T stops block voicing because they are unspecified stops that exist in a voicing contrast. While they also exist in contrast to aspirated stops, the voicing contrast is more pertinent during the closure.

Chapter 3

The voicing effect

3.1 Introduction

This chapter brings new empirical evidence to bear on another vexing central question of laryngeal contrasts: how, and why, does the laryngeal specification of a consonant affect the duration of the preceding vowel?

As previewed in Chapter 1, discussions of the 'voicing effect', in which vowels tend to be longer before voiced consonants than voiceless, have often abstracted over the phonetic cues used to differentiate those 'voiced' and 'voiceless' consonants. It could help to narrow down the possibilities for the source of the effect if there were similarly robust generalizations about the effect on vowel duration of other laryngeal specifications besides voicing. However, fewer studies examine the correlation between other laryngeal specifications and voicing. The majority of these studies involve a single language (Hindi), and their results have been somewhat inconsistent (Maddieson and Gandour 1976; M. Ohala and J. J. Ohala 1992; Maddieson 1997; Lampp and Reklis 2004; Durvasula and Luo 2012; Begus 2017).

Hindi displays a four-way laryngeal contrast between voiceless unaspirated, voiceless aspirated, voiced, and voiced aspirated stops, allowing an analysis of an effect of a stop's aspiration specification on preceding vowel duration independent from the effect of voicing. In the most recent study, Durvasula and Luo (2012) find evidence for a significant effect of aspiration specification on vowel duration: vowels are longer before aspirated classes of stops than before unaspirated classes. Durvasula and Luo term this the aspiration effect', parallel to the better known voicing effect also attested in Hindi. The results from other studies of Hindi trend in the same direction, though sometimes the durational differences are quite small, do not reach levels of significance, or are only present in certain places of articulation.

Aspiration effect (Durvasula & Luo 2012). *Vowels are longer before aspirated stop classes than before unaspirated stop classes.*

In another study examining vowel duration preceding stops with laryngeal contrasts beyond just voicing, Beguš (2017) analyzes Georgian's contrast between voiceless unaspirated, voiced, and ejective stops and finds evidence for a inverse VOT effect in which VOT duration correlates negatively with preceding vowel duration.

Inverse VOT effect (Beguš 2017). *VOT duration correlates negatively with preceding vowel duration.*

Beguš's inverse VOT effect may initially strike as contradictory to the aspiration effect found by Durvasula and Luo, because aspirated Hindi stops have both longer VOT and longer preceding vowels than their unaspirated counterparts. However, differences between Georgian and Hindi's inventories, along with differences in the studies' focus on phonological specification versus measures of phonetic cues prevent the direct comparison of the studies that produce these two proposals. The Hindi studies primarily examine the duration of vowels as a function of the following stop's phonological specification as either aspirated or unaspirated. Though Durvasula and Luo do measure and address the relationship of stop closure duration and vowel duration, they do not test for the effect of VOT, as a phonetic durational measure, on vowel duration. Beguš's study of Georgian considers both phonological specification as well as the phonetic cues that signal the phonological category, particularly VOT. However, Georgian's laryngeal contrast between voiceless aspirated, voiced (unaspirated), and ejective stops does not allow for analysis of aspiration as a phonological specification independent of any other laryngeal specification, and therefore cannot speak to Durvasula and Luo's main claim that phonological aspiration correlates to a longer preceding vowel duration. In order to assess the compatibility of Beguš's inverse VOT effect with Durvasula and Luo's aspiration effect, we must examine the effects of both phonological specification and phonetic cues signalling aspiration on preceding vowel duration, in a language with an aspiration contrast independent of voicing.

To this end, the present study analyzes the effects of both phonological specification and phonetic cues on preceding vowel duration in Nepali, an Indic language spoken in Nepal and northern India, that employs the same four-way laryngeal contrast as Hindi. Analysis of the effects of phonological specification on vowel duration provides further data on the empirical question of whether the aspiration effect exists. Examination of the effects of positive VOT duration on vowel duration, within each phonological category, tests Beguš's inverse VOT effect in a language with an aspiration contrast independent of voicing. By teasing apart the influences of phonological and phonetic factors on vowel duration, the study can contribute to discussions of the source of the aspiration effect.

The remainder of the chapter is organized around two questions: Are Durvasula and Luo's aspiration effect and Beguš's inverse VOT effect contradictory, or can both be present in a single language? How do the results reflect on possible sources for the aspiration effect?

Based on data from a production experiment with methods comparable to both Durvasula and Luo (2012) and Beguš (2017), I present evidence of an aspiration effect in Nepali, a language with phonological specification of aspiration that is independent of voicing: vowels preceding stops belonging to aspirated classes are slightly but significantly longer than vowels preceding stops belonging to unaspirated classes. I also find evidence of a within-category inverse VOT effect: within the aspirated classes, preceding vowel duration correlates negatively with VOT duration. The coexistence of both the aspiration effect and the VOT effect in Nepali shows that these two proposals are not necessarily contradictory. It also speaks to the source of the aspiration effect. Across phonological categories, phonological specification for aspiration correlates positively with vowel length. Within phonological categories, the phonetic measure of aspiration duration correlates negatively with vowel duration. This pattern of opposite distributions within- and across-category is consistent with Clayards's (2017) schematic of a cue-trading relationship, and contrary to the distribution expected of intrinsically-linked cues. While this study does not rule out articulatory or timing-based accounts for an aspiration effect, it indicates that, like the voicing effect in English, the effect of aspiration on vowel duration in Nepali is not purely physiological or automatic.

3.2 Previous studies of the aspiration effect and inverse VOT effect

Previous studies of the effect of membership in an aspirated phonological class on the duration of the preceding vowel, while largely based on Hindi, report varying results. Maddieson and Gandour (1976) test Hindi coronal stops in a sa:C frame and find evidence of an aspiration effect: vowels preceding [t^h] and [d^h] are longer than the vowels preceding their unaspirated counterparts. Maddieson (1997) expands the range of languages considered, and tests for the aspiration effect in 5 languages: Assamese, Bengali, Hindi, Marathi, and Eastern Armenian. The first four languages of the list have the same four-way contrast, while Eastern Armenian contrasts three ways with voiceless unaspirated, voiceless aspirated, and voiced stops. While the direction and significance of the vowel duration differences vary both cross-linguistically and language-internally by voicing and place of articulation, Maddieson concludes that lengthening before aspirated consonants is the predominant pattern. Of Maddieson's 49 language, voicing, and POA-specific comparison pairs, vowels before aspirated consonants are significantly longer than vowels before unaspirated consonants in 14 pairs and nonsignificantly longer in 19[1,2]. Vowels preceding unaspirated consonants are nonsignificantly longer than vowels preceding aspirated consonants in 6 pairs, and significantly longer in 2 pairs (both in Eastern Armenian).

Ohala and Ohala (1992) test the aspiration effect in Hindi stops of multiple places of articulation in a sVCa frame, and find inconsistent results across voicing and place of articulation. While vowels preceding [d^h] are slightly longer than vowels preceding [d], vowels before [t^h] and [p^h] are slightly shorter than their unaspirated counterparts. Vowels preceding [g^h]/[g], [b^h]/[b], and [k^h]/[k] are nearly identical in length. Ohala and Ohala conclude that there is no consistent difference in vowel duration due to aspiration. This may be due at least in part to a lack of statistical power in a study with a small sample size.

[1] There are no apparent patterns to the pairs showing significance vs. nonsignificance. Voiced pairs are slightly more commonly nonsignificant than voiceless, and Marathi and Hindi have a slightly higher proportion of nonsignificant pairs than Assamese and Bengali.

[2] The raw durational difference between the aspirated and unaspirated pairs is actually longer in Maddieson's study (~9ms) than in many of the other previous studies that did find significant differences. The lack of significance in this study is likely due in part to smaller sample size.

In an effort to replicate Maddieson and Gandour's (1976) results, Lampp and Reklis (2004) test Hindi velar stops in a dC frame and find a voicing effect but no significant aspiration effect, though the mean duration of vowels preceding voiced aspirated stops are slightly longer than before voiced unaspirated.

Most recently, Durvasula and Luo's (2012) study does find evidence of a significant aspiration effect in Hindi. Testing dental stops with data from 7 participants, Durvasula and Luo find that aspirated stops are slightly, though significantly, longer than their unaspirated counterparts (T: 146ms, Th: 150ms, D: 157, Dh: 160ms)[3].

Most of these studies consider vowel duration in terms of phonological specification, but Beguš (2017) considers whether the source of the effect may be phonetic cues to the phonological class distinctions, like VOT duration, rather than the phonological specification itself. Beguš (2017) measures the duration of vowels preceding each of Georgian's three-way contrasting stops: voiceless aspirated, voiced, and ejectives. Like Ohala & Ohala (1992), the stimuli put vowels in open syllables by embedding VxCy sequences in frames to create (largely) nonsense words with a CVCVxCyV shape. Beguš finds that, collapsing across vowel type and place of articulation, vowels are longest before voiced stops (96.3ms), shorter before ejective stops (87.1ms), and shortest before voiceless aspirated stops (82.3ms). These durational differences are found to be significant in a model with phonological specification as predictors. In addition to testing the effect of phonological specification on vowel duration as in previous studies summarized here, Beguš also tests the effect of VOT duration, as a continuous phonetic measurement, on preceding vowel duration. In a model that considers both laryngeal specification and VOT duration, laryngeal specification is found to be significant even while controlling for phonetic measures of VOT duration and closure duration, suggesting that the effect is in part phonological, or at least not entirely based on these two phonetic measures. This same model finds that continuous VOT duration, as a main effect, is inversely correlated with preceding vowel duration for all V and C types. With a p-value of 0.052, this effect does not quite reach a $p<.05$ level of significance, though Beguš points out that including VOT and its interactions does significantly improve the fit of the model.

As noted before by Durvasula and Luo (2012) and Beguš (2017), the results of this handful of studies that consider the effect of laryngeal specifications besides voicing on the duration of the preceding vowel are rather inconsistent. While the studies rarely report vowels that are longer before unaspirated than aspirated stops, and while the more recent studies have found evidence for a significant aspiration effect, the magnitude and significance of the aspiration effect can vary across study, language, voicing, and place of articulation. Moreover, Beguš's finding of an inverse correlation between vowel duration and VOT duration seems to contradict the findings of several of the earlier studies that aspirated classes imply longer vowels, creating an intriguing mismatch between phonological and phonetic patterns. This mismatch cannot be fully explored in Beguš's study of Georgian alone as its laryngeal contrasts to not allow evaluating the effect of aspiration specification independent of voicing. The present study adds to the body of empirical data testing the proposed aspiration effect, and tests for an inverse VOT effect in a language that can also speak to a phonologically-based aspiration effect.

[3]Durvasula and Luo do not report raw duration measures, so these numbers are estimated based on their plot.

3.3 Methods

Participants

8 participants (4 female, 4 male) contributed to this study. All were native speakers of Nepali living in Kathmandu. All spoke some amount of English, all had some familiarity with Hindi (nearly all Nepali speakers do), and some spoke an additional mother tongue language.

Stimuli

To maximize comparability with the previous studies, the stimuli used the templates sVC, pVC, and kVC in which V was /a, i, u/ and C was /p, p^h, b, b^h, t, t^h, d, d^h, k, k^h, g, g^h/ to yield 108 nonsense word test items. Each word was written in Devnagari script (Nepali's syllabic orthography) on its own cue card. The participants were shown each target word in a randomized order, and asked to produce that word in the carrier sentence in (1).

(1) X_1 (pause). mʌ ʌbʌ X_2 b^hants^{h}u: (pause) X_3
'X_1 (pause) now I say X_2: (pause) X_3.'

Acoustic analysis

The recordings were imported into Praat (Boersma and Weenink 2015), where they were hand annotated for vowel duration, closure duration, voicing duration during the closure, and post-release duration. Vowel duration was measured from the onset of modal voicing to the beginning of the closure of the coda consonant. Voicing duration was measured from the beginning of closure (marked by a sharp decrease in amplitude in the waveform signalling the end of the preceding vowel) to the end of visible periodicity in the waveform. Closure duration was measured until the release burst. Post-release duration was measured from burst release to the end of visible aperiodicity (Turk et al. 2006; Mikuteit and Reetz 2007; Berkson 2012).

3.4 Results

This study first asks whether the aspiration effect and VOT effect are contradictory, or whether they can both hold true in a single language. In order to address this question, I examine empirical evidence for the aspiration effect and empirical evidence for the VOT effect.

Evidence for the aspiration effect in Nepali

To determine the presence of an aspiration effect in Nepali, vowel durations were measured preceding stops from each of the four laryngeal classes. Because stimuli were read from written nonsense word forms, stops were categorized based on orthography, which corresponds to phonological specification. Like Durvasula & Luo found for Hindi, Nepali shows evidence of both an aspiration effect and a voicing effect. Mean vowel duration preceding voiceless aspirated and voiced aspirated stops is longer than the duration of vowels preceding their unaspirated class counterparts. Vowels preceding voiced aspirated and voiced unaspirated stops are also longer than vowels preceding the voiceless counterparts. The effect is visualized in Figure 1, and Table 2 summarizes mean durations with standard deviations for each class.

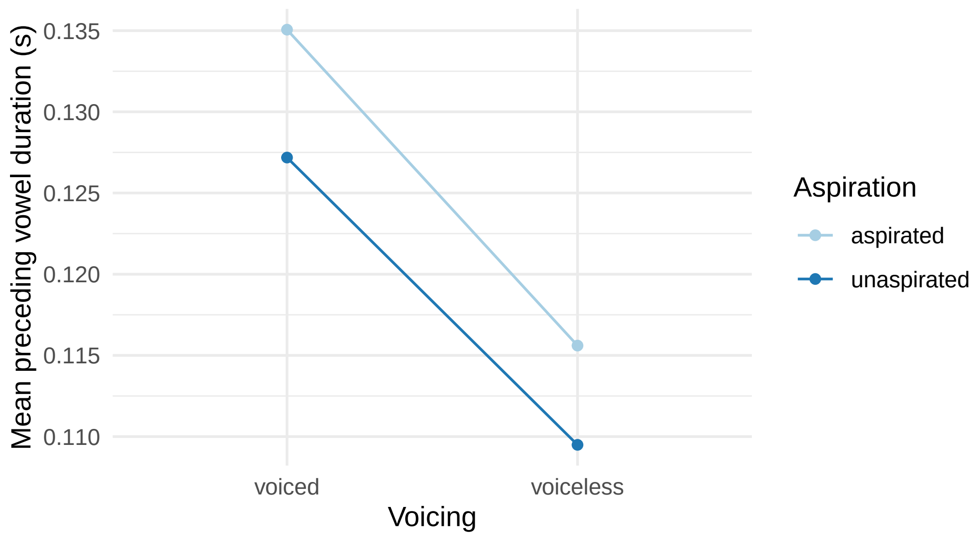


Figure 3.1: Mean vowel durations preceding each of Nepali's four laryngeal classes of stops.

Table 3.1: Mean vowel durations preceding each laryngeal class

		preceding V duration	
Class	n	mean	sd
Voiceless (T)	354	109	38
Voiceless aspirated (Th)	287	114	42
Voiced (D)	361	127	45
Voiced aspirated (Dh)	349	135	46

Also consistent with Durvasula and Luo's findings, the raw durational difference between aspirated and unaspirated stops is smaller than the difference based on voicing. Th stops are 5ms longer than T stops, and Dh stops are 8ms longer than D stops. Meanwhile, the difference between D and T stops is 18ms, and 21ms between Dh and Th stops.

A linear mixed effect model was fitted to test the patterns observed in these plots and the significance of the values reported in the table. The model considers preceding vowel duration as the dependent variable. To test for an aspiration effect, it evaluates whether a stop's membership in an aspirated vs. unaspirated class significantly affects preceding vowel duration. To test for a voicing effect, it evaluates whether a stop's membership in a voiced vs. voiceless class significantly affects preceding vowel duration. Models were fitted using the lme4 package in R (Bates et al. 2015). Fixed effect coefficients are shown with associated standard errors test statistic (t) and significance, calculated with lmerTest (Kuznetsova et al. 2015) using the Satterthwaite approximation. Random effects are not shown.

The model presented here includes as predictors aspiration specification (aspirated vs. unaspirated), voicing specification (voiced vs. voiceless), preceding vowel quality ([a], [i], [u]), place of articulation of the coda consonant, the interaction of voicing specification and vowel quality, and the interaction of voicing specification and POA. It also considers random intercepts for both word and speaker. This model was arrived at by beginning with the fullest possible model, including all possible interactions, random slopes and random intercepts. Voiceless aspirated bilabial stops [p^h] are consistently realized as fricatives in the data and were therefore excluded, causing any model that included the interactions with place of articulation and aspiration to be rank deficient. After removing these interactions to avoid rank deficiency, interactions were dropped in a stepwise manner to determine the optimal model based on Akaike Information Criteria (AIC). Aspiration and voicing specification were found to be significant in all of the non rank deficient models.

Table 3.2: Linear mixed-effect model of voicing duration with phonological voicing and aspiration class as predictors.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.12	0.008	13.982	< 0.001
Voicing (voiced)	0.019	0.002	6.047	< 0.001
Aspiration (aspirated)	0.0046	0.0015	2.998	0.003
Vowel ([a])	0.04	0.001	37.757	< 0.001
Vowel ([i])	-0.019	0.001	-19.004	<0.001
Voicing (voiced): Vowel [a]	0.003	0.001	2.78	0.006

The resulting model shows that both classification as aspirated and classification as voiced significantly increase the length of the preceding vowel. As expected, classification as voiced has a larger effect on the duration of the vowel than classification as aspirated. Vowel quality also expectedly has a significant effect on vowel duration, with [a] displaying the longest duration. POA of the following stop was not a significant predictor. The one significant interaction, between voicing and vowel quality, is also included: the voicing effect is stronger with [a] than with the other vowels, perhaps because of its greater inherent length[4].

[4]The fullest model does not converge. In the iterations of the model that include the interaction of aspiration vs. place of articulation (and are therefore rank deficient) aspiration classification is not significant. In all the iterations of the model without the aspiration vs. place of articulation interaction, aspiration specification is a significant predictor of vowel length.

Like the previous studies described in Section 2.2, the size of the effect that the present study finds is small. This raises the key question of whether a difference of 5-8ms is perceptible, which will require a future perception study to address.

Evidence for the inverse VOT effect in Nepali

Having found evidence of the aspiration effect in Nepali, I next test for Beguš's inverse VOT effect: does aspiration duration correlate inversely with vowel duration? Section 4.1 analyzed vowel duration based on the following stop's membership in a particular phonological class, and found that overall, the vowels preceding stops belonging to aspirated classes are longer than vowels preceding unaspirated classes. The analysis in this section compares the correlation between two phonetic measures while controlling for phonological representation, and finds that within phonological classes, Nepali VC sequences show an inverse correlation between vowel duration and aspiration duration, consistent with what Beguš (2017) reports for Georgian. The inverse correlation is visible in Figure 2. The tokens with longer aspiration are preceded by shorter vowels, and the tokens with shorter aspiration are preceded by longer vowels.

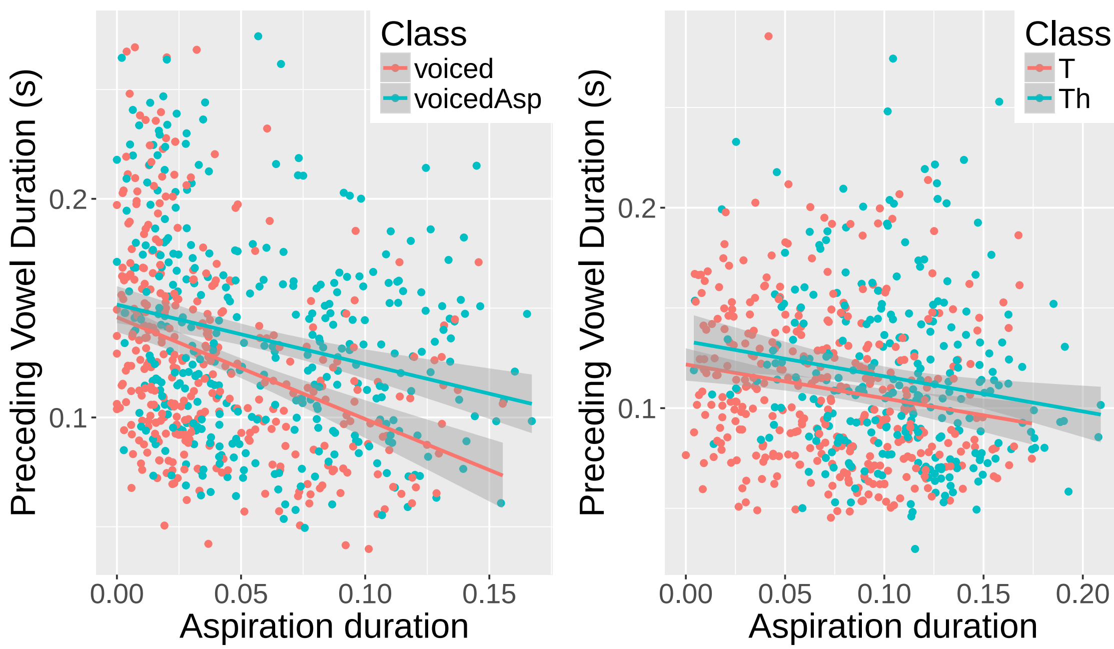


Figure 3.2: Visualization of vowel duration vs. aspiration duration for each of the voiced classes (left) and voiceless classes (right).

To test significance of these patterns, a second linear model adds positive VOT duration as a predictor of vowel duration. Like the model testing the aspiration effect, it also includes voicing specification, aspiration specification, vowel, and POA, as well as random intercepts for speaker and word. Facing the same rank deficiency constraints as the aspiration effect model, this model was again determined to be optimal based on AIC after dropping interactions in a stepwise manner. Ultimately it includes two- and three-way interactions between voicing specification, aspiration duration, and vowel quality. Significant effects are reported in Table 4.

Table 3.3: Linear mixed-effect model of vowel duration including continuous positive VOT (aaspiration duration) as a predictor.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.123	0.008	14.625	< 0.001
Voicing (voiced)	0.01	0.001	7.181	< 0.001
Aspiration (aspirated)	0.005	0.002	3.46	<0.001
Aspiration duration	-0.046	0.020	-2.253	0.024
Vowel ([a])	0.041	0.001	21.920	< 0.001
Vowel ([i])	-0.017	0.001	–8.831	<0.001
Voicing (voiced) : Vowel [a]	0.007	0.001	4.076	<0.001
Aspiration duration : Vowel [a]	-0.048	0.023	-2.064	<0.03
Voicing (voiced) : Aspiration duration : Vowel [a]	-0.089	0.023	-3.866	<0.001

In this model aspiration duration has a significant negative effect on vowel duration, confirming the patterns observed in Figure 2 and the presence of an inverse VOT effect in Nepali. The model also finds phonological specification for both voicing and aspiration to still be significant as well. As in the aspiration effect model, vowel quality is significant, and the effects of both the aspiration effect and inverse VOT effect are stronger in [a] than in the other two inherently shorter vowels.

Discussion: Implications for the source of the aspiration effect

The results presented in this study, which add to the collection of empirical data on the relationship between vowel duration and consonants with laryngeal specification besides voicing, can help bear on the question of how and why laryngeal properties can affect the duration of the vowel that precedes them. Many competing theories have been proposed to account for vowel differences due to laryngeal specification. A comprehensive summary can be found in Beguš (2017), but the theories generally fall into three categories: articulatory/physiologically-based explanations, timing-based explanations, and perceptually-based explanations. After briefly summarizing each category, I discuss how the data from Nepali bear on them.

Articulatory-based explanations

Articulatory-based explanations propose some physiological link between the articulation of the consonant and the length of the preceding vowel, making vowel length an automatic phonetic consequence of the consonant articulation. These articulatory-based accounts are, thus far, all based in the differences in articulation of voiced vs. voiceless sounds, but Maddieson & Gandour (1976) extend the predictions to aspirated vs. unaspirated sounds and conclude that none predict longer vowels before aspirated than unaspirated sounds.

One articulatory theory (Halle et al. 1967; Chomsky and Halle 1968; Mohr 1971) proposes that because the manner of vocal fold vibration during voiced consonant closures is distinct from the manner of vocal fold vibration during vowels, the extra time it takes to adjust the larynx for consonant voicing accounts for the longer vowels before voiced consonants. This theory predicts that vowels preceding aspirated consonants would be longer than the vowels preceding unaspirated consonants if the laryngeal adjustment between vowels and aspirated consonants is slower than between vowels and unaspirated consonants. Maddieson & Gandour (1976) argue that because the glottis is not any more opened for aspirated consonants than unaspirated, the transition between vowel and aspirated stop is not any more complex, and the laryngeal adjustment prediction is not borne out.

Another set of proposals concern articulatory energy expenditure (Belasco 1953; Chen 1970). Based on a premise that syllabic units or VC sequences use a relatively consistent amount of energy, consonants that require more energy/force to articulate divert energy from being spent on vowels. If voiceless consonants require more force to produce, it follows that the vowels preceding them must use less energy, which translates to shorter duration. Maddieson & Gandour argue that aspirated stops likely require more force, or are more fortis, than unaspirated stops, which again predicts longer vowels before unaspirated classes, opposite to the results of their study and the present study.

While the present acoustic study cannot weigh in on the articulatory particulars of aspirated vs. unaspirated consonants, the data presented can contribute to the discussion of an intrinsic link between laryngeal setting and preceding vowel duration more broadly. For the aspiration effect, a physiological explanation would predict an intrinsic link between vowel duration and aspiration, regardless of exactly what that link entails articulatorily. The Nepali data presented here cast some doubt on an intrinsic link. Comparing across phonological categories, aspirated classes were shown to have longer vowels than unaspirated sounds. Within categories, however, the investigation into Beguš's VOT effect found that stops with longer aspiration had shorter vowels. The opposite directions of the within-category and between-category patterns would be difficult to reconcile with an intrinsic link between presence or duration of aspiration and length of preceding vowel, suggesting that the aspiration/VOT effects are not solely articulatory or phonetic. Ultimately it is likely that multiple factors contribute to the vowel length distinction and articulatory effects may be among them, but articulatory effects cannot be the sole cause.

Timing-based explanations

Timing-based explanations follow from a position that certain units of speech are relatively consistently timed, even when the particular sounds that make up those units vary. The exact size of the isochronous unit is debated, with proposals including the syllable, the vowel + full consonant (including release) (Lindblom 1967; Farnetani and Kori 1986), and the vowel + closure of the following consonant (Slis and Cohen 1969; Coretta 2019). In many languages, voiced stops have inherently shorter closure duration than voiceless stops. In timing-based explanations, vowels are compensatorily longer before voiced stops than before voiceless stops to maintain consistent timing of V+Closure units across laryngeal classes.

Three of the studies already mentioned in this chapter test the correlations between closure duration and vowel duration preceding stops with laryngeal specifications beyond voicing, though the data, methods, and results vary. Maddieson and Gandour (1976) test timing-based explanations in Hindi's 4-way contrast by measuring vowel + consonant durations. Because the stop releases in their data were not always clear, they measure the duration from the vowel to the release of the [k] in their ab [Sa:C] kaho' carrier phrase, assuming the closure of the [k] would be consistent across laryngeal classes of the target C. They find that aspirated stops are longer than their unaspirated counterparts, which is not the inverse of the vowel durations they observe; [t] has the shortest duration and the shortest preceding vowel.

Beguš (2017) does find an inverse relationship between vowel duration and closure duration in Georgian: vowel duration decreases from voiced > ejective > voiceless aspirated, while closure duration decreases from voiceless aspirated > ejective > voiced. This is consistent with a general trend of longer vowels before shorter closures, beyond just voicing contrasts.

Durvasula and Luo (2012) also find the mean closure durations for each class to be the inverse of the mean vowel durations by class. Vowel duration, in decreasing order, is Dh > D > Th > T. Closure durations decreased in the opposite order: T > Th > D > Dh. Voiced stops have longer closure durations than voiceless, and aspirated stops have longer closures than unaspirated. While finding this cross-category inverse relationship between vowel length and closure duration consistent with an isochrony account, Durvasula and Luo find evidence for a weak positive correlation between closure and vowel durations within each laryngeal class: within a given category, stops with longer closures tend to be preceded by longer vowels.

The present study similarly finds that closure duration across classes in Nepali runs opposite to vowel duration. A plot, parallel to Figure 1 of vowel duration, shows mean closure duration by both voicing and aspiration specification. Just as Durvasula and Luo found for Hindi, closure duration of Nepali voiced stops is longer than that of voiceless stops, and closure duration of aspirated stops is longer than that of unaspirated stops.

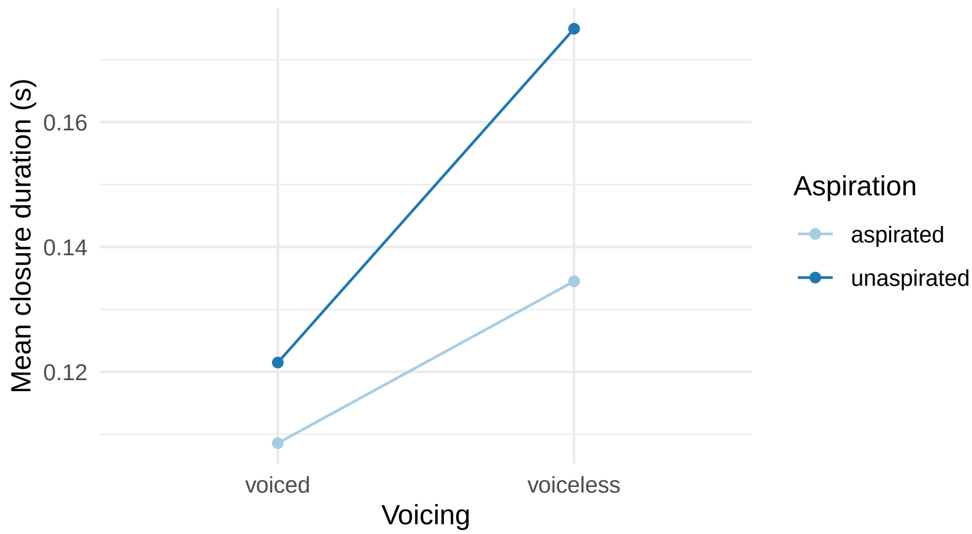


Figure 3.3: Mean closure durations for each of Nepali's four laryngeal classes of stops

Unlike Durvasula and Luo found for Hindi, Nepali does show an inverse relationship between vowel and closure durations within-class as well, as seen in Figure 4.

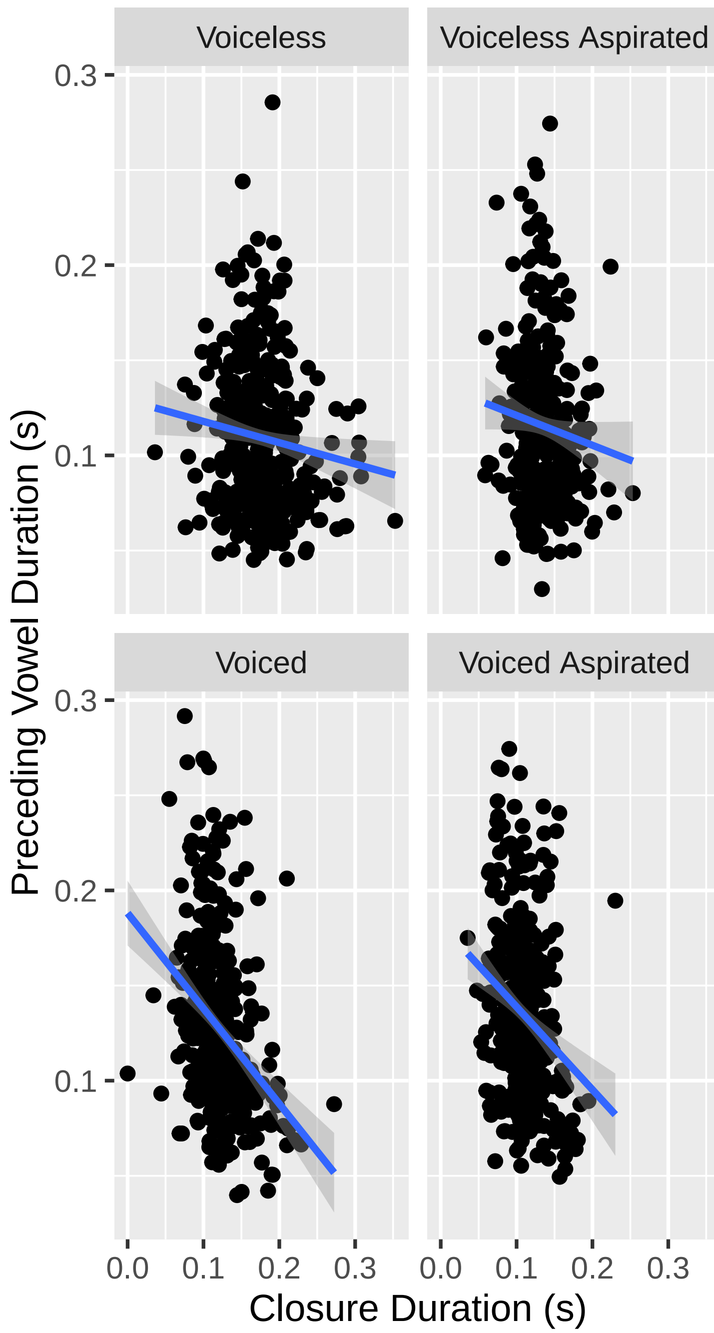


Figure 3.4: Within-category correlation between closure duration and preceding vowel duration

The significance of the inverse effect of closure duration on the length of the preceding vowel is confirmed by the model in Table 5. The model considers closure duration in addition to all previously considered predictors (voicing classification, aspiration classification, positive VOT, vowel quality, POA). Like the earlier models, all interactions were initially included, and then removed in a stepwise manner until reaching the optimal AIC. Nonsignificant interactions are not reported in the table.

Table 3.4: Linear mixed-effect model of log-transformed prevoicing duration (sec.) for D and Dh stops which showed prevoicing ($n = 212$).

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.152	0.008	17.211	< 0.001
Voicing (voiced)	0.016	0.002	5.99	< 0.001
Aspiration (aspirated)	0.008	0.004	1.805	0.071
Closure duration	-0.238	0.023	-9.522	<0.001
Positive VOT	-0.013	0.019	-0.652	0.515
Vowel ([a])	0.041	0.001	39.116	< 0.001
Vowel ([i])	-0.020	0.001	-19.871	<0.001
Voicing (voiced) : Closure Duration	0.081	0.013	-4.317	<0.001
Closure Duration : Aspiration (aspirated)	-0.069	0.036	-1.922	0.054
Voicing (voiced) : Vowel [a]	0.002	0.001	2.919	0.004

The model finds the inverse correlation between closure duration and vowel duration to be significant. Voicing classification remains a significant predictor of vowel length as well. While aspiration classification and positive VOT duration have the same positive effect on vowel duration seen in the previous models, neither effect reaches significance in this model.

Ultimately, speech units are isochronous if the units' duration remain consistent across laryngeal classes, compensating for the difference in consonant duration with vowel duration. If the source of the aspiration effect is the drive to maintain consistent duration across V+closure sequences that differ in aspiration specification, the Nepali data offers mixed support. Nepali stops yield similar durations for three of the four classes, as seen in Figure 5. Table 6 presents means and standards of deviation for vowel, closure, vowel+closure, and rhyme durations for the four stop classes.

Table 3.5: Mean durations and standards of deviation (in milliseconds) of potentially isochronous units across Nepali'd stop classes.

		vowel		closure		vowel + closure		rhyme	
	n	mean	sd	mean	sd	mean	sd	mean	sd
T	354	109	38	175	42	284	53	357	66
Th	287	114	42	134	29	249	49	354	59
D	361	127	45	120	33	247	46	308	52
Dh	349	135	46	108	27	243	46	317	56

In the voiced classes, the 8ms difference in vowel duration between Dh and D nearly compensates for the 12ms closure difference. In the voiceless classes, the 5ms difference in vowel duration between Th and T is not enough to compensate for the 40ms difference between these classes. If we consider the relevant speech unit to be the full rhyme, the vowel length in the voiced classes is not enough to overcome the closure for T and the release duration for the Th stops.

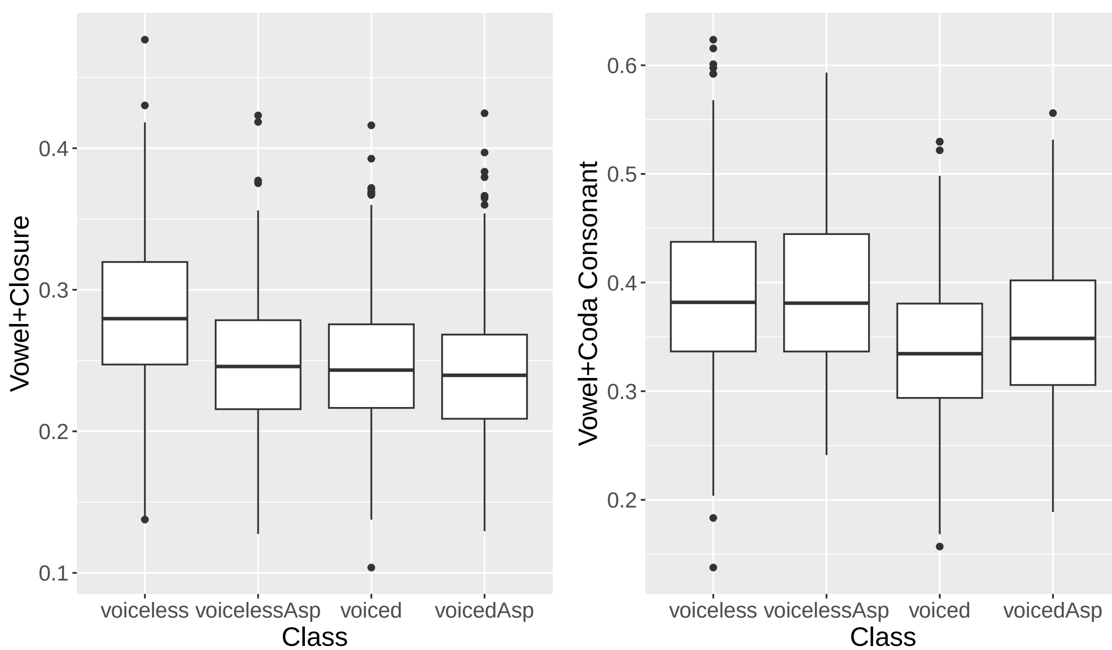


Figure 3.5: Durations of the vowel + closure of the following stop (left) and vowel + the entire following stop (right) across Nepali's stop classes

Perceptually-based explanations

Laryngeal specification, like any other phonological contrast, is realized with multiple phonetic cues. A perceptually-based explanation driven by the need to perceive the identity of the following consonant proposes that vowel duration is controlled by speakers to either enhance other perceptual cues to the laryngeal specification of the following stop, or to offset ambiguity of other deficient cues to laryngeal specification of final stops (Clayards 2017). Particularly in word-final position, when cues to laryngeal class like positive VOT are less perceptible and informative without a following vowel to host them, preceding vowel duration becomes a more reliable cue.

These cue-trading relationships are characterized by opposing between-category and within-category correlations between the two cues (Clayards 2017), precisely the pattern observed in the Nepali data. Across categories, aspirated classes are associated with longer vowels – a positive correlation between aspiration duration and vowel duration. Within categories, however, a cue-trading explanation predicts the opposite correlation. When a speaker produces a particularly ambiguous value of one cue, they might compensate by producing a more extreme value of the other cue. Within the aspirated classes, which have longer vowels overall, a token produced with short positive VOT is, by one type of cue, less perceptively distinct from the unaspirated stops. To compensate, a cue-trading explanation predicts that the speaker enhances other cues to the contrast, including making the vowel particularly long. A non-aspirated stop realized with abnormally long positive VOT might also be ambiguous, which the speaker could offset with a particularly short vowel. Thus, within-category, there is a correlation between longer VOT and shorter vowels, and shorter VOT and longer vowels – opposite to the across-category explanation. The coexistence of the aspiration effect and inverse VOT effect in Nepali, which in section 5.1 posed a challenge for an articulatory explanation of the aspiration effect, is precisely what is predicted by a perceptual, cue-trading account. While it would require further examination of the original data, several aspects of earlier studies are also potentially consistent with a perceptual account. Of all the earlier studies described in Section 2.2, Ohala and Ohala found the least evidence for an aspiration effect. They were also the only ones to analyze stops in non-coda position. The cues to aspiration are stronger in onset position, perhaps suggesting less perceptual need for the vowel duration cues. In Maddieson's (1977) study, two speakers contributed Marathi data. One participant spirantizes /p^h/ to [f], while the other does not. The participant who does not spiranize their stops displays a greater difference between /p/ and /p^h/ than the participant who does not spirantize. While we cannot make any claims based on the behavior of these two individual speakers, such a pattern could be consistent with a perceptual account if a [p]/[f] contrast is more perceptually salient without a vowel duration difference than the [p]/[p^h] contrast. Future work could test this hypothesis with a larger sample size.

3.5 Conclusions and future directions

This chapter engaged with the longstanding question of how and why laryngeal specification of a consonant affects the preceding vowel's duration. In order to assess a variety of explanations for the source of this effect, it examined the effects on vowel duration of not just voicing specification but aspiration specification as well. The results of a small number of previous studies had been somewhat mixed, but tend towards the existence of an aspiration effect: vowels are slightly longer before aspirated consonants than before unaspirated consonants. The empirical phonetic data of the present study contributes more evidence of an aspiration effect. Vowels preceding aspirated stops in Nepali are slightly longer than vowels preceding unaspirated stops.

This chapter also noted a potential tension between the studies that find evidence of a positive aspiration effect and Beguš's (2017) finding of an inverse VOT effect – that vowel duration correlates inversely with positive VOT duration of the following stop. The Nepali data considered in this study preliminarily suggest that these two effects are not necessarily in contradiction, but can coexist in a single language. This coexistence of seemingly contradictory effects is problematic only for exclusively articulatory explanations for an aspiration effect. It is consistent with perceptual explanations, and with analyses in which multiple factors and forces affect vowel duration.

While this production study may be consistent with a perceptual account of the aspiration effect, it cannot speak to whether vowel duration is in fact used by hearers to cue the aspiration contrast. To test this requires a future perception study.

Chapter 4

Laryngeal contrast weakening

4.1 Introduction

The previous chapters have dealt with the realization of contrast between different laryngeal classes, both by cues inherent to the stops themselves (voicing and aspiration durations in initial and medial positions), as well as by associated cues (preceding vowel duration) in cueing laryngeal contrasts. These chapters have argued that, in initial position, Nepali shows both a robust voicing contrast phonetically cued (primarily) by the presence vs absence of voicing during the stop closure and phonologically represented with a [voice] feature, and an aspiration contrast phonetically cued (primarily) by durational differences in post-release aspiration, and phonologically represented with a [spread] feature. The second chapter explores the extent to which each of these contrasts is enhanced or reinforced by the secondary cue of preceding vowel duration. This chapter deals with the weakening of these contrasts, investigating laryngeal neutralization in Nepali's stops word-finally. It has empirical, theoretical, and methodological aims.

Descriptions of Nepali report varying degrees of neutralization of laryngeal contrast in non-initial position, primarily loss of aspiration of the voiced aspirated Dh class. Khatiwada (2009, p. 376) describes how in spontaneous speech, voice aspirates lose aspiration in final (and inter-vocalic) position, though some speakers reportedly maintain aspiration in dentals, affricates, and velars. Nakkeerar (2011, p.21) reports more strongly for the variety of Nepali spoken in Sikkim in Northeast India that "the phonemic contrast of breathy (voiced aspirate) and non breathy (voiced) sounds is neutralized in the non-initial position" and that "if the consonants /gh, dzh, Dh, dh, bh/ do not occur in the word initial position, they are deaspirated". Following these descriptions, the emprical aims of the paper are to examine the robustness of acoustic cues to voicing and aspiration in final position through a production study, and to ascertain the degree of confusability of these contrasts in final position through a perception study. While both Khatiwada and Nakkeerar report contrast loss in non-initial positions (intervocalic as well as word-final), the studies here focus on final position. Crosslinguistically, final position displays higher degrees of laryngeal neutralization attributed to either prosodic position (Lombardi 1991, e.g.) or cue availability (Steriade 1997, e.g.), and much of the conversation surrounding laryngeal neutralization has centered on this position. Consistent with Khatiwada and Nakkeerar's reports, the studies I present in this chapter find evidence for variable loss of aspiration in Dh stops. The perception study further finds that in word-final position, stops for which the production target was Dh are often perceived as D. In addition to evidence for weakened contrast between the Dh and D classes, the production study also finds that T stops are often realized with long aspiration durations characteristic of Th stops, and that these T stops are often perceived as belonging to the Th class.

Methodologically, this chapter presents a perception study that leverages the variability present in naturally produced tokens to assess the relationships between cues and perception rather than creating manipulated stimuli as is typically done in similar perception studies (Ahmed and Agrawal 1969; Hussain and Nair 1995; Islam 2019). As elaborated further in the methods sections, this chapter presents a production study in which participants read words in a carrier phrase followed by a perception study in which a separate set of participants complete a forced choice task categorizing the same unadulterated tokens recorded in the production study. The control that is achieved in other studies by manipulating the stimuli is achieved statistically in this study in the subsequent statistical analysis, thus more closely mimicking a natural speech perception setting.

Theoretically, this study contributes to discussions of markedness and representation in laryngeal weakening in word-final representation. Laryngeal neutralization is the most extreme case of laryngeal weakening, in which laryngeal contrasts are phonologically lost and listeners can no longer reliably perceive the sound's underlying phonemic class. Previous accounts of Nepali vary in how complete the neutralization of any of the laryngeal contrasts is, and I do not argue for complete neutralization. Yet, phonetic differences between some of Nepali's laryngeal classes seem to be weakened in production and confused in perception in systematic ways. As will be elaborated on below, there is an extensive body of literature that discusses the bidirectional implications between the phonetic realizations of laryngeal contrasts word-finally and representation, asking questions about whether neutralization in languages with different types of laryngeal contrasts: is neutralization always devoicing? does neutralization always involves feature loss? is neutralization always to an unmarked class, and what is the identity of the unmarked class? Nepali, as a language with multiple laryngeal contrasts that differ in how they are weakened, provides a fruitful testing ground for exploring these questions.

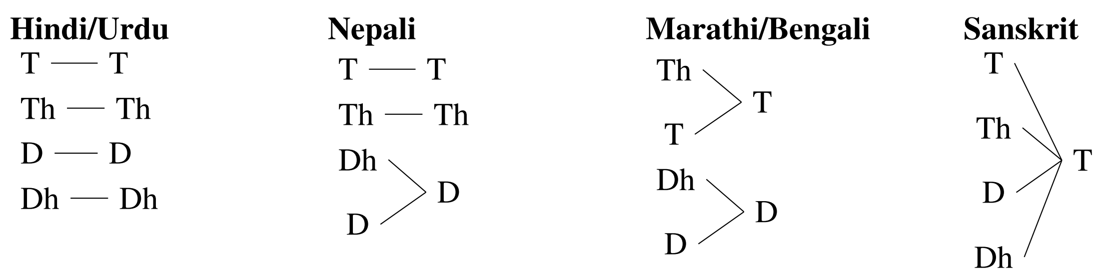
4.2 Background

Laryngeal neutralization and contrast weakening

Laryngeal contrast weakening across Indic languages

Within Indic languages with four-way laryngeal contrasts, a variety of (non)-neutralization patterns are attested. As described above, Nepali is reported to (partially) neutralize the aspiration contrast to D for only the voiced pair while maintaining the voicing contrast as well as the aspiration contrast in the voiceless stops. Elsewhere in the family, Marathi shows a loss of aspiration contrast to the unaspirated forms for both the voiced and voiceless pairs (Hussain and Nair 1995).

Chatterji (1926), Pattanayak (1966) and Lombardi (1991) report this same pattern for Bangla as well, though Islam (2019) finds through an extensive series of studies that both voicing and aspiration contrasts are maintained in Bangla. M. J. Kenstowicz (1994) and H. Dutta and M. Kenstowicz (2018) propose a dialectial difference in Bangla (which may explain the discrepancies between the Lombardi (1991) and Islam (2019) findings) in which some speakers or varieties maintain all the contrasts and some lose the aspiration contrast but keep the voicing contrast. Hindi/Urdu maintains a four-way contrast in final position (Ahmed and Agrawal 1969; J. J. Ohala 1983). Though, while confusion matrices by Ahmed and Agrawal (1969) show overall high rates of correct laryngeal class identification even in word-final position in Hindi/Urdu, they do show more confusability of the aspirated vs non-aspirated classes than the voiced vs. voiceless classes. Finally, Sanskrit shows neutralization of both the voicing and aspiration contrasts to a single T stop (I. Dutta 2007; Iverson and Salmons 2011).



Overall, the patterns attested in this set of Indic languages trend towards stability in the voicing contrast and variability in aspiration. Notably missing from this typology is a language in which the aspiration contrast is maintained while the voicing contrast is weakened or lost[1].

4.3 Production study

Methods: Production

Participants and stimuli

The acoustics of word-final stops were measured in the data collected for the study measuring vowel duration before stops of different classes reported in Chapter 3. 8 Nepali speakers living in Kathmandu participated in this study, and were recorded reading target words in carrier sentences. The target words were all of the templates sVC, pVC, and kVC in which V was /a, i, u/ and C was /p, p^h, b, b^h, t, t^h, d, d^h, k, k^h, g, g^h/ to yield 108 test items, some of which are nonce words and some of which are incidentally real words.

[1] Iverson and Salmons (2011) cite Dhasaanac, a Cushitic language of Ethiopia with voiced implosive stops specified for [voice] and [constricted] that reportedly lose the [voice] feature word finally but not the [constricted] feature (Tosco 2001)

Each word was written in Nepali's syllabic orthography (Devanagari script) on a separate cue card. The participants were shown each target word in an order randomized for vowel, initial consonant, and place of articulation of the target phoneme. Participants produced the word in the carrier sentence in (1), which is identical to the carrier sentence used in Chapters 2 and 3, and nearly identical to the carrier sentence used by Clements and Khatiwada (2007). Like the previous studies, I acknowledge that in recording read speech as opposed to spontaneous speech, it is possible that participants were influenced by the spelling and thus pronounce the sounds more 'correctly' than they would have in natural speech. However, as will be discussed in section 4.4, these recorded utterances constitute the stimuli for the perception experiment, and are more naturalistic than the sorts of stimuli typically used in perception experiments.

(1) X_1 (pause). mʌ ʌbʌ X_2 b^hants^{h}u: (pause) X_3
'X_1 (pause) now I say X_2: (pause) X_3.'

The results presented here are from the X_1 and X_3 positions of all 8 speakers, because each of these positions is preceded and followed by a pause, giving roughly comparable prosodic contexts. This yielded 1672 total tokens.

Acoustic analysis

The recordings were imported into Praat (Boersma and Weenink 2015), where they were hand-annotated for voicing, closure, and post-release duration measurements. Voicing duration was measured from the beginning of the closure (marked by a sharp decrease in amplitude in the wave-form signalling the end of the preceding vowel) to the end of visible periodicity in the waveform. Closure duration ended at the release burst if present, or at the cessation of voicing in unreleased voiced stops. Post-release duration was measured from burst release to the cessation of aperiodic noise in the spectrogram characteristic of aspiration.

Results: Production

Before turning to the primary research question regarding the overlap of realizations of D and Dh stops, I will address the prevalence of spirantization of word-final stops. Consistent with existing descriptions of Nepali (Khatiwada 2009), stops were spirantized in both studies. In total, 199 of the 1672 tokens were spirantized. Distributions across laryngeal class and place of articulation are shown in Figure 4.1.

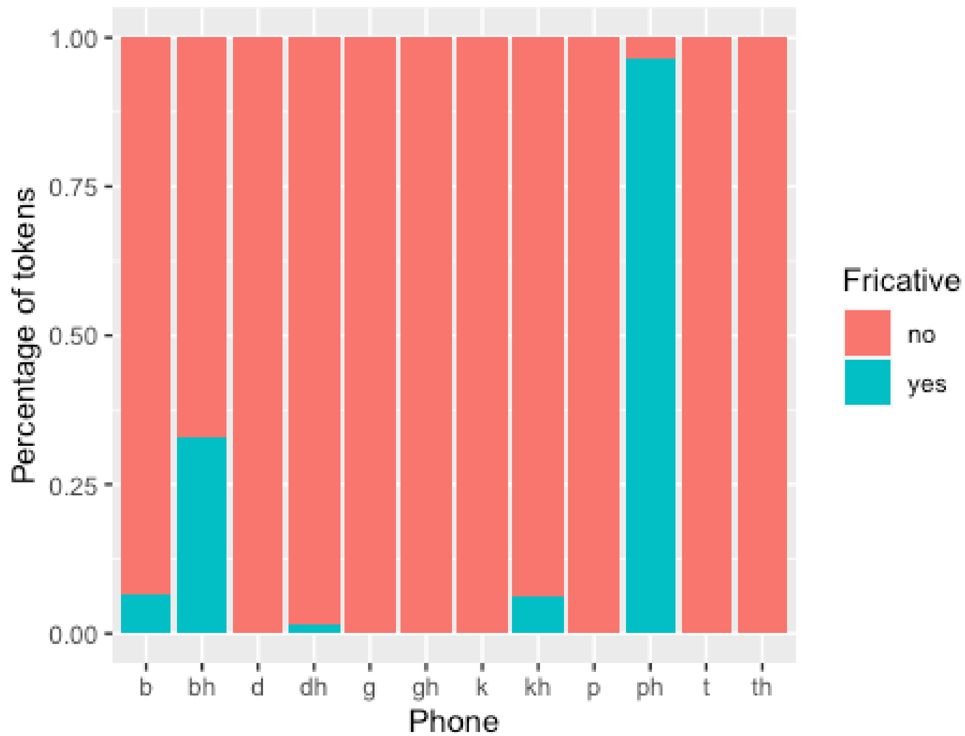


Figure 4.1: Percentage of stop tokens of each type realized as fricatives, across all speakers.

Nearly all the /p^h/ tokens are spirantized. This is consistent with a $p^h >$ sound change that appears to be in progress in Nepali. Voiceless aspirated bilabials are spirantized most in syllable- or word-final position as seen here, but are also spirantized elsewhere as seen in the Chapter XX study of word-initial stops. The phonemes that spirantize tend to be aspirated and/or bilabial, though not exclusively: besides /p^h/, we see the most spirantization in /b^h/, /p/, and /k^h/ tokens, but a small number of unaspirated stops including /d/ and /g/ spirantize as well.

The study includes enough tokens per speaker to allow us to examine by-speaker variation in spirantization frequency of non-/p^h/ stops. Rather than consistently low rates of /b^h/ spirantization across all speakers, we find that the vast majority of spirantized tokens are accounted for by 4 speakers (A, D, F, H) who spirantize /b^h/ approximately 50% of the time. The spirantized /k^h/ tokens are similarly all from participant E who produced [x] approximately 60% of the time.

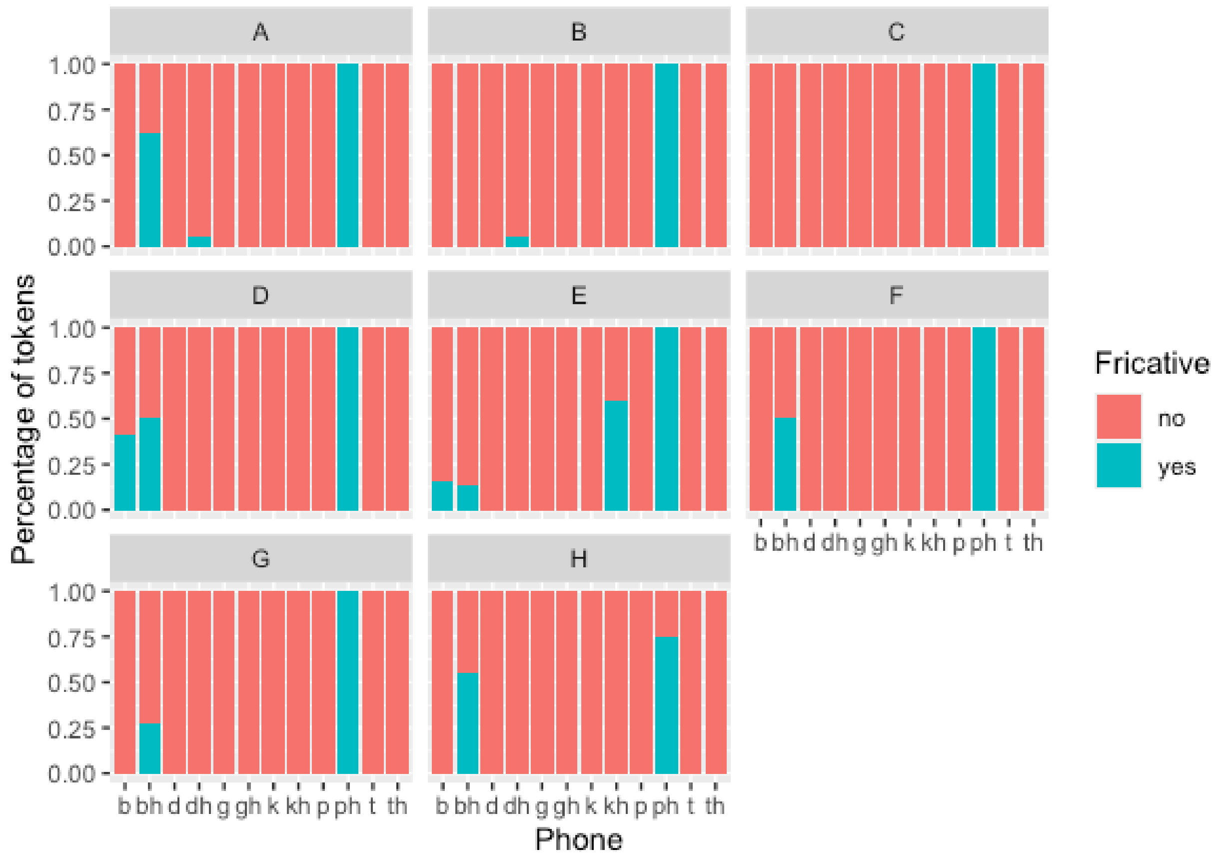


Figure 4.2: Percentage of stop tokens of each type realized as fricatives, by speaker.

These spirantized tokens will be excluded from the analysis of durational cues that differentiate the acoustics of the laryngeal classes since they lack comparable closure and aspiration durations, but we will deal extensively with spirantization in the context of contrast perceptibility in sections 4.4 and 4.5.

To determine the extent to which the productions of the various laryngeal classes overlap in final position, we will focus on two durational measures crucial to cueing the contrasts in non-spirantized stops: aspiration duration as a key cue to the aspiration contrast, and percentage of closure voicing as a key cue to the voicing contrast. In terms of the voicing contrast, we find largely non-overlapping voicing during closure proportions between the voiced and voiceless classes, as seen in Figure 4.3. Voiceless classes show voicing through 25% or less of the closure, while most of the voiced tokens have voicing through 75-100% of the closure.

The study found evidence for some amount of overlap in aspiration durations across aspirated and unaspirated stops in both the voiced and voiceless pairs in word-final position: Most D stops have an aspiration duration of 50ms or less with distributional peak around 15ms, though some display more aspiration. The distribution of Dh aspiration is more variable and appears somewhat

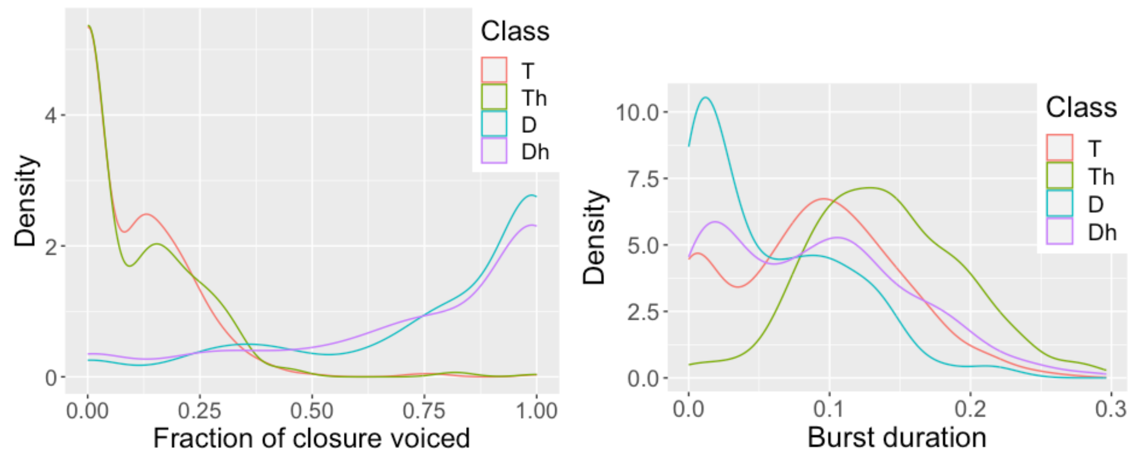


Figure 4.3: Distribution of the fraction of the closure that is voiced (left) and burst duration (right) in final position, across stop classes.

bimodal with a peak at 25ms and a second peak at 100ms, suggesting that perhaps some of the tokens are neutralized with D tokens in terms of aspiration duration, while others remain distinct and contrastive. Many of the underlyingly Dh stops fall squarely within the <50ms range characteristic of underlyingly unaspirated stops, while the remainder have durations of 50ms-200ms. In the voiceless classes, the Th stops show consistently long aspiration durations while the distribution of T stops is more spread out. Like Dh, this class also displays some bimodality, with one distributional peak (of mostly unreleased stops) at 0-30ms, and another at 100ms. Many of the T stops overlap with the Th distribution.

Like with the spirantization, some of this variation may be attributed to by-speaker variation. In the by-speaker plots below, we see that a couple of the speakers, (B and D), appear to pretty consistently realize Dh and D with the same burst durations. Other speakers (F, A) seem to have overlapping D and Dh distributions as well, though not in the prototypical D range. For the T/Th contrast the overlap of the distributions appears fairly consistent across all the speakers.

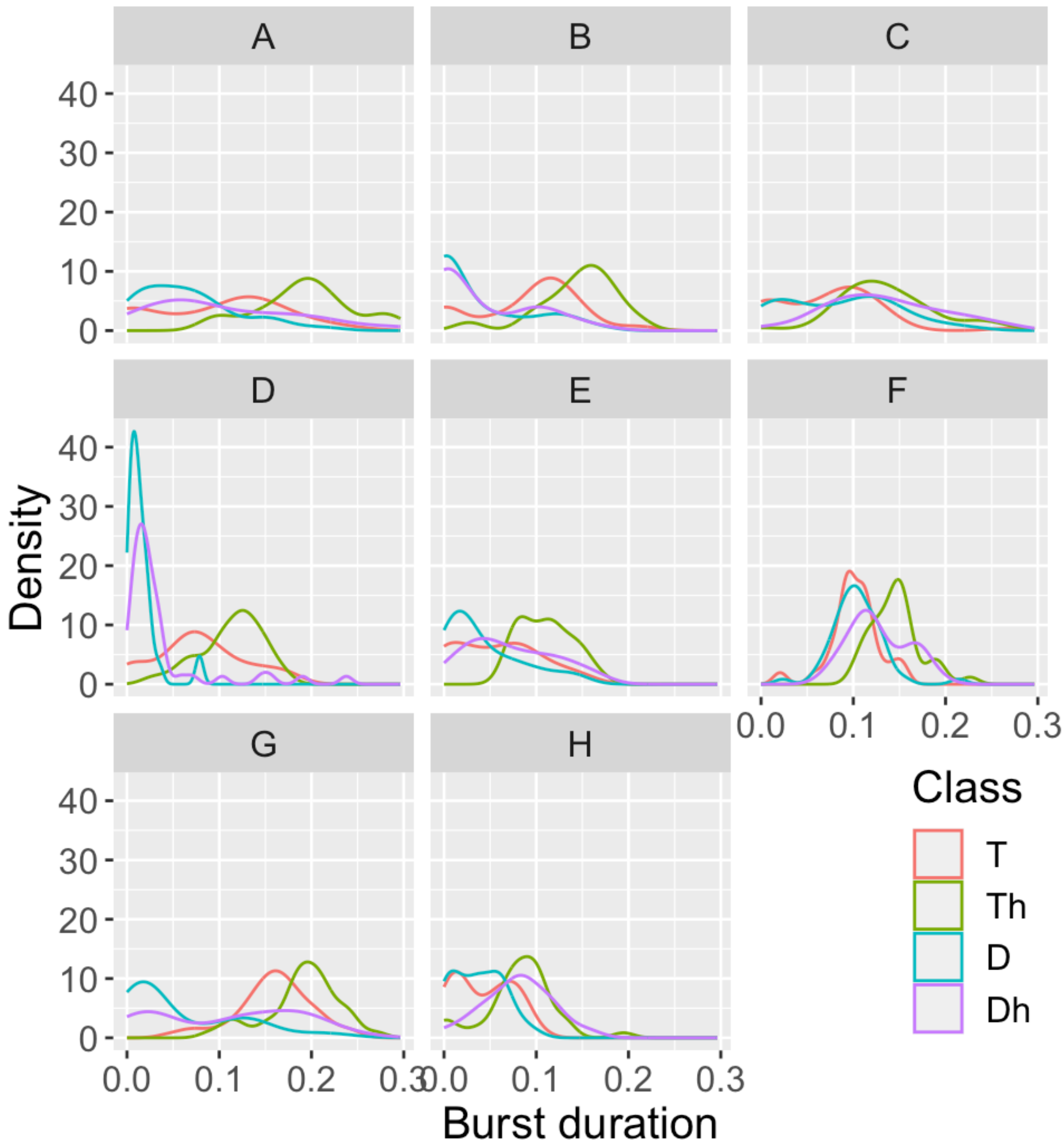


Figure 4.4: Percentage of stop tokens of each type realized as fricatives, by speaker.

We conducted the production study to test reports of neutralization of the Dh/D contrast word-finally. Due to the overlap of the aspiration duration distributions of the stop classes, we find evidence that suggesting some amount of neutralization of the aspiration duration cue in word-final position for the D/Dh contrast. We also see potential weakening of the T/Th contrast towards the Th, contrary to what has been previously reported. While the acoustics show that burst duration may not reliably distinguish the aspirated vs unaspirated classes in final position, ultimately a perception study is necessary to determine whether any contrasts are (variably) lost.

4.4 Perception study

Methods: Perception

12 Nepali-speaking participants contributed to an online perception study in the spring of 2023 examining the perceptibility of laryngeal contrasts in word-final position. In the perception study, participants were presented with audio stimuli spliced directly from the recordings from the 2018 acoustic study. Isolated target words, including those with spirantized segments, were extracted from the carrier phrases uttered by four of the participants in the production study and presented to the participants in the perception study. In each trial, participants heard a single word and were presented with a four-way forced choice discrimination task. The response options were the minimal quartet based on the written grapheme that the speaker in the production study was presented with, differing only in the laryngeal class of the final consonant. The response choices for the trials featuring a recording of the production stimulus /pat/ were [pat], [pat^h], [pad], and [pad^h], presented in Devnagari script. Participants self-paced through the study, but were alerted if their response for an audio stimulus was too slow (>4sec) or too fast (<300ms). This study was distributed online to Nepali speakers living in contexts with a matrix language of English who all reported regular use of and comfort in Nepali. The entire experiment was presented in Nepali written in Devnagari. 12 participants yielded 10,788 total tokens.

Results: Perception

Overall, 52% of the tokens were identified as the same as the grapheme the talker was presented with. As a forced-choice identification task between four items, chance is at 25%. Figure 4.5 presents a full matrix comparing the identity of the segment as presented in writing in the production study with the chosen segment in the perception study. Because participants in the perception study were only given the choice between four phones with the same place of articulation, this experiment only tests confusion between laryngeal class, not between places of articulation.

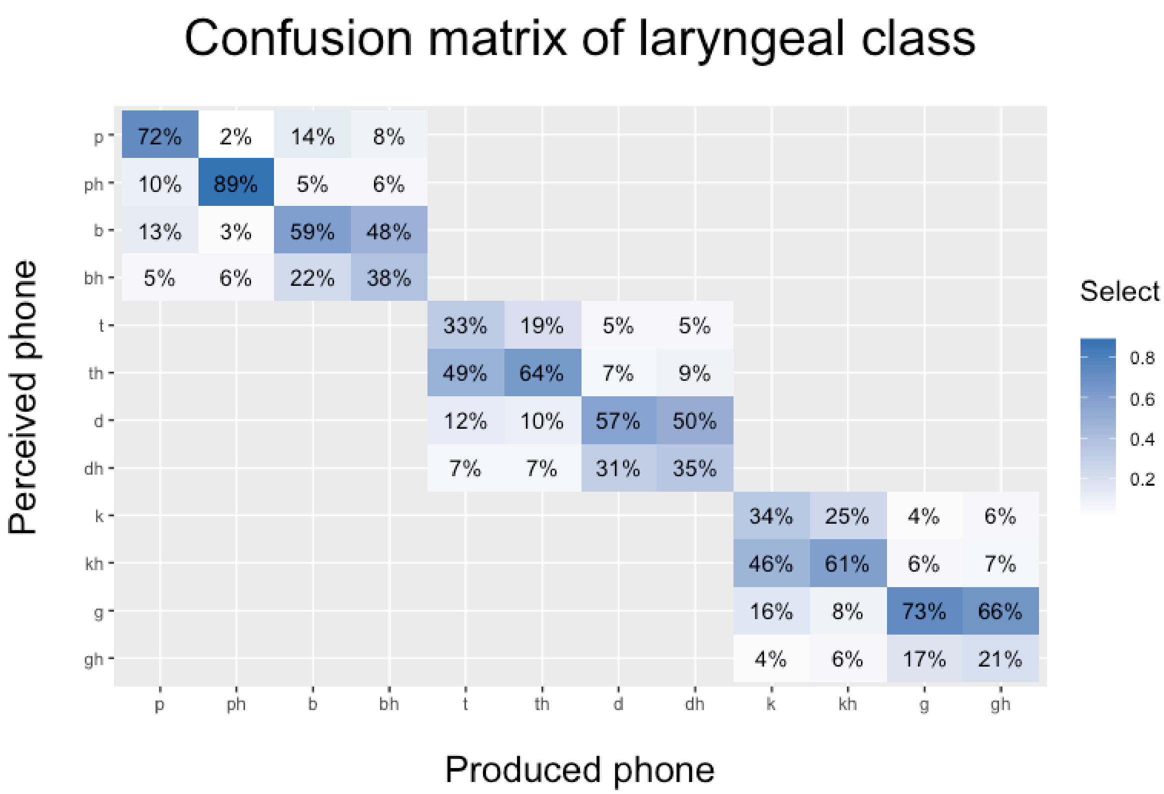


Figure 4.5: Confusion matrix showing the perceived laryngeal class (perceived phone) of stimuli based on graphemes of each type (produced phone).

Throughout these data, the aspiration contrast is much more often confused than the voicing contrast. One important exception is that /p^h/ and /p/ are rarely confused, and I will return to this below. Excluding these two phones, graphemes are identified as a phoneme with a different aspiration specification XX percent of the time, and as a phone with a different voicing specification XX percent of the time. While misidentification of aspiration is more common throughout the data, the direction of the misidentification is opposite within voiced vs. voiceless classes. Within voiceless stops, it is more common for unaspirated stops to be misidentified as aspirated than for aspirated stops to be misidentified as unaspirated: /p/ is perceived as [p^h] more than /p^h/ is heard as [p], /t/ as [t^h] more than /t^h/ as [t], and /k/ is perceived as [k^h] more often than /k^h/ is perceived as [k]. Within the voiced stops, however, the direction of the asymmetry switches: it is more likely for aspirated stops to be misidentified as unaspirated than for the unaspirated stops to be misidentified as aspirated. This pattern is at its most extreme with <g^h>, which is perceived as unaspirated 66% of the time and perceived as aspirated only 21% of the time.

In the next sections, I consider the factors that may influence whether a given token is perceived as the same grapheme that the talker was presented with to produce.

Lexical Status

One possible factor is the lexical status of the stimulus. Recall that all of the stimuli follow the templates sVC, pVC, and kVC in which V was /a, i, u/ and C was /p, p^h, b, b^h, t, t^h, d, d^h, k, k^h, g, g^h/. These templates yielded 108 different stimuli types, some of which are nonce words and some of which are incidentally real words. It is possible that when presented with a set of four words in the perception study, lexical bias may influence the listener to perceive and/or choose a real word more often than a nonce word, meaning that real words would be more often identified correctly than nonce words. Indeed, real words were correctly identified slightly more often than nonce words, as shown in 4.1.

Table 4.1: Percentage of tokens correctly identified, for real words vs. nonce words

	number of tokens	% correct
Real word	2501	59%
Nonce word	8287	50%

Spirantization

The disproportionate accuracy with which /p^h/ tokens were perceived as [p^h] suggests that spirantization is an important factor in categorization, because Nepali is undergoing a sound change in which /p^h/ is often realized as a fricative ([f] or [ɸ]). Phonologically, spirantization occurs more in non-initial positions, though it is extremely common in initial position as well (see Chapter 2). The sociolinguistic factors that influence adoption of the sound change are beyond the scope of this paper, but are an important area for future research. In this study, /p^h/ tokens were realized as fricatives in final position 100% of the time. While the bilabial voiceless aspirated stops are the most frequently spirantized, figure 4.1 showed that other stops spirantize as well, particularly /b^h/ and /d^h/. While the number of tokens is much smaller than for the voiceless bilabial, correct identification of these graphemes was higher when the tokens were spirantized than when they were not: The 528 /b^h/ tokens realized as stops were identified as [b^h] rather than [b] 33% of the time, whereas the 210 /b^h/ tokens realized as fricatives were correctly identified as [b^h] rather than [b] 70% of the time. Similarly (though with even smaller numbers), the 782 /d^h/ tokens realized as stops were identified as [d^h] rather than [d] 40% of the time, whereas the 19 /d^h/ tokens realized as fricatives were identified as [d^h] rather than [d] 73% of the time. In sum, spirantization seems to strongly influence identification as the aspirated grapheme. Section 4.5 will discuss this phenomenon further, proposing spirantization as a strategy for contrast enhancement.

Durational acoustic cues

A primary purpose of perception studies is to determine which acoustic cues influence perception and categorization. Often, this is achieved by manipulating stimuli to create tokens that with varying values for the cues under investigation. In the present study, this variability was achieved naturally: As seen in Figures 4.3 and 4.4 show a lot of variability in the aspiration (and to a lesser extent voicing) durations across the productions of stop tokens of the same class. It is predicted that Dh tokens with shorter aspiration will be classified more often as D than Dh tokens with longer aspiration, and that T tokens with longer aspiration durations will be classified more often as Th than T tokens with shorter aspiration. Figure 4.6 shows perceived class as a function of this aspiration duration measure, visualizing the extent to which aspiration duration correlates with class categorization in perception. A few trends emerge, visually, from this plot. First, many of the tokens realized with very little aspiration are classified as the D class. Second, most of the tokens at the highest aspiration duration range are classified as Th. Third, most of stops perceived as T class have more than 50ms of aspiration, and while many of the stops perceived as the Dh class have aspiration durations between 75-150ms, many also have aspiration durations of less than 30ms.

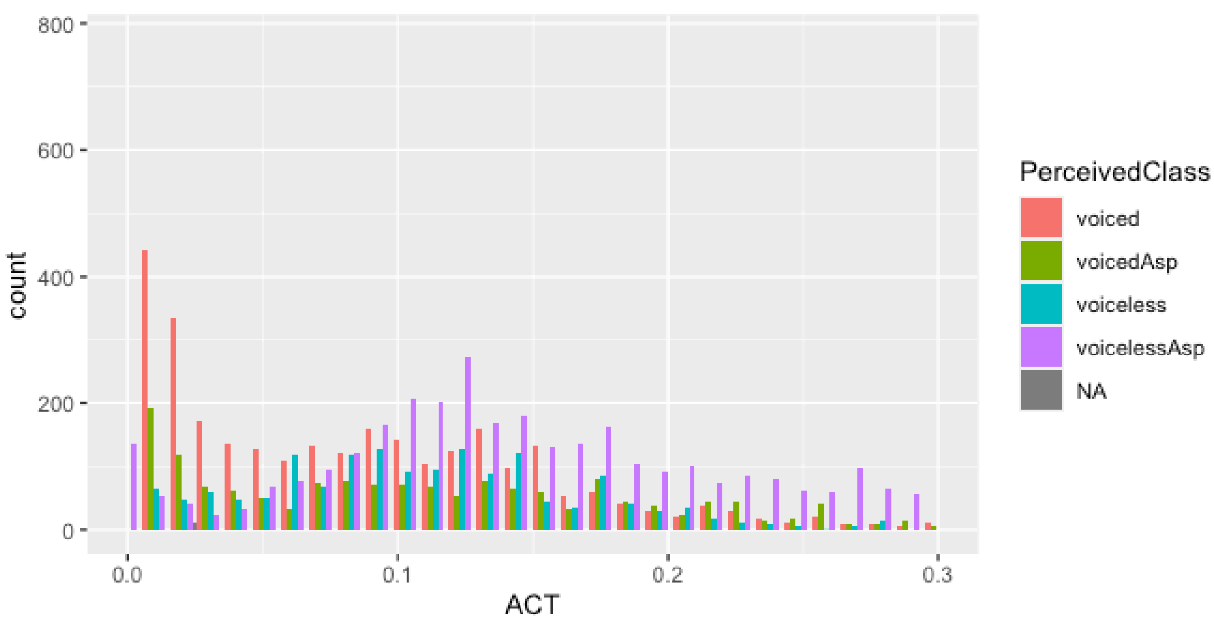


Figure 4.6: Number of tokens perceived as each class, as a function of aspiration duration.

Two logistic regression mixed effects models were fitted to test whether durational acoustic cues such as voicing duration and aspiration duration, as well as other factors, significantly influence the categorization of the token by the listener. Because the voicing contrast word-finally is rarely confused in Nepali, the models assess only on the aspiration contrast. The first model, presented in 4.2 considers only the voiceless subset of the data (only the stimuli corresponding to voiceless graphemes), and predicts the aspiration specification of the perceived class. It includes as predictors aspiration duration and voicing duration of the stimulus (both scaled), place of articulation of the coda consonant, the identity of the preceding vowel ([a], [i], [u]), and the interaction between aspiration duration and POA. It also considers random intercepts for both reader and listener. Despite the possibility that lexical status of the stimulus has an effect on how often the stimulus was identified correctly, this model does not consider lexical status. Because the model predicts the frequency with which tokens are identified as aspirated, including lexical status would not tell us how often real words are identified correctly, but rather how often real words are identified as aspirated, which is not relevant to the research question. Laryngeal class of the grapheme is not included as a predictor either, since listeners do not have access to that information. Because all of the voiceless aspirated bilabials are spirantized, all bilabials are excluded from the model of the voiceless subset of the data. Both models were fitted using the lme4 package in R (Bates et al. 2015). Fixed effect coefficients are shown with associated standard errors test statistic (z) and significance, calculated with lmerTest (Kuznetsova et al. 2015) using the Satterthwaite approximation. Random effects are not shown.

Table 4.2: Generalized linear mixed-effect model predicting the aspiration specification of the perceived class, for only stimuli produced from voiceless graphemes. Bilabials are excluded ($n = 3,547$).

	Estimate	Std. Error	z value	Pr (>|z|)
(Intercept)	0.940	0.260	3.610	<0.001
POA (velar)	-0.187	0.076	-2.452	0.0142
ACT (scaled)	0.463	0.059	7.868	<0.001
Vowel [i]	-0.354	0.096	-3.694	<0.001
Vowel [u]	0.096	0.099	0.964	0.335
Voicing Duration (scaled)	0.072	0.044	1.663	0.096
POA (velar): ACT	-0.087	0.079	-1.107	0.268

The resulting model shows that aspiration duration does have a significant positive effect. Tokens that have a longer aspiration are more likely to be perceived and classified as aspirated phonemes. POA also has a significant effect: velar stops are somewhat less likely to be perceived as aspirated than alveolar stops. Velar stops are also slightly less likely to be perceived as aspirated at higher ACT values than alveolar stops, though this does not reach significance. Stops following [i] are significantly less likely to be perceived as aspirated than stops following [a], but there is no effect for [u]. Finally, as voicing duration increases, tokens are slightly (but not significantly) more likely to be perceived as aspirated than unaspirated.

The second model, presented in 4.3, considers just the data corresponding to the stimuli from voiced graphemes. It includes the same fixed and random effects as the voiceless model. Unlike the voiceless model, this model does include the bilabials because a far smaller percentage of voiced aspirated bilabial stops are spirantized. POA is therefore a three-level factor in this model.

Table 4.3: Generalized linear mixed-effect model predicting the aspiration specification of the perceived class, for only stimuli produced from voiced graphemes. All places of articulation are included ($n = 5,293$).

	Estimate	Std. Error	z value	Pr (>|z|)
(Intercept)	-0.368	0.336	-1.094	0.274
POA (velar)	-0.704	0.080	-8.753	<0.001
POA (alveolar)	0.088	0.762	1.155	0.248
ACT (scaled)	0.077	0.035	2.220	0.026
Vowel [i]	0.171	0.077	2.221	0.026
Vowel [u]	0.192	0.077	2.481	0.131
Voicing Duration (scaled)	0.082	0.042	1.930	0.054
POA (velar): ACT	-0.803	0.168	4.782	<0.001
POA (alveolar): ACT	0.567	0.177	3.201	0.001

As in the voiceless model, ACT has a significant positive effect. Tokens with longer aspiration are more likely to be classified as Dh than as D. Voicing duration has a slightly but not significantly positive effect as well: tokens with longer voicing durations are slightly more likely to be perceived as aspirated. Velar tokens are classified as aspirated less than bilabials, which could again be due to listeners' unmet expectations of long aspiration for velar stops. Opposite to the voiceless model, stops preceded by [i] and [u] vowels are significantly more likely to be classified as aspirated than stops preceded by [a] vowels.

Summary of results

Both the production study and the perception study found evidence consistent with Khatiwada (2009) and Nakkeerar (2011)'s reports that the aspiration contrast between Dh and D is weakened or variably realized as D in final position. In the production study, Dh stops were realized with widely varying ACT durations, showing a somewhat bimodal distribution in which many of the Dh stops had less than 50ms of ACT and overlapped with the prototypical D productions. In the perception study, stops that were produced from voiced aspirated graphemes were perceived as plain voiced stops on average 54% of the time and as voiced aspirated an average of 31% of the time. Furthermore, analysis of the factors influencing listeners' categorization of the stops found that aspiration duration is a significant factor, suggesting that realization with less aspiration does indeed weaken the contrast.

Both studies also found evidence of a weakening of the T/Th contrast in final position, which has not been reported in previous studies of Nepali. Opposite to the pattern in voiced stops, T stops are often realized with aspiration durations typical of Th stops, and are perceived as aspirated by listeners. Further work should examine the realization of Nepali laryngeal contrasts in more spontaneous speech. The laboratory setting and read-speech context of the present study likely influenced production. Participants may have maintained the contrasts more than they would in naturalistic speech (Khatiwada 2009). Participants also may have exaggerated the release of the stops, which would particularly affect the results for the T/Th contrast.

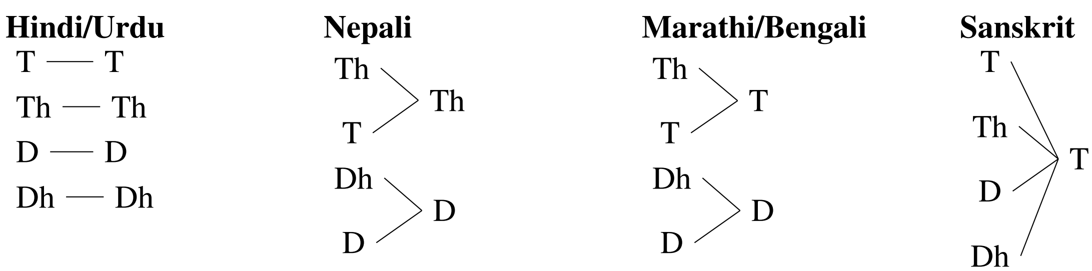
4.5 Discussion

Asymmetries in contrast weakening

The patterns of contrast weakening in Nepali are characterized by two asymmetries. First, the aspiration contrast weakens while the voicing contrast remains robust. The second asymmetry is in the direction of the aspiration contrast's weakening: the voiceless T and Th classes display weakening towards the aspirated Th class, while the voiced D and Dh classes display weakening towards the unaspirated D class. I propose that the first asymmetry can be explained by cue availability, and the second by markedness. The important exceptions to these patterns of weakening and confusion are the spirantized tokens, which will be discussed extensively below. This section discusses the non-spirantized stops.

Asymmetry #1: Weakening of the aspiration contrast independent of the voicing contrast

The first asymmetry, in which the aspiration contrast weakens while the voicing contrast remains robust, is consistent with the typology of weakening patterns in Indic languages presented in Section 4.2, and repeated here with an updated pattern for Nepali based on the present studies.



Like H. Dutta and M. Kenstowicz (2018) do for Assamese, which also maintains a robust voicing contrast while shifting the aspiration contrast in its four-way contrast, I turn to Steriade's 1997 theory of Licensing by Cue to explain Nepali's pattern and this Indic typology more broadly. In a Licensing by Cue framework, contrasts are neutralized in positions where the cues to realize the contrasts are unavailable or compromised. In word-final position, the absence of a following vowel eliminates important cues to the aspiration contrast including (positive) VOT and formant values at the onset of the following vowel. Primary cues to Nepali's voicing contrast, however, including voicing during closure and duration of the preceding vowel (shown in Chapter 3 to differ significantly between voiced and voiceless classes) do not rely on following vowel and are therefore still available in word-final position. Particularly because the vast majority of the tokens (at least in this study of read speech) are released and the closure duration is therefore perceptible, the voicing contrast is somewhat protected. Steriade proposes the implicational universal that languages that neutralize a contrast in a given position will also neutralize that contrast in all positions with a less informative cue package. I propose that languages that neutralize a contrast with a more informative cue package in a given position will also neutralize a contrast in that same position with a less informative cue package. An Indic language that contradicts this hypothesis would show neutralization of the voicing contrast independent of the aspiration contrast, which the literature has not yet attested.

Recall from Chapter 2, the Q Theoretic representation of Nepali's stops:

(2) With features:

T	(t [ ]	t [ ]	t) [ ]
D	(d [voice]	d [voice]	d) [voice]
Th	(t [ ]	t [ ]	h) [spread]
Dh	(d [voice]	d [voice]	h) [spread]

By linearly ordering the laryngeal features to voicing and aspiration, a Q Theoretic representation matches the phonetics of the cues to laryngeal contrast, captures Nepali's asymmetry in weakening the aspiration contrast independent of the voicing contrast, and predicts a typology in which voicing contrasts do not neutralize independent of aspiration contrasts.

Asymmetry #2: Direction of aspiration contrast weakening

I propose that Nepali's second asymmetry, in which the voiceless classes weaken in the direction of the aspirated stop while the voiced classes weaken in the direction of the unaspirated stop, is driven by markedness. Keeping with foundational tenets of laryngeal neutralization, I propose that contrast weakening in Nepali is always towards the unmarked class. However, the marked status of [spread] differs between the voiceless and voiced pairs. Th is the unmarked class in the T/Th contrast, while D is the unmarked class in the voiced contrast. This proposal draws from insights that markedness is not universal (e.g. Rice 1999; Vaux and Samuels 2005; Broselow 2018) but rather that preferred laryngeal specification varies across languages depending on inventories, behaviors, and realizations. In true voicing languages that contrast T and D, the outcome of voicing neutralization is typically T. In aspirating languages that contrast T with Th, the outcome of aspiration neutralization may be Th rather than T (Westbury and P. A. Keating 1986; Vaux and Samuels 2005). Nepali joins the set of languages in which Th is less marked than T, but simultaneously treats D as less marked than Dh. The markedness of Dh stops is supported by their rarity crosslinguistically compared to plain voiced stops, and by diachronic patterns. While Dh stops have been lost in a number of Indic languages including Punjabi, Kashmari, Shina, and Romani, there are no Indic languages that lack plain voiced stops (Masica 1993).

While beyond the scope of this dissertation, an Optimality Theory analysis capturing the typology of Indic neutralization patterns should be pursued in future work. The relative ranking of the markedness constraints would be fixed according to the viability of each contrast in word-final position, and faithfulness constraints to the underlying contrasts would be ranked relative to these markedness constraints. In a language like Sanskrit, markedness of both aspiration and voicing would outrank faithfulness. In Hindi/Urdu, faithfulness would outrank all markedness. In Marathi, faithfulness would be ranked lower than aspiration contrast markedness but above voicing contrast markedness. Nepali motivates a markedness constraint specific to Dh to be ranked higher than the general constraint against aspiration contrasts word-finally.

Spirantization as contrast enhancement

The exceptions to these patterns of confusion and contrast weakening are the spirantized tokens. <p^h> tokens are realized as fricatives 100% of the time in this study. These spirantized <p^h> tokens are almost never (2% of the time) perceived as [p]. Moreover, (non-spirantized) <p> tokens are perceived as [b] more often than they are perceived as [p^h], unlike other places of articulation. Together, this suggests that not only do the speakers who participated in this study reliably demonstrate the /p^h/ > /ɸ sound change, but that listeners *expect* </p^h/> to be spirantized and are much less likely to categorize a (non-spirantized) <p> as [p^h], even with a long ACT duration, than they are for other places of articulation. Figure 4.7 shows the number of tokens produced from T graphemes categorized as each class, according to the ACT duration, at each place of articulation. We see that while many alveolar and velar tokens with longer ACT durations are classified as Th, very few bilabial tokens are classified as Th even with ACTs of 100ms or more.

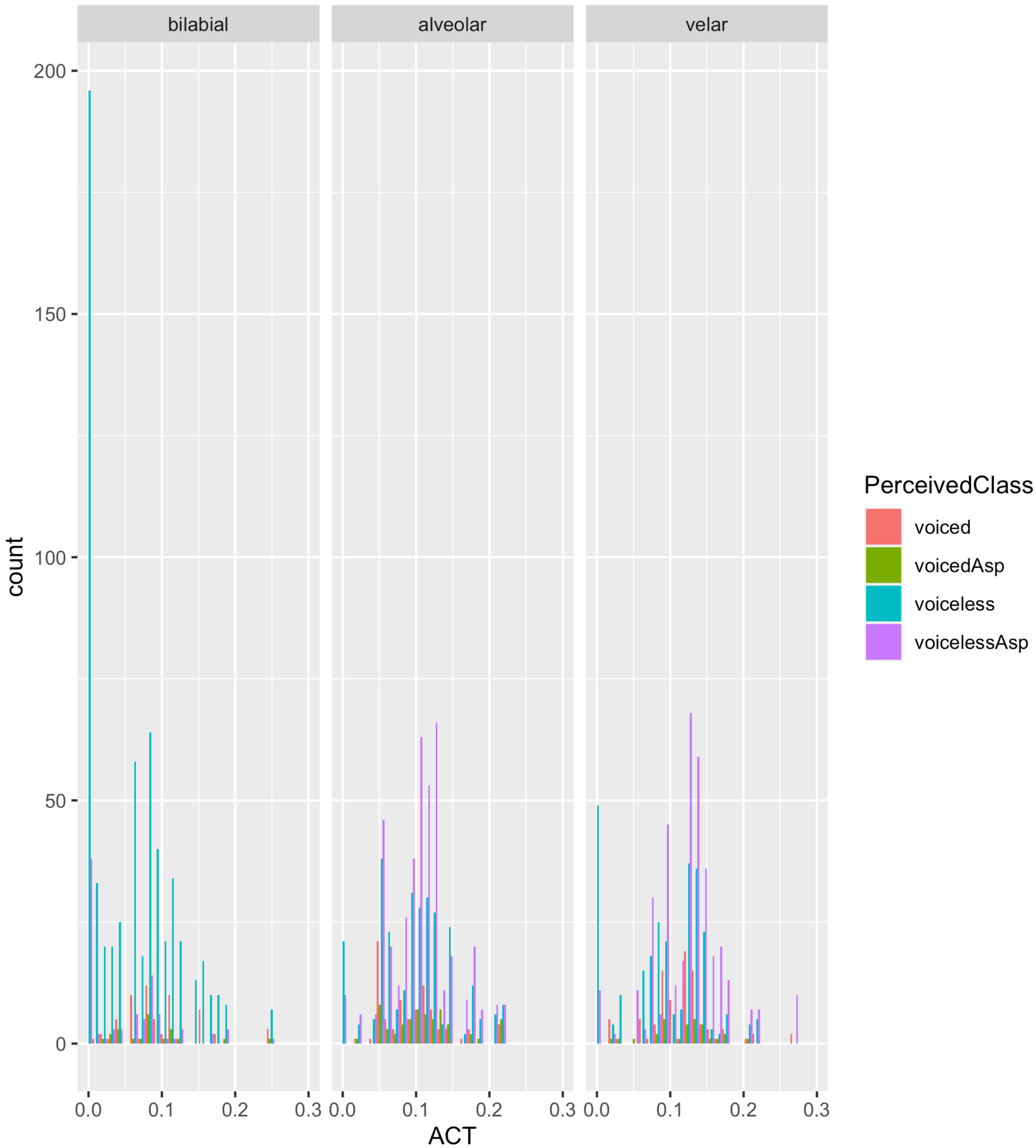


Figure 4.7: Number of tokens produced from T graphemes perceived as each class, as a function of aspiration duration, for each place of articulation.

The contrast between /p/ and /p^h/ may therefore be best described currently (at least for the participants in these production and perception studies) as a difference in continuance ([cont])rather than a difference in [spread] feature alone. Moreover, the high rate of identification of the /p/ vs. /p^h/ contrast evident in the confusion matrix suggests that this [cont] contrast may be more perceptible than the [spread] contrast, and that spirantization may be a tactic for contrast enhancement. This shift from a contrast solely cued by laryngeal features to a contrast cued by continuance makes functional sense. Like all languages with four-way laryngeal contrasts, Nepali's stop and affricate inventory is extremely crowded (20 stops and affricates at four places of articulation), while its fricative inventory is quite small (just /s/ and /h/). Shifting from /p^h/ to /ɸ/ relieves some of the pressure on the stop inventory.

4.6 Conclusions

The production and perception studies presented in this chapter provide phonetic support for previous descriptions of laryngeal contrast weakening in non-initial position in Nepali. While the voicing contrast remains robust in both production and perception, the aspiration contrast is weakened. Dh stops are often realized without much aspiration, and are often perceived as D stops. T stops are often realized with as much aspiration as Th stops, and are often perceived as Th stops. These patterns further support the Q Theoretic feature representation in which laryngeal features are distributed across the subsegments of each stop, corresponding to the temporal ordering of the cues and the greater vulnerability of the [spread] contrast than the [voice] contrast in the absence of a following vowel. Another major finding of the study is that /p^h/ is consistently realized as [ɸ], and /p/ and /p^h/ are very rarely confused. While further research should investigate the status and spread of the p > ɸ sound change across regions and speakers of Nepali, the participants in these studies appear to both produce and rely on spirantization as the primary cue to this phoneme.

Chapter 5

Conclusions

5.1 Summary

This dissertation has investigated three core areas related to laryngeal contrasts in light of Nepali's four-way laryngeal contrast: laryngeal realist representations, the voicing effect, and laryngeal neutralization. Much of the debate surrounding these topics centers on the extent to which phonological representations match the phonetic realizations of laryngeal contrast. In languages with a two-way contrast, this debate can be theoretical: a two-way contrast may be represented with a single feature, and that feature may either match the phonetics of the contrast or be abstracted from it. It is possible to frame each of these phenomena as a matter of 'voicing' regardless of the phonetics, even if laryngeal realists argue that it is inadvisable. A language like Nepali, however, which employs all possible combinations of voicing and aspiration, cannot be represented with a single feature (without using a feature with four values) and forces us to consider each of these phenomena in relation to the phonetics that cue the contrasts.

In an acoustic analysis of Nepali stops in initial and intervocalic positions, Chapter 2 found that laryngeal realist diagnostics broadly hold up when applied to Nepali. Based on the proposed correspondence between prevoicing and a [voice] feature and long-lag VOT and a [spread] feature, we arrived at a representation in which Nepali's T stops are unspecified for either laryngeal feature, D is specified for [voice], Th for [spread], and Dh for both [voice] and [spread]. Speech rate diagnostics corroborate this representation, finding that long-lag VOT is a controlled cue corresponding to a specified feature for Th and Dh, and prevoicing acts as a controlled cue corresponding to a specified feature for D and Dh. The second laryngeal realist diagnostic, which proposes a connection between laryngeal specification and the amount of voicing during intervocalic stop closures, makes incompatible predictions for stops specified for both [voice] and [spread]: [voice] stops are predicted to involve active voicing during the closure, while [spread] stops are predicted to actively block voicing during closure. Nepali Dh stops show active voicing throughout the closure, which motivated a Q Theoretic representation in which the subsegments corresponding to the closure are specified for [voice], while the subsegment corresponding to the release is specified for [spread].

Chapter 3 investigated the 'voicing' effect. The voicing effect, in which vowels are longer before 'voiced' stops, has been found in both aspirating and voicing languages, and various sources for the effect have been proposed, including articulatory explanations that tie the lengthening of the preceding vowel to the articulation of the following consonant. Examining the laryngeal effects on preceding vowels in a language with both phonetic voicing and aspiration could help disentangle the possible sources of the effects. I found that vowels before stops belonging to voiced classes in Nepali are significantly longer than vowels before stops in voiceless classes, while vowels before aspirated stop classes are very slightly (but significantly) longer than vowels before the unaspirated stop classes. I simultaneously found evidence for Beguš's inverse VOT effect which states that vowel duration correlates inversely with the VOT of the following stop. At first these results may seem contradictory, showing longer vowels before aspirated categories but an inverse correlation between vowel length and VOT duration. While a perception experiment is needed to test this, the acoustic patterns seen in the Nepali data here are consistent with a pattern of cue-trading. Across categories, aspirated classes are associated with longer vowels – a positive correlation between aspiration duration and vowel duration. Within categories, however, stops with potentially ambiguous tokens with shorter VOT have longer vowels. While a perception experiment would be necessary to test whether listeners make use of a longer vowel in the absence of clear aspiration contrast, Nepali casts doubts on a purely articulatory explanation for the voicing and aspiration effects.

Chapter 4 investigated the third core phenomenon, laryngeal neutralization. Production and perception studies showed that while the voicing contrast remains robust in final position, the aspiration contrast is weakened: Dh stops are realized without very much aspiration and often perceived as D, while many T stops are realized with long aspiration and often perceived as Th. Nepali shows that not all laryngeal neutralization is devoicing, and requires a theory of laryngeal contrast weakening that allows one laryngeal contrast to weaken independent of the other. Nepali, along with the broader typology of Indic languages, suggests that voicing is a less vulnerable contrast in final position than aspiration. The asymmetric direction of the aspiration contrast weakening furthermore suggests that markedness is not universal, but dependent on the contrasts, the language, and the inventory, and that the same feature [spread] may be the marked member in some contrasts (D/Dh) and the less marked in another contrast (T/Th), even within a single language.

5.2 Contributions and implications

This dissertation contributes data from phonetic studies to the existing body of literature about Nepali. It provides evidence for claims about the weakening of the aspiration contrast that had previously just been reported impressionistically. It also contributes to the literature on laryngeal contrasts by forcing more precise consideration of often-assumed claims of laryngeal realism and markedness to account for a language with both laryngeal contrasts central to realist theories. It affirms the utility of a phonetically-grounded phonological representation, and proposes an even tighter connection between phonetics and phonology with a Q Theoretic representation in which stops are specified for laryngeal features on a subsegmental level.

5.3 Remaining questions and future work

This dissertation focused on primary durational cues to voicing and aspiration contrasts: the duration of voicing during stop closure, and the duration of post-release aspiration (and, in Chapter 3, the durations of the stop closure and preceding vowel). Laryngeal contrasts are known to be cued by a much larger set of cues including non-durational spectral cues to voice quality. I focused on the durational cues as these are central to theories of laryngeal realism, but further work should examine the additional cues, particularly in relation to contrast weakening.

This dissertation also made claims about Indic neutralization patterns based on a sample of languages described in the literature so far, but more production and perception studies should be done on a broader set of languages with four-way laryngeal contrasts to test the typology and markedness patterns suggested in Chapter 4.

Bibliography

Abaglo, Poovi and Diana Archangeli (1989). 'Language-particular underspecification: Gengbe /e/ and Yoruba /i/'. In: *Linguistic Inquiry* 20, pp. 457–480.

Abramson, Arthur S and DH Whalen (2017). 'Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions'. In: *Journal of Phonetics* 63, pp. 75–86.

Acharya, Jayaraj (1991). *A descriptive grammar of Nepali and an analyzed corpus*. Georgetown University Press.

Ahmed, Rais and SS Agrawal (1969). 'Significant features in the perception of (Hindi) consonants'. In: *The Journal of the Acoustical Society of America* 45.3, pp. 758–763.

Archangeli, Diana (1988). 'Aspects of underspecification theory'. In: *Phonology* 5.2, pp. 183–207.

Archangeli, Diana and Douglas Pulleyblank (1989). 'Yoruba vowel harmony'. In: *Linguistic inquiry* 20.2, pp. 173–217.

Asikin-Garmager, Eli (2017). 'Rate effects in Hindi and the phonological specification of voiced aspirates'. Ms., University of Iowa.

Avery, Peter and William J Idsardi (2001). 'Laryngeal dimensions, completion and enhancement'. In: *Distinctive feature theory*. Ed. by T.A. Hall and U. Kleinhanz. Berlin: Mouton de Gruyter, pp. 41–70.

Barr, Dale J et al. (2013). 'Random effects structure for confirmatory hypothesis testing: Keep it maximal'. In: *Journal of Memory and Language* 68.3, pp. 255–278.

Bates, Douglas et al. (2015). 'Fitting Linear Mixed-Effects Models Using lme4'. In: *Journal of Statistical Software* 67.1, pp. 1–48. DOI: `10.18637/jss.v067.i01`.

Beckman, Jill (handout). 'Privative laryngeal features and passive voicing: Evidence from Hindi'. mfm handout.

Beckman, Jill, Pétur Helgason et al. (2011). 'Rate effects on Swedish VOT: Evidence for phonological overspecification'. In: *Journal of Phonetics* 39.1, pp. 39–49.

Beckman, Jill, Michael Jessen and Catherine Ringen (2013). 'Empirical evidence for laryngeal features: Aspirating vs. true voice languages'. In: *Journal of Linguistics* 49.02, pp. 259–284.

Begus, Gasper (2017). 'Effects of ejective stops on preceding vowel duration'. In: *The Journal of the Acoustical Society of America* 142.4, pp. 2168–2184.

Belasco, Simon (1953). 'The influence of force of articulation of consonants on vowel duration'. In: *The Journal of the Acoustical Society of America* 25.5, pp. 1015–1016.

Berkson, Kelly Harper (2012). 'Capturing Breathy Voice: Durational Measures of Oral Stops in Marathi'. In: *Kansas Working Papers in Linguistics* 33, pp. 27–46.

Berkson, Kelly Harper (2013). ‘Phonation types in Marathi: An acoustic investigation’. PhD thesis. University of Kansas.

Blevins, Juliette (2006). ‘A theoretical synopsis of Evolutionary Phonology’. In: *Theoretical Linguistics* 32, pp. 117–166.

Boersma, Paul and David Weenink (2015). *Praat [Computer program]*. version 6.0.05. http://www.praat.org/.

Braunschweiler, Norbert (1997). ‘Integrated cues of voicing and vowel length in German: A production study’. In: *Language and Speech* 40.4, pp. 353–376.

Broselow, Ellen (2018). ‘Laryngeal contrasts in second language phonology’. In: *Phonological typology. Berlin: de Gruyter Mouton*, pp. 312–340.

Browman, Catherine P and Louis M Goldstein (1986). ‘Towards an articulatory phonology’. In: *Phonology* 3, pp. 219–252.

Brown, Jason (2016). ‘Laryngeal assimilation, markedness and typology’. In: *Phonology* 33.3, pp. 393–423.

Chatterji, Suniti Kumar (1926). *The origin and development of the Bengali language*. Vol. 2. Calcutta University Press.

Chen, Matthew (1970). ‘Vowel length variation as a function of the voicing of the consonant environment’. In: *Phonetica* 22.3, pp. 129–159.

Chomsky, Noam and Morris Halle (1968). *The sound pattern of English*. Harper & Row.

Clayards, Meghan (2017). ‘Individual talker and token covariation in the production of multiple cues to stop voicing’. In: *Phonetica* 75.1, pp. 1–23.

Clements, George N (1985). ‘The geometry of phonological features’. In: *Phonology* 2.01, pp. 225–252.

Clements, George N and Rajesh Khatiwada (2007). ‘Phonetic realization of contrastively aspirated affricates in Nepali’. In: *Proceedings of the 16th International Congress of Phonetic Sciences*. Ed. by J. Trouvain and W. J. Barry, pp. 629–632.

Coretta, Stefano (2019). ‘An exploratory study of voicing-related differences in vowel duration as compensatory temporal adjustment in Italian and Polish’. In: *Glossa: a journal of general linguistics* 4.1.

Davis, Katharine (1994). ‘Stop voicing in Hindi’. In: *Journal of Phonetics* 22.2, pp. 177–193.

Durvasula, Karthik and Qian Luo (2012). ‘Voicing, aspiration, and vowel duration in Hindi’. In: *Proceedings of meetings on acoustics*. Vol. 18. 1. AIP Publishing.

Dutta, Hemanga and Michael Kenstowicz (2018). ‘The phonology and phonetics of laryngeal stop contrasts in Assamese’. In: *From sounds to structures: Beyond the veil of Maya*, pp. 30–64.

Dutta, Indranil (2007). ‘Four-way stop contrasts in Hindi: An acoustic study of voicing, fundamental frequency and spectral tilt’. PhD thesis. University of Illinois at Urbana-Champaign.

Farnetani, Edda and Shiro Kori (1986). ‘Effects of syllable and word structure on segmental durations in spoken Italian’. In: *Speech communication* 5.1, pp. 17–34.

Flemming, Edward (2005). ‘Deriving natural classes in phonology’. In: *Lingua* 115.3, pp. 287–309.

Gafos, Adamantios I (2002). ‘A grammar of gestural coordination’. In: *Natural language & linguistic theory* 20.2, pp. 269–337.

Hale, Mark and Charles Reiss (2008). *The phonological enterprise*. Oxford University Press.

Halle, Morris, Kenneth N Stevens and Alan V Oppenheim (1967). 'On the mechanism of glottal vibration for vowels and consonants'. In: *Massachusetts Institute of Technology, Research Laboratory of Electronics, Quarterly Progress Report* 85, pp. 267–277.

Harris, John (1994). *English sound structure*. Blackwell.

Haspelmath, Martin (1993). *A grammar of Lezgian*. 9. Walter de Gruyter.

Hauser, Ivy (2019). 'Effects of phonological contrast on within-category phonetic variation'. PhD thesis. University of Massachusetts, Amherst.

Honeybone, Patrick (2005). 'Diachronic evidence in segmental phonology: the case of obstruent laryngeal specifications'. In: *The internal organization of phonological segments*. Ed. by M. van Oostendorp and J. van de Weijer. Mouton de Gruyter, pp. 319–354.

House, Arthur S and Grant Fairbanks (1953). 'The influence of consonant environment upon the secondary acoustical characteristics of vowels'. In: *The Journal of the Acoustical Society of America* 25.1, pp. 105–113.

Hussain, Sarmad and Rami Nair (1995). 'Voicing and aspiration contrasts in Hindi and Urdu'. In: *31st Meeting of the Chicago Linguistics Society*.

Inkelas, Sharon and Stephanie S Shih (2017). 'Looking into Segments'. In: *Proceedings of the Annual Meetings on Phonology*. Vol. 4.

— (2016). 'Re-representing phonology: consequences of Q Theory'. In: *Proceedings of NELS*. Vol. 46. 2, pp. 161–174.

Iosad, Pavel (2012). 'Representation and variation in substance-free phonology: A case study in Celtic'. PhD thesis. University of Tromsø.

Islam, Md Jahurul (2019). 'Phonetics and Phonology of'Voiced-Aspirated'Stops: Evidence from Production, Perception, Alternation and Learnability'. PhD thesis.

Iverson, Gregory K and Joseph C Salmons (1995). 'Aspiration and laryngeal representation in Germanic'. In: *Phonology* 12.03, pp. 369–396.

— (2011). 'Final devoicing and final laryngeal neutralization'. In: *The Blackwell Companion to Phonology*. Ed. by M. van Oostendorp et al. Vol. 3. Wiley-Blackwell, pp. 1622–1643.

— (2003). 'Legacy specification in the laryngeal phonology of Dutch'. In: *Journal of Germanic Linguistics* 15.01, pp. 1–26.

Jakobson, Roman, Gunnar Fant and Morris Halle (1952). *Preliminaries to speech analysis: The distinctive features and their correlates*. Tech. rep. 13. Massachussets Institute of Technology Acoustics Laboratory.

Jansen, Wouter (2004). 'Laryngeal contrast and phonetic voicing: A laboratory phonology approach to English, Hungarian, and Dutch'. PhD thesis. University of Groningen.

Jessen, Michael and Catherine Ringen (2002). 'Laryngeal features in German'. In: *Phonology* 19.02, pp. 189–218.

Keating, Patricia (1984). 'Universal phonetics and the organization of grammars'. In: *UCLA Working Papers in Phonetics* 59, pp. 35–49.

Keating, Patricia A (1984). 'Phonetic and phonological representation of stop consonant voicing'. In: *Language* 60 (2), pp. 286–319.

Kenstowicz, Michael J (1994). *Phonology in generative grammar*. Vol. 7. Blackwell Cambridge, MA.

Kessinger, Rachel H and Sheila E Blumstein (1997). 'Effects of speaking rate on voice-onset time in Thai, French, and English'. In: *Journal of Phonetics* 25.2, pp. 143–168.

Khatiwada, Rajesh (2007). 'Nepalese retroflex stops: a static palatography study of inter-and intra-speaker variability.' In: *INTERSPEECH*, pp. 1422–1425.

— (2009). 'Nepali'. In: *Journal of the International Phonetic Association* 39.03, pp. 373–380.

Kluender, Keith R, Randy L Diehl and Beverly A Wright (1988). 'Vowel-length differences before voiced and voiceless consonants: An auditory explanation'. In: *Journal of phonetics* 16.2, pp. 153–169.

Kuznetsova, Alexandra, Per Bruun Brockhoff and Rune Haubo Bojesen Christensen (2015). *lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package).* R package version 2.0-25. URL: `http://CRAN.R-project.org/package=lmerTest`.

Lampp, Claire and Heidi Reklis (2004). 'Effects of coda voicing and aspiration on Hindi vowels'. In: *The Journal of the Acoustical Society of America* 115.5_Supplement, pp. 2540–2540.

Lindblom, Björn EF (1967). 'Vowel duration and a model of lip mandible coordination'. In: *Speech Transmission Laboratory Quarterly Progress Status Report* 4, pp. 1–29.

Lisker, Leigh and Arthur S Abramson (1964). 'A cross-language study of voicing in initial stops: Acoustical measurements'. In: *Word* 20.3, pp. 384–422.

Lombardi, Linda (1991). *Laryngeal features and laryngeal neutralization*. Garland.

— (1999). 'Positional faithfulness and voicing assimilation in Optimality Theory'. In: *Natural Language & Linguistic Theory* 17.2, pp. 267–302.

Maddieson, Ian (1997). 'The handbook of phonetic sciences'. In: ed. by William J Hardcastle, John Laver and Fiona E Gibbon. John Wiley & Sons. Chap. Phonetic Universals, pp. 619–639.

Maddieson, Ian and Jack Gandour (1976). 'On 'explaining' vowel duration variation'. In: *UCLA Working Papers in Phonetics* 31, pp. 46–52.

Magloire, Jo&euml l and Kerry P Green (1999). 'A cross-language comparison of speaking rate effects on the production of voice onset time in English and Spanish'. In: *Phonetica* 56.3-4, pp. 158–185.

Masica, Colin P (1993). *The indo-aryan languages*. Cambridge University Press.

Mielke, Jeff (2008). *The emergence of distinctive features*. Oxford University Press.

Mikuteit, Simone and Henning Reetz (2007). 'Caught in the ACT: The timing of aspiration and voicing in East Bengali'. In: *Language and Speech* 50.2, pp. 247–277.

Mitleb, Fares M (1984). 'Voicing effect on vowel duration is not an absolute universal'. In: *Journal of Phonetics* 12.1, pp. 23–27.

Möbius, Bernd (2004). 'Corpus-based investigations on the phonetics of consonant voicing'. In: *Folia Linguistica* 38.1-2, pp. 5–26.

Mohr, Burckhard (1971). 'Intrinsic variations in the speech signal'. In: *Phonetica* 23.2, pp. 65–93.

Nakkeerar, R (2011). 'Nepali in Sikkim'. In: *Linguistic Survey of India—Sikkim, Part II*. Language Division Office of the Registrar General & Census Commissioner, pp. 23–120.

Ohala, John J (1983). 'The origin of sound patterns in vocal tract constraints'. In: *The production of speech*. Ed. by P.F. MacNeilage. Springer, pp. 189–216.

Ohala, M. and John J Ohala (Oct. 1992). ‘Phonetics universals and Hindi segment duration’. In: *Proceedings of the International Conference on Spoken Language Processing*. Banff, Alberta, Canada, pp. 831–834.

Pape, Daniel and Luis MT Jesus (2014). ‘Production and perception of velar stop (de) voicing in European Portuguese and Italian’. In: *EURASIP Journal on Audio, Speech, and Music Processing* 2014.1, p. 6.

Pattanayak, Debi Prasanna (1966). ‘A controlled historical reconstruction of Oriya, Assamese, Bengali, and Hindi’. In: *(No Title)*.

Pind, Jörgen (1995). ‘Speaking rate, voice-onset time, and quantity: The search for higher-order invariants for two Icelandic speech cues’. In: *Attention, Perception, & Psychophysics* 57.3, pp. 291–304.

Pokharel, Madhav Prasad Sharma (1989). ‘Experimental analysis of Nepali sound system’. PhD thesis. University of Pune.

Poon, Pamela G and Catherine A Mateer (1985). ‘A study of VOT in Nepali stop consonants’. In: *Phonetica* 42.1, pp. 39–47.

Ramsammy, Michael and Patrycja Strycharczuk (2016). ‘From phonetic enhancement to phonological underspecification: hybrid voicing contrast in European Portuguese’. In: *Papers in Historical Phonology* 1, pp. 285–315.

Reiss, Charles (2017). ‘Substance free phonology’. In: *The Routledge Handbook of Phonological Theory*. Ed. by S. J. Hannahs and Anna R. K. Bosch. Routledge, pp. 425–452.

Rice, Keren (1999). ‘Featural markedness in phonology: variation’. In: *Glot international* 4.7, pp. 3–6.

Ridouane, Rachid, George N Clements and Rajesh Khatiwada (2010). ‘Language-independent bases of distinctive features’. In: *Tones and Features: Phonetic and Phonological Perspectives*. Ed. by J.A. Goldsmith, E. Hume and L. Wetzels. Walter de Gruyter, pp. 264–291.

Rubach, Jerzy (1990). ‘Final devoicing and cyclic syllabification in German’. In: *Linguistic Inquiry* 21.1, pp. 79–94.

Schwarz, Martha (2017). ‘Realization and representation of Nepali laryngeal contrasts: voiced aspirates and Laryngeal Realism’. MA thesis. McGill University.

Schwarz, Martha, Myriam Lapierre et al. (2019). ‘Recent advances in Q Theory: Segment strength’. In: *Proceedings of the Linguistic Society of America* 4, pp. 51–1.

Schwarz, Martha, Morgan Sonderegger and Heather Goad (2019). ‘Realization and representation of Nepali laryngeal contrasts: voiced aspirates and Laryngeal Realism’. In: *Journal of Phonetics* 73, pp. 113–127.

Shih, Stephanie and Sharon Inkelas (2014). ‘A subsegmental correspondence approach to contour tone (dis) harmony patterns’. In: *Proceedings of the Annual Meetings on Phonology*. Vol. 1. 1.

Slis, Iman Hans and Antonie Cohen (1969). ‘On the complex regulating the voiced-voiceless distinction I’. In: *Language and speech* 12.2, pp. 80–102.

Solé, M-J (2007). ‘Controlled and mechanical properties in speech: A review of the literature’. In: *Experimental Approaches to Phonology*. Ed. by M-J Solé, P. Beddor and M. Ohala. Oxford: Oxford University Press, pp. 302–321.

Steriade, Donca (1993). ‘Closure, release, and nasal contours’. In: *Phonetics and phonology* 5, pp. 401–470.

— (1994). ‘Complex onsets as single segments: the Mazateco pattern’. In: *Perspectives in phonology*, pp. 203–291.

— (1997). ‘Phonetics in phonology: the case of laryngeal neutralization’. In.

Stevens, Kenneth N (1998). *Acoustic Phonetics*. MIT Press.

Tosco, Mauro (2001). *The Dhaasanac language: Grammar, texts, vocabulary of a Cushitic language of Ethiopia*. Rudiger Koppe Verlag.

Turk, Alice, Satsuki Nakai and Mariko Sugahara (2006). ‘Acoustic segment durations in prosodic research: A practical guide’. In: *Methods in empirical prosody research*. Mouton de Gruyter, pp. 1–28.

Vaux, Bert and Bridget Samuels (2005). ‘Laryngeal markedness and aspiration’. In: *Phonology* 22.3, pp. 395–436.

Westbury, John R and Patricia A Keating (1986). ‘On the naturalness of stop consonant voicing’. In: *Journal of linguistics* 22.1, pp. 145–166.

Whelpton, John (2005). *A history of Nepal*. Cambridge University Press.

Wiese, Richard (1996). *The phonology of German*. Clarendon Press.

Wilson, Colin (2001). ‘Consonant cluster neutralisation and targeted constraints’. In: *Phonology* 18.1, pp. 147–197.

Yu, Alan CL (2004). ‘Explaining final obstruent voicing in Lezgian: Phonetics and history’. In: *Language*, pp. 73–97.