Realization and representation of Nepali laryngeal contrasts: Voiced aspirates and laryngeal realism

Martha Schwarz*

Department of Linguistics, University of California, Berkeley, Dwinelle Hall #2650, Berkeley, CA 94720

Morgan Sonderegger

Department of Linguistics, McGill University, 1085 Dr Penfield Avenue, Montreal, Quebec, H3A 1A7, Canada

Heather Goad

Department of Linguistics, McGill University, 1085 Dr Penfield Avenue, Montreal, Quebec, H3A 1A7, Canada

Abstract

Theories of laryngeal realism argue for a tight correspondence between a segment's phonetic cues and the (laryngeal) phonological features that represent it. As such, the 'p'/'b' contrast in French, expressed phonetically by vocal fold vibration during the stop closure, is represented by a [voice] feature while the 'p'/'b' contrast in English, expressed phonetically by contrasting long and short lag VOT, is represented by a [spread] feature. Laryngeal realist literature focuses on whether a given segment is best represented by [voice] or [spread], and proposes a set of criteria and tests by which to diagnose the representation. In this study we push laryngeal realist theory in a new direction – to segments proposed to be specified for both [voice] *and* [spread] features – a combination which poses challenges to the current diagnostics. To do so, we analyze acoustic data from Nepali, an Indic (a.k.a. Indo-Aryan) language with a single class of stops described as both voiced and aspirated. We apply the same criteria and diagnostics

*Corresponding author at: Tel: (510) 642-2757; fax: (510) 643-5688.

Email addresses: martha_schwarz@berkeley.edu (Martha Schwarz*), morgan.sonderegger@mcgill.ca (Morgan Sonderegger), heather.goad@mcgill.ca (Heather Goad)

used in laryngeal realism. We find support for the proposed representation, with a caveat that the [voice] feature appears 'stronger' than [spread].

Keywords: Laryngeal realism; Nepali; Indo-Aryan; Indic; voiced aspirates; laryngeal timing

---

1. Introduction

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. Clements, 1985; Flemming, 2005; Jakobson et al., 1952; Mielke, 2008; but see Hale & 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. To answer these questions, the literature proposes diagnostics based on phonetic data and phonological behavior. In this study we examine how well these criteria extend to a language with a single segment arguably specified for both [voice] *and* [spread].

This study examines Nepali as spoken in the northeast Indian state of Sikkim. Nepali is an Indic (a.k.a. Indo-Aryan) language with a four-way laryngeal contrast proposed to utilize only [voice] and [spread] features, including a stop class proposed to be specified with both [voice] and [spread] in the laryngeal realism literature (described below). 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.

1.1. Two approaches to representing laryngeal contrast

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. Abramson & Whalen, 2017; Lisker & Abramson, 1964). Since 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 cord 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 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). A large body of research has discussed this issue's bearing on the link between phonological representation and phonetic realization. Should this two-way contrast should 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 & Halle, 1968; Keating, 1984; Lombardi, 1991, 1999; Rubach, 1990; Wiese, 1996), 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 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. Avery & Idsardi, 2001; Beckman et al., 2013; Brown, 2016; Harris, 1994; Honeybone, 2005; Iverson & Salmons, 1995, 2003; Jessen & Ringen, 2002; Vaux & Samuels,

[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.

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

2005) 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: 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

[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 & Pulleyblank (1989) on [atr] in Yoruba vs. Masaai and Abaglo & Archangeli (1989) on [high] in Yoruba vs. Gengbe). This does not hold for privative systems, as discussed below in the text.

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 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 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 paper 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 the next sections.

Table 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]

1.2. 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', 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 et al. (2011, 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 et al. (2011, 2013) turn similar ideas into a pair of feature specification diagnostics. Beckman et al. (2011) review literature on the effects of speech rate on word-initial VOT durations in voicing and aspirating languages (e.g. Kessinger & Blumstein, 1997; Magloire & Green, 1999; Pind, 1995), 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. 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 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 et al., 2013, 259; Jansen, 2004; Jessen & Ringen, 2002). Meanwhile, Beckman 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); Pape & 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. Iverson & Salmons, 2011; Lombardi, 1991). 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.

All three types of evidence – phonetic realization, diagnostics of control, and phonological patterning/markedness – are important within laryngeal realism, though different studies place greater importance on different types. This study will focus on the first two types of evidence – phonetic realization and diagnostics of control – not because we consider them to be more important than phonological patterning, but because we see four-way contrasting Indic languages as particularly challenging for the diagnostics, and a particularly apparent gap in the phonetic measurability arguments for laryngeal realism. We leave evidence from phonological processes and markedness arguments to future work.

1.3. 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 (Beckman et al., 2011; Kessinger & Blumstein, 1997; Pind, 1995), as shown in Table 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 et al., 2011). Beckman 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 et al. (2011) to propose overspecification.[5]

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 & Salmons (1995) propose the feature representation in Table 3. This representation uses the same [voice] and [spread] features as English, French, and Thai, but exploits every logical combination of them.

Table 3: 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 4.

[5]See also Ramsammy & 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 4: 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 4.

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 3. 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 paper is organized as follows. Section 2 provides background on Nepali and on phonetic cues relevant to four-way contrasts. Section 3 explains the methods of data collection and analysis. Sections 4–6 present the results of the study and their implications for feature representation. Section 4 addresses research question 1, examining the phonetic realization of stops in initial position. Section 5 addresses research question 2, applying the diagnostics of control. Section 6 concludes.

2. Background

2.1. Background on Nepali

Nepali is an Indic language spoken primarily in Nepal and northern India. The data for this study comes from the variety spoken in Sikkim, in northern India. The focus of this study is Nepali's four-way laryngeal contrast between voiceless, voiced, voiceless aspirated, and voiced aspirated stops. While this contrast exists in both stops and affricates, this study considers only stops to enable comparison to work on laryngeal contrasts in other languages (see Clements & Khatiwada, 2007 for an acoustic study of Nepali affricates). The stop inventory is shown in Table 5.

Table 5: Nepali stop inventory

bilabial		alveolar		retroflex		velar	
p	pʰ	t	tʰ	ʈ	ʈʰ	k	kʰ
b	bʰ	d	dʰ	ɖ	ɖʰ	g	gʰ

Throughout this paper we will label the [p, t, ʈ, k] series 'voiceless' or 'T', the [pʰ, tʰ, ʈʰ, kʰ] series 'voiceless aspirated' or 'Th', the [b, d, ɖ, g] series 'voiced' or 'D', and the [bʰ, dʰ, ɖʰ, gʰ] series 'voiced aspirated' or 'Dh'.

2.2. 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) (Berkson, 2012; Clements & Khatiwada, 2007; Davis, 1994; Dutta, 2007; Mikuteit & Reetz, 2007). 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 (Clements & Khatiwada, 2007; Poon & Mateer, 1985) as well as on related languages including Hindi (Davis, 1994; Dutta, 2007; Lisker & Abramson, 1964), Marathi (Berkson, 2013), and Bengali (Mikuteit & Reetz, 2007).

2.2.1. Voice Onset Time

The classic durational measure used to distinguish stop classes from each other is voice onset time (VOT) (Lisker & 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 & Abramson, 1964; Poon & Mateer, 1985), as schematized in Table 6.

Table 6: 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.

2.2.2. 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). 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 (Davis, 1994; Dutta, 2007; Lisker & Abramson, 1964), 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 (Beckman et al., 2013; Iverson & Salmons, 1995), as will be discussed further in Section 5.2.

2.2.3. 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

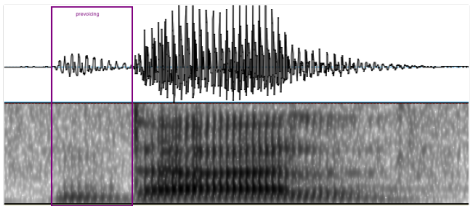

Figure 1: Example of prevoicing/negative VOT, for the Nepali word [dal] ‘lentils’.

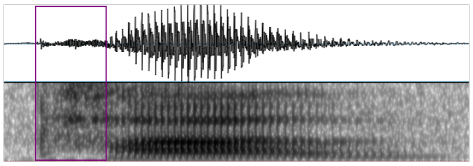

Figure 2: Example of post-release/positive VOT, for the Nepali word [t^{h}al] ‘plate’.

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 2).

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 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 (Berkson, 2012; Mikuteit & Reetz, 2007).

Mikuteit & 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 & 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 visu-

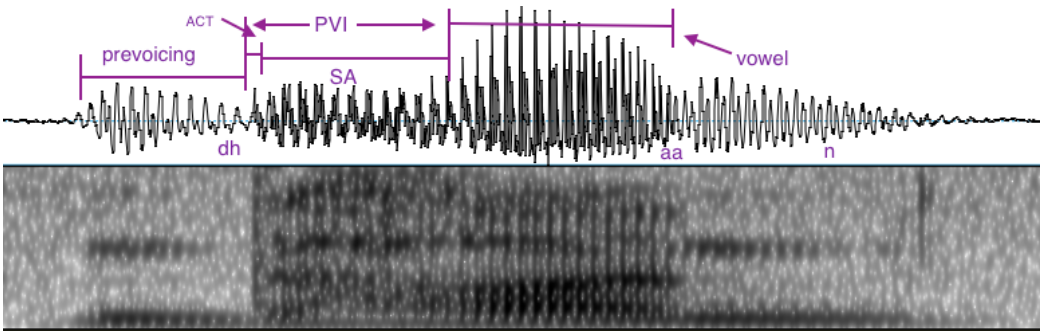

Figure 3: Example of post-release on a voiced aspirated stop, with various annotation schemes marked, for the Nepali word [d^{h}an] 'rice paddy'.

ally, 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. In the present study, we follow Berkson and use the PVI measure.

3. 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.

3.1. Participants

The data for this study was collected from 17 speakers (10 male, 7 female), all 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 were also all bilingual in 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.

3.2. Stimuli

The stimuli consist of 32 Nepali words, corresponding to each of the 16 stops in Table 5, 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.[6] 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 & 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.'

[6]We expect prosody to affect stop realization, but the factors governing word-level stress/prominence in Nepali are not very clear (Clements & Khatiwada, 2007; Acharya, 1991). 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.

The results presented here are from the X_1 and X_3 positions of all 17 speakers, because each of these positions is preceded and followed by a pause, giving roughly comparable prosodic contexts. This yielded a total of 559 tokens with the segment in initial position.[7] 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.

3.3. Acoustic analysis

The recordings were imported into Praat (Boersma & Weenink, 2015), where they were hand-annotated for voicing, closure, and post-release duration measurements, as pictured in Figure 4. 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 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.

4. 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 5) confirm the (non-)significance of patterns discussed in the empirical data here.

4.1. 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

[7] 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.

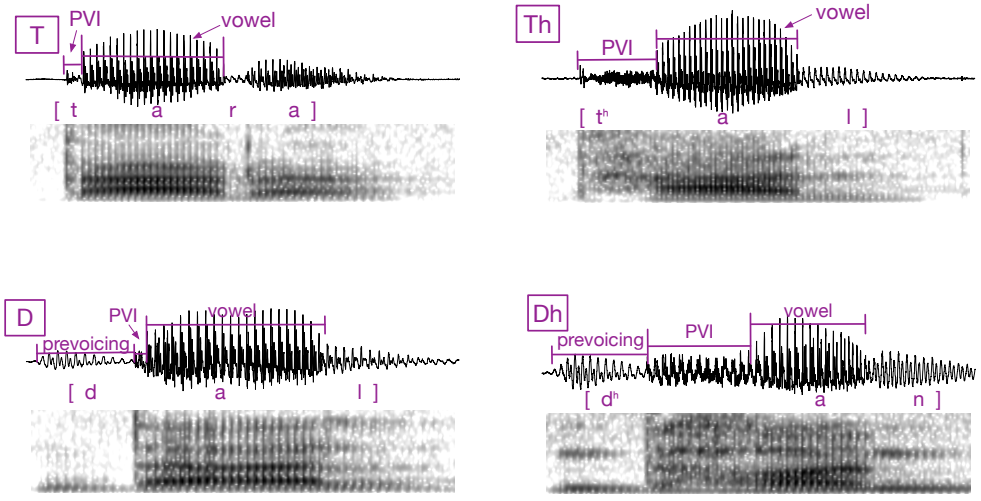

Figure 4: Examples of annotation. Top left: [tara] 'star'. No prevoicing, end of stop closure marked by a burst with very short interval of aperiodic aspiration before voicing begins. Top right: [t^hal] 'plate'. No prevoicing, ACT is much longer than for [tara], but still marked by aperiodic high frequency 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].

(Lisker & Abramson, 1964; Poon & Mateer, 1985). In Figure 5 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 7 summarizes the means and standards of deviation of these measures for each class.

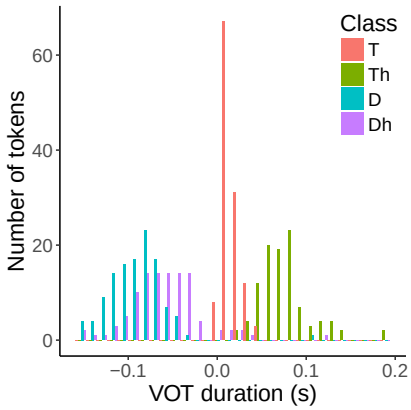

Figure 5: Distribution of VOT durations in initial position, for each stop class.

Table 7: 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 6 (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 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.

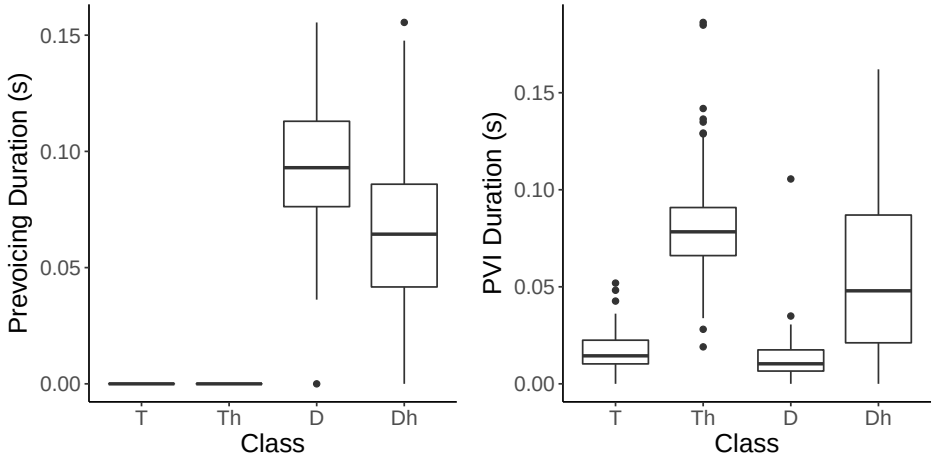

Figure 6: Prevoicing duration (left) and PVI duration (right) in initial position, across stop classes.

Figure 6 (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 5.

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 7 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 5.

4.2. 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 realized with phonetic prevoicing

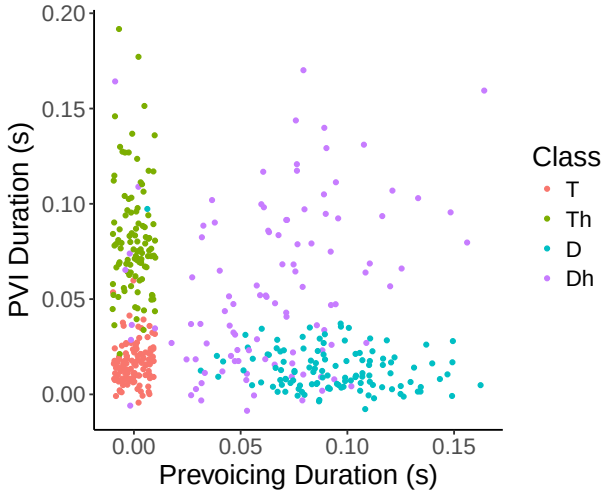

Figure 7: Prevoicing duration versus PVI duration for stops in initial position, for each stop class. Points are jittered for visibility.

(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 & 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 6), 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. The difference can be seen in Figure 4, comparing [t^{h}al] to [d^{h}an]. 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 & 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 8, 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 & Salmons's (1995) representation.

Table 8: 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 arrived at these feature representations using the same criterion as previous laryngeal realist studies: consistent use of a phonetic cue corresponding to a feature. However, it is worth explicitly discussing briefly whether the criteria need to be revised in light of data from four-way contrasting languages in which each feature is used for two classes – specifically, how relevant relative duration of the cues is. Both D and Dh show consistent presence of prevoicing, which is the criterion for a [voice] feature. But recall that the prevoicing duration of D is significantly longer than that of Dh. None of the languages studied thus far include two different classes specified for [voice], so laryngeal realism has not yet had to address this possibility. The spirit of cue-to-feature correspondence might suggest that this difference in prevoicing length should be captured in the representation as well. The same issue arises with the [spread] feature. Both Th and Dh usually have PVI of at least 30 msec, but the PVI of Th is significantly longer than that of Dh (see mixed effect model in section 5.1.2).

Whether these voicing and PVI durational differences are reflected in the representation is essentially a question concerning where to draw the line on which phonetic detail to capture, and this is an outstanding issue which various laryngeal realist theories disagree on. It may be relevant whether, for example, shorter prevoicing when followed by long PVI is physiologically inevitable rather than phonologically controlled, and further work should investigate whether these durational differences are typologically consistent and whether there is any evidence language-internally that the durational differences have any phonological implications. If so, laryngeal realist logic may suggest that these consistent phonetic differences in voicing and aspiration cues between distinct laryngeal classes warrant differences in feature representation, and the formal machinery would need to be extended to capture these differences. Based on the data currently available, we support the feature representations in Table 8 as the most direct result of applying current realist criteria to Nepali.

5. 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 et al. (2011, 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.

5.1. Speech rate effects in initial position

Recall that Beckman 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 (Beckman et al., 2011; Davis, 1994): [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 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 & Blumstein, 1997; Magloire & Green, 1999) and Thai (Kessinger & 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 & Blumstein, 1997), Spanish (Magloire & Green, 1999), and Thai (Kessinger & Blumstein, 1997).

We now test the predictions of this diagnostic against Iverson & 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 8 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.[8] Figure 8 (left) plots prevoicing durations across speech

[8]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

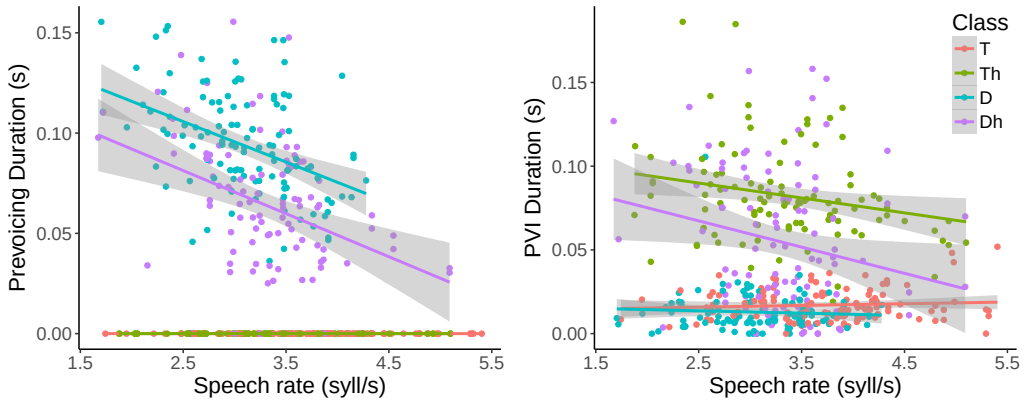

Figure 8: Effect of speech rate on prevoicing duration (left) and PVI duration (right) for stops in initial position.

rates and shows negatively-sloped trend lines for the voiced and voiced aspirated classes, suggesting longer prevoicing durations at slower speech rates. Figure 8 (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 5.1.1 and 5.1.2) 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

attested in this data is comparable to the rates reported in Beckman et al. (2011), who did explicitly prompt fast speech. We use the term 'fast' to describe the faster rates in the data, but recognize that this is slower than the fast end of the continuum of spontaneous speech

Table 9: 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

average value of other predictors, for an average speaker and word.

5.1.1. Prevoicing duration

The only stop classes that ever display prevoicing are the voiced and voiced aspirated classes (see Figures 6 (left) and 8 (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 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 9.

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 6. 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 signific-

antly shorter prevoicing durations than bilabial stops ($p < 0.01$), consistent with the difficulty of maintaining voicing for less anterior constrictions (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 ($p = 0.19$).

5.1.2. PVI duration

Based on Figure 8, 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 aspirated 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).[9] The model's fixed effects are summarized in Table 10.

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 4.1 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

[9]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.

Table 10: 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

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 & Abramson, 1964). (Recall that PVI is closely related to positive VOT.)

5.1.3. Summary

The speech rate diagnostic provides evidence for the Iverson & 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.

5.2. Passive vs. active voicing in medial position

5.2.1. 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 & Jesus (2014) provide additional data, finding that European Portuguese *ptk* stops also have 20-30% of the closure voiced[10]. 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 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 11.

[10]One purpose of Pape & 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 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 et al. (2013).

Table 11: Intervocalic voicing in Russian and German based on data from Beckman 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

5.2.2. 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 9, and summarized in Table 12.

Table 12: 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

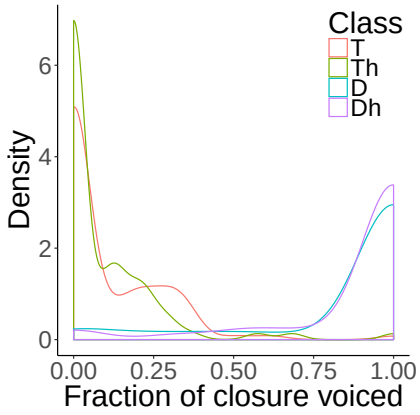

Figure 9: Distribution of the voicing fraction for stops in medial (intervocalic) position, for each stop class.

expected of [voice]-specified stops 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 9 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 9. 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 9 is clear: belonging to a class specified for [voice] is the main predictor of voicing proportion.[11]

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 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.

6. Conclusions

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]. At the beginning of the paper, we set out to extend the theory of laryngeal realism to an area where it has 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 set to answering this question based on two types of evidence: phonetic realization in

[11] 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.

initial position, and diagnostics of control in initial and medial position. We found that many aspects of the theory transfer to doubly-specified segments without issue (initial realization, speech rate diagnostic), but that intervocalic voicing poses a challenge.

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. Intervocalically, T stops and Dh stops each patterned like the stops in a [voice] contrasting language rather than a [spread] contrasting language, suggesting the strength of the [voice] feature over [spread]. We mentioned two directions to pursue an account for the asymmetry, by attributing the prominence of voicing to the intervocalic context or by capturing the asymmetry in the representation, but leave this to future work.

Acknowledgements

This research was funded in part by a Mitacs Globalink Research Award to Martha Schwarz; SSHRC grants #435-2017-092 to Morgan Sonderegger and #435-2015-0490 to Heather Goad and Lydia White; and FRQSC grants #183356 to Morgan Sonderegger and #430-2014-00018 to Lydia White, Heather Goad, and colleagues. We are grateful for comments from audiences at Manchester Phonology Meeting 25, the Annual Meeting on Phonology 2017, and three anonymous reviewers. We especially thank Shrijana Chhetri, Samar Sinha, the Nepali Department at Sikkim University, and its students for their participation and the facilities that allowed us to collect this data.

References

Abaglo, P., & Archangeli, D. (1989). Language-particular underspecification: Gengbe /e/ and Yoruba /i/. *Linguistic Inquiry*, *20*, 457–480.

Abramson, A. S., & Whalen, D. (2017). Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics*, *63*, 75–86.

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

Archangeli, D. (1988). Aspects of underspecification theory. *Phonology*, *5*, 183–207.

Archangeli, D., & Pulleyblank, D. (1989). Yoruba vowel harmony. *Linguistic inquiry*, *20*, 173–217.

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

Avery, P., & Idsardi, W. J. (2001). Laryngeal dimensions, completion and enhancement. In T. Hall, & U. Kleinhanz (Eds.), *Distinctive feature theory* (pp. 41–70). Berlin: Mouton de Gruyter.

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

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

Beckman, J. (handout). Privative laryngeal features and passive voicing: Evidence from Hindi. Mfm handout.

Beckman, J., Helgason, P., McMurray, B., & Ringen, C. (2011). Rate effects on Swedish VOT: Evidence for phonological overspecification. *Journal of Phonetics*, *39*, 39–49.

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

Berkson, K. H. (2012). Capturing breathy voice: Durational measures of oral stops in Marathi. *Kansas Working Papers in Linguistics*, *33*, 27–46.

Berkson, K. H. (2013). *Phonation types in Marathi: An acoustic investigation*. Ph.D. thesis University of Kansas.

Boersma, P., & Weenink, D. (2015). Praat [computer program]. Version 6.0.05. http://www.praat.org/.

Brown, J. (2016). Laryngeal assimilation, markedness and typology. *Phonology*, *33*, 393–423.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. Harper & Row.

Clements, G. N. (1985). The geometry of phonological features. *Phonology*, *2*, 225–252.

Clements, G. N., & Khatiwada, R. (2007). Phonetic realization of contrastively aspirated affricates in Nepali. In J. Trouvain, & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 629–632).

Davis, K. (1994). Stop voicing in Hindi. *Journal of Phonetics*, *22*, 177–193.

Dutta, I. (2007). *Four-way stop contrasts in Hindi: An acoustic study of voicing, fundamental frequency and spectral tilt*. Ph.D. thesis University of Illinois at Urbana-Champaign.

Flemming, E. (2005). Deriving natural classes in phonology. *Lingua*, *115*, 287–309.

Hale, M., & Reiss, C. (2008). *The phonological enterprise*. Oxford University Press.

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

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

Iosad, P. (2012). *Representation and variation in substance-free phonology: A case study in Celtic*. Ph.D. thesis University of Tromsø.

Iverson, G. K., & Salmons, J. C. (1995). Aspiration and laryngeal representation in Germanic. *Phonology*, *12*, 369–396.

Iverson, G. K., & Salmons, J. C. (2003). Legacy specification in the laryngeal phonology of Dutch. *Journal of Germanic Linguistics*, *15*, 1–26.

Iverson, G. K., & Salmons, J. C. (2011). Final devoicing and final laryngeal neutralization. In M. van Oostendorp, C. Ewen, B. Hume, & K. Rice (Eds.), *The Blackwell Companion to Phonology* (pp. 1622–1643). Wiley-Blackwell volume 3.

Jakobson, R., Fant, G., & Halle, M. (1952). *Preliminaries to speech analysis: The distinctive features and their correlates*. Technical Report 13 Massachussets Institute of Technology Acoustics Laboratory.

Jansen, W. (2004). *Laryngeal contrast and phonetic voicing: A laboratory phonology approach to English, Hungarian, and Dutch*. Ph.D. thesis University of Groningen.

Jessen, M., & Ringen, C. (2002). Laryngeal features in German. *Phonology*, *19*, 189–218.

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language*, *60*, 286–319.

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

Kuznetsova, A., Bruun Brockhoff, P., & Haubo Bojesen Christensen, R. (2015). *lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package).*. R package version 2.0-25.

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

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

Lombardi, L. (1999). Positional faithfulness and voicing assimilation in optimality theory. *Natural Language & Linguistic Theory*, *17*, 267–302.

Magloire, J. l., & Green, K. P. (1999). A cross-language comparison of speaking rate effects on the production of voice onset time in English and Spanish. *Phonetica*, *56*, 158–185.

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

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

Möbius, B. (2004). Corpus-based investigations on the phonetics of consonant voicing. *Folia Linguistica*, *38*, 5–26.

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

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

Pape, D., & Jesus, L. M. (2014). Production and perception of velar stop (de) voicing in European Portuguese and italian. *EURASIP Journal on Audio, Speech, and Music Processing*, *2014*, 6.

Pind, J. (1995). Speaking rate, voice-onset time, and quantity: The search for higher-order invariants for two Icelandic speech cues. *Attention, Perception, & Psychophysics*, *57*, 291–304.

Poon, P. G., & Mateer, C. A. (1985). A study of VOT in Nepali stop consonants. *Phonetica*, *42*, 39–47.

Ramsammy, M., & Strycharczuk, P. (2016). From phonetic enhancement to phonological underspecification: hybrid voicing contrast in European Portuguese. *Papers in Historical Phonology*, *1*, 285–315.

Reiss, C. (2017). Substance free phonology. In S. J. Hannahs, & A. R. K. Bosch (Eds.), *The Routledge Handbook of Phonological Theory* (pp. 425–452). Routledge.

Ridouane, R., Clements, G. N., & Khatiwada, R. (2010). Language-independent bases of distinctive features. In J. Goldsmith, E. Hume, & L. Wetzels (Eds.), *Tones and Features: Phonetic and Phonological Perspectives* (pp. 264–291). Walter de Gruyter.

Rubach, J. (1990). Final devoicing and cyclic syllabification in German. *Linguistic Inquiry*, *21*, 79–94.

Schwarz, M. (2017). *Realization and representation of Nepali laryngeal contrasts: voiced aspirates and Laryngeal Realism*. Master's thesis McGill University.

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

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

Vaux, B., & Samuels, B. (2005). Laryngeal markedness and aspiration. *Phonology*, *22*, 395–436.

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

Appendix: List of stimuli

[pat]	*leaf*
[phal]	*flower*
[bag^{h}]	*tiger*
[b^{h}at]	*rice*
[tara]	star
[t^{h}al]	*plate*
[dal]	*lentils*
[d^{h}an]	*rice (paddy)*
[kam]	*work*
[k^{h}am]	*envelope*
[gʌp^{h}]	*chat*
[g^{h}ʌr]	*house*