JOURNAL ARTICLE

Contrast implementation affects phonetic variability: A case study of Hindi and English stops

Ivy Hauser

Department of Linguistics & TESOL, University of Texas at Arlington, US

ivy.hauser@uta.edu

There is a large body of work in phonetics and phonology demonstrating sources and structure of acoustic variability, showing that variability in speech production is not random. This paper examines the question of how variability itself varies across languages and speakers, arguing that differences in extent of variability are also systematic. A classic hypothesis from Dispersion Theory (Lindblom, 1986) posits a relationship between extent of variability and phoneme inventory size, but this has been shown to be inadequate for predicting differences in phonetic variability. I propose an alternative hypothesis, Contrast-Dependent Variation, which considers cue weight of individual phonetic dimensions rather than size of phonemic inventories. This is applied to a case study of Hindi and American English stops and correctly predicts more variability in English stop closure voicing relative to Hindi, but similar amounts of lag time variability in both languages. In addition to these group-level between-language differences, the results demonstrate how patterns of individual speaker differences are language-specific and conditioned by differences in phonological contrast implementation.

1. Introduction

It is well-established that phonetic realization of phonological categories is variable both within and between speakers. There is a large body of work on sources of variation in speech production, which include speaking rate, phonetic context, and sociocultural factors (to name only a few). While sources of variation are relatively well-studied, there is less work on factors that condition differences in *extent* of variation. A classic proposal from Lindblom (1986) states that phonetic realization of phones in larger phonemic inventories should exhibit less within-category variation relative to realization of phones in smaller inventories. This hypothesis is based in Dispersion Theory (Liljencrants & Lindblom, 1972), which posits that phonemic inventories are optimized for perceptual distinction. This proposal is intuitive, but existing studies comparing variability in differently-sized phonemic inventories have largely failed to find unqualified support for the prediction (e.g., Bradlow, 1995). In particular, it does not account for the fact that extent of variability can differ across individual speakers and phonetic dimensions (Recasens & Espinosa, 2006), nor does it account for the effects of context and phonological processes (Renwick, 2012).

In this paper, I propose an alternative hypothesis, Contrast-Dependent Variation, which posits a relationship between cue weight of individual phonetic dimensions and extent of phonetic variability. By considering cue weight of individual phonetic dimensions, this hypothesis addresses multiple areas in which the inventory-size approach has been shown to be inadequate. Contrast-Dependent Variation acknowledges the multidimensionality of

phonemic contrast by predicting different amounts of variability across different phonetic dimensions according to their cue weight. This approach also implicitly takes phonological context into account, as cue weights differ across contexts.

I test the predictions of Contrast-Dependent Variation by comparing within-category within-speaker acoustic variation of stop consonants in Hindi and American English, focusing on the dimensions of voiceless lag time (positive voice onset time) and closure voicing. In a laboratory speech production experiment, Hindi and English long lag stops show similar amounts of within-category variation in lag time, but English phonologically voiced stops show significantly more variation in closure voicing relative to Hindi phonologically voiced stops both within- and between-speakers.

These results demonstrate how sources and structure of variability are language-specific and conditioned by differences in phonological contrast implementation. In particular, individual speaker difference as a source of variability differs between-language according to the differences in cue weight. In the results here, structured patterns across speakers and vowel contexts emerge in the English closure voicing data, but not in the Hindi data. While phonological voicing is the best predictor of presence of closure voicing in Hindi, individual speaker is the best predictor of closure voicing in English.

2. Background

2.1 Cue weighting

Speech sound contrasts simultaneously incorporate many co-varying acoustic cues. For example, American English phonologically voiced and voiceless stops can differ on dimensions such as voice onset time (VOT), fundamental frequency (f0) at the onset of the following vowel, and presence/absence of prevoicing. Lisker (1986) notes that at least 16 co-varying acoustic cues are involved in this contrast. While speakers and listeners can make use of many phonetic dimensions in contrast realization, cues differ in their relative strength. Cue weighting quantifies the degree to which an individual dimension contributes to overall perception or production of a contrast.

Weighting cues in production data is frequently done by applying a classification algorithm (e.g., discriminant analysis, logistic regression) where the relevant cues are predictors (e.g., Garellek & White, 2015; Kim & Clayards, 2019). Strength of each predictor is taken to be a metric of cue weight. Differences in cue weighting patterns for the same phonological contrast have been observed in production between native speakers of the same language (Shultz, Francis, & Llanos, 2012), native and non-native speakers (Schertz, Cho, Lotto, & Warner, 2015), non-native speakers with different levels of L2 exposure (Kong & Yoon, 2013), and speakers of a language undergoing sound change (Bang, Sonderegger, Kang, Clayards, & Yoon, 2018; Coetzee, Beddor, Shedden, Styler, & Wissing, 2018; Kuang & Cui, 2018).

Cue weighting in perception is typically examined with perceptual experiments where a single phonetic dimension is manipulated to determine effects on listener discrimination and/or categorization. Cue weighting in perception "tends to reflect community production norms on a broad level" (Schertz & Clare, 2020, p.2), though relative differences in perceptual cue weighting are still observed among individual speakers of the same language (e.g., Chandrasekaran, Sampath, & Wong, 2010; Idemaru, Holt, & Seltman, 2012; Kong & Edwards, 2016; Clayards, 2018). In addition, relative perceptual weights can be manipulated in the lab by altering properties of the training data to which participants are exposed, including extent of within-category variability (Holt & Lotto, 2006; Clayards, Tanenhaus, Aslin, & Jacobs, 2008).

For the purposes of the current proposal, the crucial distinction is between *primary cues*, the cues with the highest relative weight, and *secondary cues*, the cues with lower weights

than the primary cue. While individual variation is to be expected in relative weights of secondary cues, based on previous work on Hindi and English stops (summarized in Section 3) the primary versus secondary status of cues to stop voicing is expected to be consistent across speakers within each language (i.e., individual variation in which cues are primary is not expected).[1] In this paper, I analyze multiple cues to stop voicing in Hindi and English to demonstrate that cue status (primary versus secondary) in each language must be considered to account for differences in extent and structure of phonetic variability.

2.2 Phonetic variability

There is a large body of work on sources of phonetic variability in speech production, which include speaking rate (e.g., Baese-Berk & Morrill, 2015), recent language exposure (e.g., Babel, 2012; Nielsen, 2011), word frequency (e.g., Gahl, 2008; Warner & Tucker, 2011), and speaking style (e.g., Krause & Braida, 2004), to name only a few. While there are many factors which may contribute to phonetic variability in production, this section focuses specifically on individual differences, variability in lab speech, and hypotheses about extent of within-category phonetic variability.

2.2.1 Individual differences

Individual speaker differences are well-documented in many phonetic spaces including vowel formant frequencies among native speakers (K. Johnson, Ladefoged, & Lindau, 1993; Wright, 2004; Ferguson & Kewley-Port, 2007) and L2 learners (Baker & Trofimovich, 2006), voice onset time (Allen, Miller, & DeSteno, 2003; Scobbie, 2006; Theodore, Miller, & DeSteno, 2009; Chodroff & Wilson, 2017), and sibilant center of gravity (Newman, Clouse, & Burnham, 2001; Tabain, 2001), among others. In some cases, the phonetic values for a particular category produced by one speaker may almost entirely overlap with values from a different category produced by another speaker (e.g., Newman et al., 2001; Hillenbrand, Getty, Clark, & Wheeler, 1995, on fricatives and vowels in English, respectively).

A growing body of work demonstrates that although phonetic values from different speakers can show a great deal of overlap, individual variation is often systematic across contrasts and phonetic dimensions. In vowels, cross-category correlations between talker-specific formant values have been observed (Nearey, 1989; Rose, 2010). In stops, Chodroff, Godfrey, Khudanpur, and Wilson (2015) observe correlations in mean VOT values within-speakers across different stop categories of English. Bang and Clayards (2016) observe similar correlations in individual VOT values from different stop categories, and also found correlations between VOT and fricative duration. Tanner, Sonderegger, and Stuart-Smith (2020) examine individual differences in Japanese stop production, and similarly find covariation within cues across speakers. However, they also find that between-cue relationships across speakers are weaker, and suggest that this is because structure in individual differences differs according to language-specific phonological contrast implementation.

This paper tests additional hypotheses about how within-category within-speaker variation might be structured. While previous work has demonstrated that individual differences in phonetic values are often systematic across different phonological categories,

---

[1] It is worth noting that while this may be a reasonable assumption based on previous work on stop contrasts in these languages (see Section 3), individual differences in primary versus secondary cue status are certainly possible, especially in situations involving current sound change (e.g., Kuang & Cui, 2018; Lee & Jongman, 2019; Schertz & Clare, 2020). In addition, while the distinction between primary and secondary cues is clear and consistent for the contrasts examined in this paper, it is not the case that all phonological contrasts in all languages maintain such a clear distinction between primary and secondary cues.

the results here show that language and individual differences in extent of variation are also systematic and conditioned by phonological contrast implementation.

2.2.2 Lab speech and phonetic variability

Speakers adopt the use of clear or hyperarticulated speech in a variety of contexts. Most important to the results in this paper is the use of clear speech in lab contexts. The experiments were laboratory studies in which many potential sources of variation in spontaneous speech (phonetic context, lexical frequency, etc.) were controlled. Differences in lab speech and spontaneous speech are well documented, including a tendency for hyperarticulation by default (Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988; Harnsberger, Wright, & Pisoni, 2008; though see Xu, 2010).

The use of clear speech or hyperarticulation generally involves decreased speaking rate, increased pitch range, and increased acoustic distance between contrasting segments (Picheny, Durlach, & Braida, 1986; Bradlow & Bent, 2002; Smiljanić & Bradlow, 2005). The increased acoustic distance between contrasting segments can affect various acoustic cues depending on the contrast. In English clear speech, VOT increases for voiceless stops but does not change for voiced stops (Chen, 1980; Picheny et al., 1986; J. J. Ohala, 1994a; Krause & Braida, 2004). However, speakers may use different strategies to enhance contrast leading to between- and within-speaker variability even in clear speech situations (Warner & Tucker, 2011).

There has been less work on how extent of variability may differ in clear speech, though clear speech is often assumed to be less variable than relaxed speech,[2] and there is some evidence for this. Chen (1980) describes clear speech as having tighter clustering of vowels within categories. More recently, DiCanio, Nam, Amith, García, and Whalen (2015) find more variability in F1/F2 in spontaneous versus elicited speech in Mixtec. This paper presents results of laboratory studies (where people frequently tend towards clear speech) where differences in extent of variation are present between languages and speakers. This suggests that it is not the case that speakers always minimize within-category variation in clear/laboratory speech contexts.

2.2.3 Hypotheses about extent of within-category variability

Though relatively more work has been done on sources of phonetic variability, multiple theories of speech production do make predictions about extent of phonetic variability. The main hypothesis examined in this paper stems from Dispersion Theory (e.g., Liljencrants & Lindblom, 1972), which posits that phonetic realizations are optimized to preserve contrast for ease of perceptual distinction (see Section 2.3 for a full review of this hypothesis and its predictions about extent of phonetic variability).

There are multiple alternatives to Dispersion Theory which also make predictions about extent of phonetic variability. Under the framework of Quantal Theory, Stevens and Keyser (2010) suggest a relationship between variability and typological frequency—typologically common sounds should require less articulatory precision, resulting in increased articulatory variability. Keating (1983) similarly proposes that some segments, such as sibilant fricatives, may have articulatory targets which are fixed, requiring high levels of articulatory precision. Both predictions about the relationship between articulatory precision and extent of variability have mixed support in the literature (Blake, 2019; Iskarous, Shadle, & Proctor, 2011; Tabain, 2001).

---

[2] While it might be intuitive to hypothesize that clear speech would be less variable than relaxed speech, a reviewer points out that it is also possible that clear speech could be more variable than relaxed speech because it is generally less practiced by talkers. To our knowledge this has not been examined directly.

Other factors influencing extent of within-category variability may be unrelated to contrast or articulatory precision requirements. For example, Vaughn, Baese-Berk, and Idemaru (2018) demonstrate that language background affects extent of phonetic variability, with the phonological system of an L1 potentially having a systematic effect on extent of variability in an L2. Work in sociolinguistics also shows differences in extent of variability according to the range of speech styles utilized by different speakers (e.g., Eckert, 2000). Following this, Sonderegger, Bane, and Graff (2017) suggest that individual differences in 'phonetic plasticity' observed in their medium-term study of spontaneous English speech may therefore be due to individual differences in style shifting. In the present study, all speakers were native speakers of the target language and all data come from a laboratory speech task, where it is unlikely that speakers will be engaging in style shifting within the task itself.

Overall, multiple frameworks make predictions about differences in extent of variability between speakers and languages, but the literature provides generally mixed evidence, with no theory able to account for all observed differences in extent of variability. While there are many potential reasons why extent of variability might differ across individual speakers, the main focus of the current paper is to examine factors conditioning differences in extent and sources of variability between languages. Previous hypotheses from Dispersion Theory are most relevant for this question, as it makes predictions about between-language differences in extent of variability.

2.3 Dispersion Theory

Dispersion Theory (DT; Liljencrants & Lindblom, 1972; Lindblom, 1986; Schwartz, Boë, Vallée, & Abry, 1997) was originally formulated to make predictions about the relative typological frequency of vowel inventories. The intuition behind DT is that vowel spaces are optimized for perceptual distinction. Liljencrants and Lindblom (1972) propose maximal contrast as an organizing principle in vowel inventories, defining the vowel space using two phonetic dimensions: F1 and F2', which is a combination of F2 and F3. Their model correctly predicts cross-linguistic frequency of /i a u/ for three-vowel inventories but has more discrepancies with predictions for larger inventories.

Several updates have been made to the original formulation of DT to address these and other discrepancies. Lindblom (1986) adds multiple revisions including the concept of sufficient instead of maximal dispersion. Dispersion from sufficient contrast predicts that languages with more phonological categories should occupy an overall larger phonetic space and have tighter categories (i.e., less within-category variation) within that space. Schwartz et al. (1997) later expand the dispersion calculation used in Liljencrants and Lindblom (1972) to include intra-vowel spectral information in addition to inter-vowel distances. Their Dispersion-Focalization Theory (DFT) includes an energy function with two perceptual components: a global dispersion term quantifying inter-vowel distance, and a local focalization term quantifying intra-vowel spectral salience based on proximity of formants.

DT ideas have also been formulated in Optimality Theory (OT; Prince & Smolensky, 1993/2004) with the use of constraints that explicitly demand distance between members of a phonological inventory (e.g., Flemming, 1996). The goals of the system are to maximize distinctiveness of contrast while minimizing articulatory effort and maximizing the number of total contrasts in the system. OT formulations of DT have successfully been used to model palatalization (Padgett, 2001) and vowel reduction (Padgett & Tabain, 2005; Flemming, 2004), but the approach has been criticized for being 'teleological' (Boersma & Hamann, 2008). Constraints that demand types of contrast must evaluate entire inventories/languages, departing from traditional formulations of OT.

Other approaches have attempted to derive emergent dispersion effects, rather than explicitly demanding dispersion by grammatical or other factors. For example, Boersma and Hamann (2008) show that when production and perception are modeled with bidirectional phonetic cue constraints, dispersion emerges without constraints specifically demanding it. They note that change towards a more dispersed inventory has been observed in diachronic change between medieval and present Polish sibilants. Hall (2011) uses contrastive feature specification to show that dispersion effects can emerge when phonological representations only specify contrastive features. Predictable features are then enhanced during phonetic realization, creating a dispersion effect.

Engstrand and Krull (1994) extend the ideas of DT to account for cross-linguistic differences in durational correlates of vowel quantity, departing from previous DT work focusing largely on vowel quality. They use the DT idea of sufficient contrast to explain their results showing that Estonian and Finnish speakers preserve the durational correlate of quantity more relative to Swedish speakers. This is because quantity contrasts are mainly based on duration in Finnish and Estonian while Swedish quantity is also correlated with vowel quality and diphthongization. They argue that contrasts which use a highly exploited feature dimension should require more 'precise signal information' relative to contrasts on a less exploited dimension. This is supported with their data, which shows more between-category dispersion in the durational dimension in Estonian and Finnish relative to Swedish. I make a similar argument in this paper, but concerning within-category dispersion rather than between-category dispersion. There should be relatively less within-category variation when a dimension is employed as a primary cue (i.e., a highly exploited feature dimension) relative to when it is employed as a secondary cue.

2.3.1 Dispersion Theory and consonants

DT was originally formulated to make predictions about vowel inventories (as in Liljencrants & Lindblom, 1972), but the issue of dispersion among consonants has also been investigated. If speakers aid listener perception by constraining variation in crowded phonetic spaces, we might expect to see the prediction hold for all types of speech sounds.

Most of the literature on typological frequency of consonant inventories has revolved around maximal use of available features, proposed as an organizing principle by J. Ohala (1979), and later formalized by Clements (2003) as Feature Economy. The economy model does predict the ubiquity of the typologically common /bilabial-coronal-velar/ stop system in actual and randomly generated inventories (Mackie & Mielke, 2011). The principle of feature economy differs from the principle of maximizing perceptual distinction in DT. J. Ohala (1979) claims that maximizing perceptual distinction would result in consonant inventories like [ɓ k' ts ɬ m r ɟ], which are unattested. This claim is countered by Lindblom (1986), who suggests that it need not necessarily be the case that vowel and consonant systems are organized by different principles when considering sufficient instead of maximal contrast.

In a further elaboration of the idea of sufficient contrast as a organizing principle in consonant inventories, Lindblom and Maddieson (1988) propose a relationship between consonant inventory size and complexity of consonant articulation. They divide consonants into three sets: basic articulations, elaborated articulation, and complex articulations (combinations of elaborated articulations). These sets are proposed to correlate with inventory size; smaller inventories typically only use basic articulations, and larger inventories make use of the elaborated and complex articulations.

Despite the focus on alternative organizing principles, some work has shown evidence for acoustic dispersion in consonants. Boersma and Hamann (2008) propose a framework

in which dispersion is emergent in sibilant systems by modeling perception and production with bidirectional cue constraints. In work on stops, Schwartz, Boë, Badin, and Sawallis (2012) examine inventory dispersion using a large data set of 50,000 stop tokens generated from a vocal tract model. They claim that the typologically common stop consonant inventory /b d g/ should be viewed as a perceptually optimal and dispersed structure just like the typologically common vowel inventory /i a u/. However, their results show that pharyngeals or epiglottals should be included in the most dispersed inventory "in terms of raw acoustic dispersion" (p.28). They argue that the space which should be considered is modulated by articulatory considerations, namely Frame-Content Theory (MacNeilage, 1998), which functionally excludes pharyngeal and epiglottal stops from the dispersion calculations. In this revised space, Schwartz et al. are able to revive the dispersion account as a major factor contributing to stop system organization.

This literature suggests that while dispersion can be argued to be an organizing principle in consonant inventories, extending the ideas from vowel inventories is not straightforward. Existing work on consonant dispersion has also mostly focused on the DT predictions about between-category dispersion and inventory organization, rather than the related prediction about within-category dispersion and extent of variability, which is examined here.

2.3.2. Previous work on within-category variation and DT

Lindblom (1986, p. 33) proposes an intuitive hypothesis about the relationship between phonological contrast and phonetic variation in vowel inventories: "the phonetic values of vowel phonemes should exhibit more variation in small than in large systems." This hypothesis assumes that distributions must be tightened in a more crowded space to avoid overlap between categories and preserve perceptual distinction. A language with relatively fewer categories exploiting a single phonetic space has room for within-category variation while maintaining separation between categories. The prediction arises from the assumption that speakers aid listener perception by producing speech sounds that are sufficiently (but not maximally) perceptually distinct.

Most of the work investigating the hypotheses in Lindblom (1986) has focused on the related prediction about inventory size, rather than the prediction about variation. The prediction that larger vowel inventories should occupy larger phonetic spaces is supported by data from comparisons between German (14 vowels) versus Greek (5) (Jongman, Fourakis, & Sereno, 1989) and English (11) versus Spanish (5) (Bradlow, 1995), as well as a large-scale typological corpus study by Becker-Kristal (2010). However, Livijn (2000) compares 28 languages and finds that languages with 4-8 vowels have comparably sized phonetic spaces and space only increases with 11 or more vowels. Similarly, Gendrot and Adda-Decker (2007) compare the vowel spaces of eight languages and find that larger inventories do not have expanded vowel spaces. In addition, Recasens and Espinosa (2006) examine multiple dialects of Catalan and find that the maximal formant range of point vowels is constant across dialects, regardless of inventory size. However, distances between individual vowels do vary according to dialect and vowel pair, which they argue provides partial support for DT predictions. While earlier work on tone systems includes evidence of larger tone systems using a relatively larger F0 space (Maddieson, 1977), Alexander (2010) compares the tone spaces of five languages and finds that tone space size differed as a function of type of tone language (e.g., level versus contour-tone systems) rather than number of tones.

Specific investigations of Lindblom's prediction about within-category variability have differed in terms of speech sounds and types of variation examined. Some work has focused

on token-by-token variability within a single phonological context.[3] Bradlow (1995) compares vowels in English (14 vowels) and Spanish (5), and does not find any significant between-language differences in extent of within-category variability. In contrast, Blake (2019) does find support for an effect of inventory size when comparing variation of /s/ in Spanish relative to English and Catalan, which have larger sibilant inventories. Results show significantly more variation in /s/ center of gravity in Spanish, despite previous claims that sibilants generally require high articulatory precision (Keating, 1983).

Another line of work has examined the DT variation prediction from the angle of coarticulatory variation. Manuel (1990) compares extent of vowel-to-vowel coarticulation in data from Ndebele and Shona (5 vowels) with data from Sotho (7 vowels), and finds less anticipatory coarticulation in Sotho. Following DT predictions, Manuel proposes that the extent of vowel-to-vowel coarticulation is less in languages with larger inventories where coarticulation may cause confusion of contrastive phones. Later work by Renwick (2012), however, demonstrates that these effects are sensitive to phonological context and alternations, and advocates a more nuanced approach. Renwick compares vowel coarticulation in Romanian (7 vowels) and Italian (5 or 7 vowels depending on the analysis) and finds more coarticulatory variability in Romanian. However, Romanian's phonological processes can account for the exaggeration of coarticulatory effects.

Renwick categorizes two different types of variability. Context-dependent variability is triggered by different coarticulatory contexts and therefore somewhat predictable. Italian exhibits less coarticulation, and therefore less context-dependent variability relative to Romanian. Context-independent variability is the measure of precision in which productions reach their acoustic targets. When coarticulatory context is taken into account, Italian shows greater context-independent variability relative to Romanian. Neither the predictions of DT nor Manuel (1990) completely line up with these results, which demonstrate that vowel inventory size alone cannot predict levels of coarticulation and phonological processes must also be taken into account.

Recasens and Espinosa (2006) similarly examine both contextual and token-by-token variability in multiple dialects of Catalan, and find that patterns of variability differ across vowels and formants, and are not directly related to inventory size. They argue that contextual variability is related to articulatory requirements of vowel production while token-by-token (context-independent) variability is related to precision in hitting a target context for a particular vowel in a particular context. They do not observe overall less variability in Majorcan, which adds a schwa to the other dialect's 7-vowel systems. Rather, they only observe less variability in Majorcan mid low vowels, partially confirming the predictions of DT. They suggest that Majorcan schwa is specified for a mid central target, which causes repulsion of peripheral vowels near the mid central region.

Following Recasens and Espinosa (2006) and Renwick (2012), I also demonstrate that inventory size is not enough to predict variability patterns, adding a case study from stop consonants, and proposing an alternative which considers cue weight instead of inventory size. Recasens and Espinosa's finding that patterns of variability differ across individual segments and cues suggests the need for an approach like Contrast-Dependent Variation, which evaluates each phonetic dimension separately. As different factors contribute to contextual variability versus token-by-token variability, Contrast-Dependent Variation is only intended to account for relative differences in token-by-token variability (i.e., non-contextual, context-independent). While I do not compare rates of variability across contexts in this paper, focusing on cue weights instead of inventory sizes implicitly takes

---

[3] Either by only collecting data in a single context or by including phonological context as a predictor in statistical modeling.

phonological context into account, as cue weights differ across phonological contexts. This follows Renwick (2012), who demonstrates that phonological context influences patterns and extent of variability.

2.3.3 Phonetic spaces in Dispersion Theory

Most work on DT carries implicit assumptions about the relevant space for understanding dispersion. The space for analysis is often assumed to be a subset of the phonemic inventory defined by a shared phonological feature. For example, work on consonant dispersion looks for dispersion within consonant inventories (rather than, e.g., between consonants and vowels). The spaces in which dispersion is examined are often subsets of the consonant inventory as in Boersma and Hamann (2008) with voiceless sibilant fricatives and Schwartz et al. (2012) with voiced stops. As with vowel inventories, these subsets are defined (either implicitly or explicitly) by phonological features which refer to particular segment classes.

The approach here differs from previous approaches as the predictions of Contrast-Dependent Variation refer to individual phonetic dimensions instead of phoneme inventory subsets. The focus on phonetic dimensions instead of inventory size provides an alternative which captures the fact that speech sound contrasts are multidimensional, with differing cue weights across sounds and languages. The hypothesis is general and testable across multiple types of speech sounds, allowing for investigation of the relationship between cue weight and variability along potentially any phonetic dimension regardless of whether that dimension is a cue to consonant or vowel contrasts.

3 Predictions

3.1 Hindi background

Hindi is one of several Indo-Aryan languages which exhibit a four-way laryngeal contrast on stops (**Table 1**). Dutta (2007) cites UPSID (Maddieson & Disner, 1984) which contains ten languages from six families with the four-way contrast. In Hindi, the four-way contrast occurs at four places of articulation: bilabial, dental, retroflex, and velar. Voice onset time (VOT) has frequently been analyzed as a phonetic correlate to these stop contrasts (Lisker & Abramson, 1964; Abramson & Lisker, 1967; Poon & Mateer, 1985). VOT is a duration measure of the onset of voicing relative to the release of the stop occlusion, and is often implemented as a continuum of negative and positive values. Lead voicing before the stop closure is coded as negative VOT and lag voicing which begins after the stop closure is coded as positive VOT (Lisker & Abramson, 1964; Cho & Ladefoged, 1999).

Using the single VOT measure for lead and lag voicing has been challenged. In particular, VOT has been recognized as inadequate for languages like Hindi which have stops that are produced with lead voicing and aspiration (Lisker & Abramson, 1964; Schiefer, 1986;

Table 1: Consonant inventory of Hindi (M. Ohala, 1983).

	Labial	Dental	Retroflex	Palatal	Velar	Glottal
Stop	p b	t d	ʈ ɖ		k g	
Aspirated stop	pʰ bʰ	tʰ dʰ	ʈʰ ɖʰ		kʰ gʰ	
Affricate				tʃ dʒ		
Fricative	f v	s z		ʃ		h
Nasal	m	n		ɲ	ŋ	
Rhotic		r	ɽ			
Approximant		l		j		

Dixit, 1989). Mikuteit and Reetz (2007) use data from East Bengali (another language with a four-way contrast) to argue that lead voicing and lag voicing should not be considered part of the same continuum. They instead propose separate duration measures of after closure time (duration from release to onset of voicing; lag time), onset voicing (start of glottal pulsing to release in initial stops), and connection voicing (closure duration in medial stops).

Following this analysis, I consider lag time (traditionally known as positive VOT) to be a separate phonetic dimension from lead time (traditionally known as negative VOT). In this paper, I use *lag time* to indicate the duration between the stop burst and onset of voicing, *closure duration* to indicate the duration of the stop closure, and *closure voicing* to indicate the duration of periodic voicing during the stop closure. See Section 6 for further discussion of how voicing was measured and analyzed in the production experiment.

In terms of distinctive features, Hindi is typically described as fully crossing all values of two features [± voice] and [± spread glottis] (Dutta, 2007), shown in **Table 2**. I will refer to instances of [+voice] as *phonologically voiced stops* and instances of [+spread glottis] as *phonologically aspirated stops*. There is some debate in the literature about the exact feature specification of the voiced aspirates (e.g., Benguerel & Bhatia, 1980; Dixit, 1989; Dutta, 2007). There is also debate about whether these features should be binary or privative, a controversy which is not specific to Hindi (e.g., Honeybone, 2005; Schwarz, Sonderegger, & Goad, 2019). The questions examined in this paper do not hinge on any particular feature representations and I discuss implications for feature specification in Section 6.4.1.

3.2 English background

American English has two contrasting stop consonants at three places of articulation: bilabial, alveolar, and velar. These can be seen in the English consonant inventory given in **Table 3**. In American English, lag time is the primary cue to the stop contrast and other phonetic cues such as F0 frequently co-vary with lag time (e.g., Lisker & Abramson, 1967; Keating, 1984; Lisker, 1986). Because lag time is the primary cue, English is often considered to be an aspirating language instead of a true voicing language. Despite this, phonemic representations of English typically use the IPA symbols for voiceless and voiced stops /t d/.

Table 2: Feature specifications for stops in Hindi.

	[–spread glottis]	[+spread glottis]
[–voice]	/t/	/tʰ/
[+voice]	/d/	/dʰ/

Table 3: Consonant inventory of English (e.g., Quirk, Greenbaum, Leech, & Svartvik, 1972).

	Labial	Dental	Alveolar	Post-alveolar	Palatal	Velar	Glottal
Stop	p b		t d			k g	
Affricate				tʃ dʒ			
Fricative	f v	θ ð	s z	ʃ ʒ			h
Nasal	m		n			ŋ	
Approximant			l ɹ		j	w	

There is some disagreement on which phonological features should be used to distinguish English stops. Laryngeal realism takes the position that phonological features should reflect phonetic realization in word-initial (or another prominent) position (e.g., Jessen & Ringen, 2002; Honeybone, 2005; Beckman, Jessen, & Ringen, 2013). Under this view, the feature distinguishing the two English stops is [spread glottis] (features are also typically privative in laryngeal realism). The laryngeal relativist view takes a more abstract approach focusing on cross-linguistic similarities. In this view, two-way stop contrasts are typically represented with [(±)voice], and phonetic implementation can differ across languages (e.g., Keating, 1984; Kingston & Diehl, 1994; Lombardi, 1994; Cyran, 2011). **Table 4** shows these potential representations of the English stop phones and their common phonetic realizations in word-initial position. In this paper, I will assume the [±voice] analysis and revisit the question of feature representation in Section 6.4.1. In all discussion that follows, I refer to the English short lag stops /b d g/ as phonologically voiced and the English long lag stops /p t k/ as phonologically voiceless.

Despite the use of voiceless lag time as the primary cue to stop voicing in English, many studies have reported prevoicing on English phonologically voiced stops, which is assumed to be the primary cue to voicing in true voicing languages. Flege (1982) summarizes previous studies (Lisker & Abramson, 1964; Lorge, 1967; Zlatin, 1974; Smith, 1978; Westbury & Niimi, 1979) in which 20–57% of English stops are produced with prevoicing, and also reports results in which more than half of all phonologically voiced stops produced by ten male speakers of American English were produced with prevoicing. Docherty (1992) reports prevoicing incidence from five male speakers of British English. In that study, on average, duration of voicing during stop closures was 51% (of closure) for [b], 58% for [d], and 66% for [g]. Deterding and Nolan (2007) also find similar results in a later study of seven British English speakers.

In more recent work, Davidson (2016) documents prevoicing on American English stops, and finds prevoicing variation in connected read speech to be influenced by linguistic factors such as adjacent sounds and lexical stress. There is also a growing body of work in sociolinguistic literature documenting prevoicing in Southern American English varieties (in utterance-initial and medial contexts), sometimes with higher incidence among male and African-American speakers (Jacewicz, Fox, & Lyle, 2009; Elston et al., 2016; Herd, Torrence, & Carino, 2016; Hunnicutt & Morris, 2016). Overall, previous work on production of stop voicing in American English suggests that use of prevoicing is common but inconsistent, with potentially higher incidence in particular varieties and phonological contexts.

Prevoicing has been shown to influence perception of English stop voicing contrasts in syllable-final position (Hillenbrand, Ingrisano, Smith, & Flege, 1984; Hogan & Rozsypal, 1980; Wardrip-Fruin & Peach, 1984). Pisoni, Aslin, Perey, and Hennessy (1982) also demonstrate that English speakers can learn to reliably discriminate between word-initial prevoiced stops and voiceless unaspirated stops in the lab with only a few minutes of exposure training. While lag time is consistently shown to be the primary cue to the word-initial voicing contrast in American English (e.g., Lisker & Abramson, 1967; Keating,

Table 4: Representations of English stops.

Phoneme	Distinctive feature (realist)	Distinctive feature (relativist)	Common phonetic realizations in word-initial position
/b/	–	[+voice]	[b p b̥]
/p/	[sp. gl.]	[–voice]	[pʰ]

1984; Lisker, 1986), these results suggest that prevoicing has at least some degree of perceptual relevance as a secondary cue.

3.3 Predictions for the present study

In this section, I compare the predictions of Contrast-Dependent Variation with an application of Lindblom's (1986) DT hypothesis about the inverse relationship between inventory size and extent of phonetic variability. The two hypotheses are summarized in **Table 5**.

The main assumption behind DT is that phonetic realizations are optimized to preserve perceptual distinction, which can be done by increasing between-category dispersion and avoiding category overlap. Under this framework, Lindblom (1986) makes a concrete prediction about the relationship between within-category variation and inventory size. As in most work in DT, the focus is on vowels, but the general assumption of preserving perceptual distinction could also be applied to consonants (see Section 2.3 for a review of previous work in this area). In extending Lindblom's hypothesis about variation to consonants, we might consider the stop inventory to be the relevant 'system' as Lindblom considers the vowel inventory to be the relevant 'system' (Lindblom, 1986, p. 33). Lindblom's hypothesis relies on inventory size to make predictions, and does not distinguish between phonetic dimensions. If the stop inventory is understood to be the relevant system, Lindblom's hypothesis predicts less variation in Hindi relative to English because Hindi has a larger stop inventory. One particular prediction that could be drawn from this is that we would expect voiceless aspirated stops in Hindi to vary less in lag time. Expected results under this prediction are shown in **Figure 1**.

I advocate for an alternative approach which considers individual phonetic dimensions to be the relevant 'system,' using relative differences in cue weight to make predictions about relative differences in extent of variability (Contrast-Dependent Variation; Hauser, 2019). For a given phonetic dimension, we expect less variability in languages where that dimension is used as a primary cue relative to languages where that dimension is used as a secondary cue. Under this hypothesis, no difference in lag time variation is expected between the two languages. This is because because both languages employ lag time as a primary cue for distinguishing short and long lag stops in word-initial utterance-medial position, the context of elicitation in this study. However, we do expect more voicing variation in English relative to Hindi, as Hindi uses closure voicing as a primary cue in this context. While closure voicing does co-vary with lag time (and other cues) as a secondary cue for English stop contrasts, English does not use closure voicing as a primary cue to distinguish any phonological contrasts. The primary cues of Hindi and English are sketched in **Table 6**.

Table 5: Hypotheses about within-category variation summarized.

| Hypothesis	Summary	Domain	Phonetic space
Lindblom (1986, p. 33)	"The phonetic values of vowel phonemes should exhibit less variation in small systems than in large systems."	vowels	F1/F2 assumed
Contrast-dependent variation (Hauser, 2019, p. 27)	For a given phonetic dimension X, we expect less group-level within-speaker variability in languages in which X is employed as a primary cue relative to languages in which X is employed as a secondary cue.	general	any
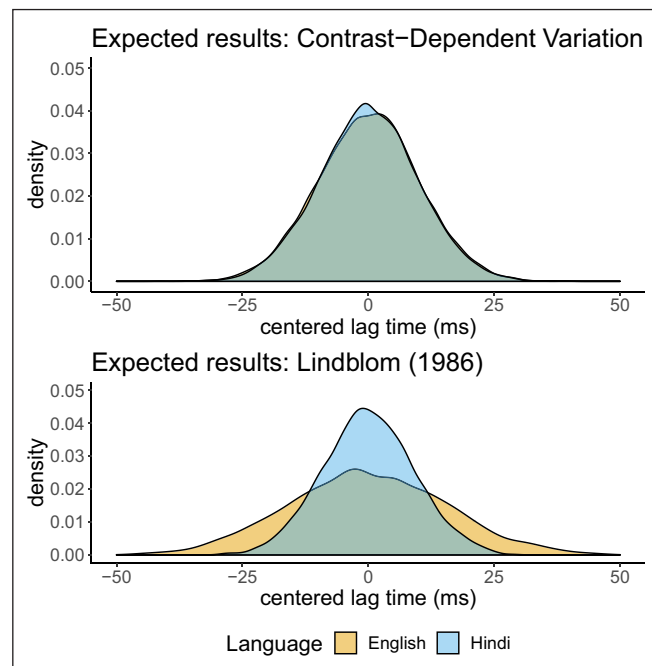
Figure 1: Predicted voiceless lag time distributions for two hypotheses in Hindi and English. Distributions pictured are schematic examples of possible distributions predicted under each hypothesis, generated with the rnorm function from the stats package in base R (R Core Team, 2013).

Table 6: Primary cues in Hindi (top) and English (bottom) stops.

↑	← voiceless aspiration (lag time) →	
closure	/t/	/tʰ/
voicing	/d/	/dʰ/
↓	←voiced aspiration →	
	← voiceless aspiration (lag time) →	
	/d/	/t/

4 Experimental design

4.1 Participants

All speakers were between the ages of 18-30 and recruited from student populations of the University of Massachusetts Amherst. Most of the English speakers were undergraduates enrolled in introductory linguistics courses and most of the Hindi speakers were graduate students in various fields. In the first round of data collection, nine speakers of each language were recorded.

The task was a production task which involved reading phrases off a computer screen. Therefore, native speakers with poor reading skills spoke unnaturally during the task and produced many speech errors. Any participants who expressed difficulty with the task and/or paused before the stimulus leaving silence for more than 1.5 seconds on at least 75% of the phrases were removed from the analysis. Five Hindi speakers and one English speaker were excluded according to these criteria. Two Hindi speakers were additionally removed from the analysis because they were L2 speakers of Hindi (which was determined by their answers to a demographic questionnaire about language background). Two English speakers were additionally removed because they did not complete the task. After exclusions, data from two Hindi speakers from the first round of data collection were retained.

To replace the Hindi speakers which were excluded in the first round, a second round of data collection was conducted with a few adjustments. The call for participants was circulated only in Hindi orthography to ensure the participants were comfortable with reading in addition to speaking. Additionally, the experimenter was always a native speaker of Hindi who only spoke Hindi to the participants throughout the experiment. This helped in resolving confusion among the participants about L1/L2 status of Hindi before they participated. These were the only differences in the procedure of the experiment between the first round and the second round of data collection. The two speakers whose data were retained from the first round of collection did not systematically differ in extent of lag time or closure voicing variance relative to those in the second round of collection. After the second round of data collection, recordings from six speakers of each language were available for analysis.

4.2 Stimuli

The goal was for stimuli to be as similar as possible between the two languages. The stimuli were C_1VC_2 words and non-words where C_1 was a stop and V was one of [i a u].[4] The coda consonant of the stimulus (C_2) was in most cases a stop. If there were no stops available that could make a phonotactically natural word or non-word, then a fricative was used. If there were no fricatives available, then a sonorant was used. Eliciting only monosyllabic words avoided any effects of stress placement. All stimuli were recorded in a uniform carrier phrase: "Say X again" in English and "Dobara X doharao" (repeat X again) in Hindi. The carrier phrases placed the target words in focused environments in both languages. The stimuli were all developed in consultation with native speakers to assure phonotactic wellformedness.

Real words and non-words were used in both the Hindi and English stimuli. Hindi stimuli were crossed according to the following factors: consonant (16 levels) × vowel context (3 levels) × word status (2 levels: word/non-word) for a total of 96 distinct stimuli. English stimuli were crossed according to: consonant (6 levels) × vowel context (3 levels) × word status (4 levels: high frequency/low frequency/non-word/has C_1 minimal pair) for a total of 72 distinct stimuli. Example stimuli are given in **Table 7**.

The English stimuli were crossed according to word frequency and minimal pair status, using data obtained from the English Lexicon Project (Balota et al., 2007). At the time of initial data collection, similar lexical statistics were not readily available for Hindi, so quantitative word frequency data was only included in the initial English stimuli selection. However, such materials have become available in the time since data collection, allowing for a post-hoc analysis of word frequency effects in both languages, using Hindi data from WorldLex (Gimenes & New, 2016). The analyses of lexical statistics in both languages showed no significant effect of word status or word frequency on lag time or closure voicing (statistical models are provided in the Appendix). Therefore, I do not include word status or frequency statistics as factors in any of the analyses that follow. Statistical models do include item as a random effect, when appropriate, to account for any idiosyncratic effects of particular words.

---

[4] These transcriptions implicitly assume a low-back merger between [a]/[ɑ] and [ɔ] for some of the English stimuli, such as 'bog,' by categorizing those vowels with [a] (potentially affecting 4 items out of 66 total stimuli). All English speakers acquired English in New England, where the low-back merger is common but variable (e.g., Clopper, Pisoni, & De Jong, 2005; D. E. Johnson, 2007). All words were presented to participants in English orthography, so any participants with a low-back distinction may have used something closer to [ɔ] in some words. However, we would not expect these vowel quality differences to systematically affect preceding stop voicing or lag time. In addition, random effects for item were included in initial statistical models (when appropriate), which would capture any idiosyncratic effects associated with particular words.

Table 7: Example stimuli.

Language	C$_1$	vowel	stimulus (IPA)	lexical status
Hindi	b	i	bit	word
Hindi	k^h	i	k^{h}il	word
Hindi	b^h	u	b^{h}ut	word
Hindi	d	a	dag	word
English	p	i	pis	word-hi
English	t	a	tab	non-word
English	t	u	tub	word-hi
English	b	a	bat	word-low

4.3 Recording

The participants were all recorded in a sound-attenuated booth using Audacity software (Audacity Team, 1999–2021). The recordings were done using an M-Audio Fast Track Pro Mobile Audio Interface and a Shure SM10A head-worn microphone. The recordings were sampled at a rate of 44.1 kHz with a bit depth of 16. The participants were presented with stimuli in the relevant orthography on a laptop computer inside the booth. They were asked to produce the phrases as naturally as possible. All experimenters were trained to give feedback which encouraged natural production.[5] The stimuli were recorded in four separate blocks, each with a different random order, totaling four repetitions of each stimulus for analysis.

The recordings from each speaker were first scanned by the author and/or a native speaker research assistant for speech errors. After speech error exclusions, there were a total of 3663 tokens available for analysis. The recordings were force aligned using the Montreal Forced Aligner (McAuliffe, Socolof, Mihuc, Wagner, & Sonderegger, 2017), which creates Praat (Boersma, 2001) textgrids marking boundaries at the word and segment level. I used the English pre-trained model (originally trained on the LibriSpeech corpus) for aligning the English data. The dictionary of the model was updated with the addition of non-words used in our stimuli. No pretrained model was available for Hindi, but MFA also allows for alignment using only the data set. I used this feature to train a model on the Hindi data and align the Hindi data. The aligned text grids in both languages were spot checked for accuracy. More detailed hand adjustments were not yet done at this stage as none of these boundaries would be directly used to extract any measurements.

5 Lag time

In this section, I detail how lag time was analyzed for the phonologically voiceless stops in both languages and summarize the comparative lag time results. In accordance with the predictions of Contrast-Dependent Variation, there was no significant difference in extent of group-level within-speaker lag time variability between the two languages.

5.1 Analysis

Many dialects of Hindi are currently undergoing (or have undergone) a merger between the voiceless aspirated labial stop /p^h/ and the voiceless labiodental fricative /f/, where both are produced as [f] (Dutta, 2007). All of the speakers in this study consistently

---

[5] This included things like suggesting the participant speak as if they were talking to a friend and not giving a presentation, suggesting they say the phrase "in one breath" to discourage pausing before the stimulus, etc.

produced the fricative, so I only compare the coronal and velar stops in this paper. The coronal category includes the dental and retroflex stops in Hindi and alveolar stops in English. Results do not change if the English alveolar stops are compared with only the dental stops or only the retroflex stops.

The force aligned textgrids were used as input to AutoVOT (Keshet, Sonderegger, & Knowles, 2014) which allowed for automatic measurement of lag time intervals. Prior to running AutoVOT, I extended the MFA boundaries of each long lag stop by 31ms on each side to create the intervals in which AutoVOT would measure VOT, following Chodroff (2018). I then ran AutoVOT, measuring lag time from the start of the burst to the onset of voicing. This procedure was used to measure lag time for the voiceless short and long lag stops in both languages. The intervals created by AutoVOT were all hand-checked and hand-corrected as needed by the author or a trained research assistant.

Example tokens are shown in **Figures 2–3**. In both figures, the short lag tokens are on the left and the long lag tokens on the right. As expected, a difference in the duration of aspiration between the short and long lag tokens can be seen in both languages. This section focuses on lag times for the long lag stops.

To abstract over differences in mean values between speakers and vowel contexts, lag time values were centered around means within-speaker, within-category, and within-vowel context. A standard outlier rejection method was applied before analysis, excluding tokens with a *z*-score greater than |3| (Well, Myers, & Lorch, 2010). This removed 33 of 3663 total tokens.

5.2 Results

In **Figure 4**, I show the distribution of lag time values for long lag stops in both languages at coronal and velar places of articulation. These plots use the centered lag time values, collapsed over speakers. Lindblom's hypothesis predicts less within-category variation in
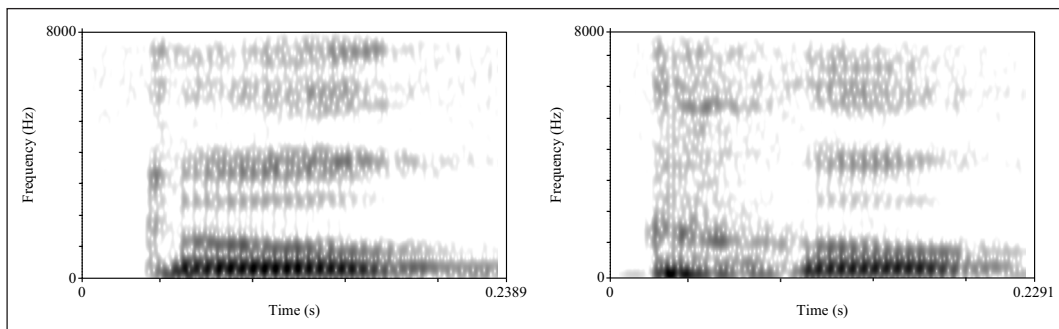
Figure 2: Voiceless short and long lag stops in Hindi. Left: CV sequence from token of /tup/. Right: CV sequence from token of /tʰup/.
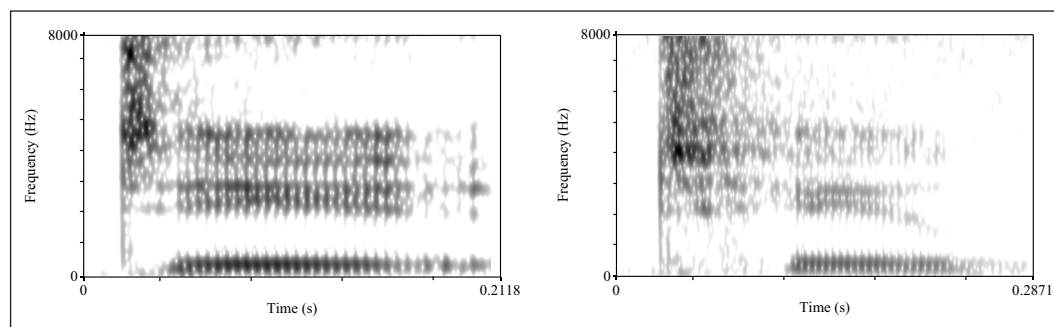
Figure 3: Short lag and long lag stops in English. Left: CV sequence from token of /dit/. Right: CV sequence from token of /tip/.
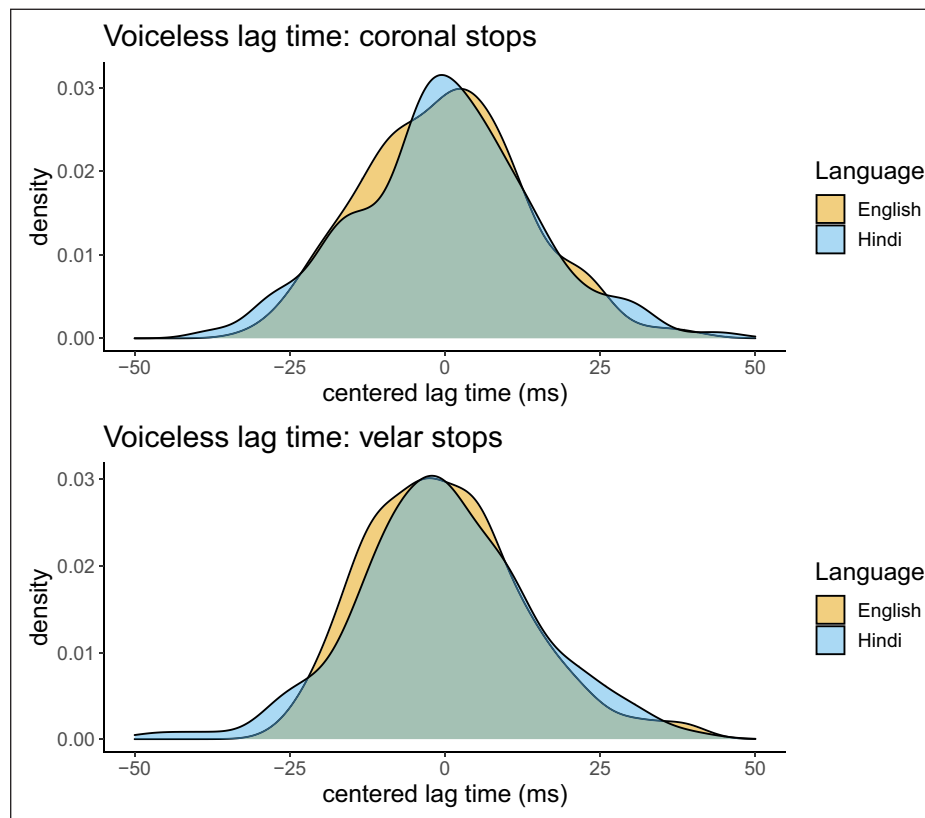
Figure 4: Experimental results for coronal and velar long lag stops. Lag time values are centered within speaker and vowel context.

Hindi (expected results shown in **Figure 1**). If this were the case, the English distributions would be wider than the Hindi distributions in the results. However, in **Figure 4**, the English distributions do not appear to be wider than the Hindi distributions for either place of articulation. In fact, it appears that the Hindi data might actually be slightly more variable than the English data, though this difference is insignificant.

To quantify the effect of language, I use a mixed effects linear regression where within-speaker within-category lag time variance is the dependent variable. This follows Vaughn et al. (2018) who used within-category variance as a dependent variable to test for differences in group-level within-speaker variability. Variance was calculated within-speaker within-category and within vowel context (e.g., variance in speaker e-02's productions of /t/ before /i/, etc.), over about 40 tokens in each condition. The number of tokens differs slightly across conditions because a small number of tokens were excluded, and participants occasionally skipped stimuli. The coefficient of variation was then calculated over the tokens in each condition, resulting in 90 observations of within-speaker within-category variance.

Language, Place of Articulation, and Vowel Context were also included as fixed effects with random intercepts for speaker. Although Language is the main effect of interest, other factors were included to ensure that a significant effect of language would not be due to covariation with other factors. It is possible that stop place of articulation and vowel quality may independently influence extent of variation. No random slopes for speaker were included as this additional model structure was not justified by the research question. I am interested in the main effect of language, and speaker is fully nested within language. R (R Core Team, 2013) was used for all statistical analyses. The lmer function in the lme4 package (Bates, Sarkar, Bates, & Matrix, 2007) was used for the regression model, with LmerTest to obtain *p* values (Kuznetsova, Brockhoff, & Christensen, 2017). Place was

Table 8: Fixed effects table for linear mixed effects regression. Dependent variable: within-category within-speaker lag time variation (quantified by coefficient of variation). Predictors: language, place of articulation, V (vowel context), random intercepts for speaker. Model intercept is English coronal _/a/ context. Standard error values are similar between effects when there is similar n in each level of the factor.

Fixed effects	Estimate (*SE*)	95% CI	df	t	*p*
(Intercept)	16.15 (2.25)	11.83 – 20.46	15.49	7.18	<0.001***
Language-Hindi	2.63 (2.88)	−2.97 – 8.23	10.56	0.91	0.380
Place-velar	−1.18 (1.06)	−3.88 – 0.25	75.45	−1.71	0.091
V-/i/	−2.56 (1.26)	−5.00 − −0.11	75.45	−2.04	0.050*
V-/u/	−3.32 (1.26)	−5.78 − −0.87	75.45	−2.64	0.010*

coded as a categorical variable with two levels: coronal and velar. Default dummy coding contrast structure was used with English coronal _/a/ context as the reference level.

Lindblom's (1986) hypothesis predicts less group-level within-speaker variation in Hindi relative to English, therefore we would expect a significant effect of Language in the model. Under Contrast-Dependent Variation (proposed here) we do not expect this difference in group-level within-speaker variation, therefore we would expect no significant effect of Language in the model. The model output in **Table 8** shows no significant effect of Language.

5.3 Interim discussion: Lag time

Despite the difference in number of stop phonemes in the two languages, the amount of group-level within-speaker lag time variability of voiceless aspirated stops is similar. This is not expected under the most direct implementation of Lindblom (1986) which predicts less variation in languages with more phonemes. Under Contrast-Dependent Variation, similar amounts of lag time variation in Hindi and English are expected, as both employ lag time as a primary cue. In the data here, there is no significant effect of Language on group-level within-category variability. As with any null effect, it could always be the case that the sample was too small to observe any significant effects. However, a significant difference in voicing variation (Section 6) was found, so this model would have detected differences of the same magnitude in lag time variation if they were present.[6]

These results can also be interpreted as providing empirical evidence for the division of lag time and lead time into separate dimensions (as in Mikuteit & Reetz, 2007), as prevoicing and lag time pattern differently. Lag time variation is similar in both languages, but (as shown in Section 6) closure voicing variation differs between Hindi and English. Analyzing prevoicing and lag time as separate phonetic dimensions captures these differences.

6 Closure voicing

In this section, I discuss the analysis of closure voicing, beginning with Section 6.1 detailing the methods of analysis. Section 6.2 compares extent of voicing variation between the two languages. In accordance with the predictions of Contrast-Dependent Variation, there is

---

[6] To determine what effect size would have been detectable given this experimental design, I conducted simulations using simr in R (Green & MacLeod, 2016) with a rage of possible effect sizes. The smallest effect size which would be detectable with 80% power with this sample size is d = 1.5. This would be considered a 'large' effect by most standards (e.g., Gaeta & Brydges, 2020 suggest a 'large' effect for speech research is d ≥ 0.95). However, the effect size observed in Section 6 for voicing variability is large at d = 2.01, which is to be expected given previous literature documenting voicing variation in English. Crucially, these results demonstrate that lag time variability patterns differently from voicing variability in Hindi and English.

more variability in closure voicing in English relative to Hindi, both within- and between-speakers. In Section 6.3, I compare sources of variation between the two languages, including individual differences and vowel context effects, and model these results using regression and model comparison.

6.1 Analysis

Closure duration and closure voicing were hand measured for all stops. Fifty-seven tokens with stop closures longer than 300 ms were excluded. Outlier rejection of tokens with a z-score greater than $|3|$ also excluded 47 tokens. Closure duration was measured from the offset of the preceding vowel until the stop burst. Vowel offset was determined by lack of all formant structure except the lowest formant, following Turk, Nakai, and Sugahara (2006). Closure voicing was measured as the portion of the stop closure which contained periodicity in the waveform, indicating voicing. The percentage of the closure containing voicing was calculated from the measurements of closure duration and closure voicing. Operationalizing voicing with a percentage measurement follows previous work on voicing in English (e.g., Docherty, 1992; Davidson, 2016). These data were also classified according to three categorical bins: no prevoicing (voicing through 0–25% of the stop closure), partial prevoicing (25–90%), and full prevoicing (90–100%). The classification of full prevoicing as voicing through 90% or more of the closure duration follows the categorization in Beckman et al. (2013).

For tokens that are only partially voiced, the percentage measurement does not convey the shape of that voicing, or where in the closure the voicing is present. Davidson (2016) distinguishes multiple shapes of partial voicing for obstruents. 'Bleed' describes voicing that continues from the preceding segment but dissipates some time during closure, before the stop burst. 'Trough' describes voicing that continues from the preceding segment, dissipates, and then reappears before the stop burst. 'Hump' describes cases where voicing does not continue from the preceding segment, then appears in some middle interval of the closure, and then dissipates again before the burst. Lastly, 'Negative VOT' describes voicing that starts in the middle of the stop closure and continues into the burst. In this paper, I only analyze cases of bleed, the pattern displayed in almost all partially prevoiced stops in both languages. Some cases of negative VOT, trough, and hump were observed, but most fell into the group of tokens excluded on the basis of long closure durations (>300 ms). Thirty additional cases of trough were also excluded, all with closures in the 250–300 ms range (just missing the criteria for exclusion on the basis of closure duration). This was done because the percentage measurement does not capture the differences between bleed and trough, and there were not enough trough tokens to analyze shape differences systematically. After all exclusions, 3512 tokens were available for voicing analysis.

Example tokens are shown in **Figures 5–6**. The Hindi tokens in **Figure 5** show voicing before the stop closure which continues through the burst into the vowel. Phonetically voiced and voiceless realizations of the English phonologically voiced stops were observed. The English token in **Figure 6** differs from the other phonologically voiced English example token shown in **Figure 3**. Voicing starts before the stop burst in **Figure 6**, but after the stop burst in **Figure 3**.

6.2 Results: Extent of voicing variation

In this section, I analyze extent of within-category variation of phonologically voiced stops in both languages. Because this analysis is comparative and there are no voiced aspirated stops in English, the Hindi voiced aspirated stops have been excluded from the
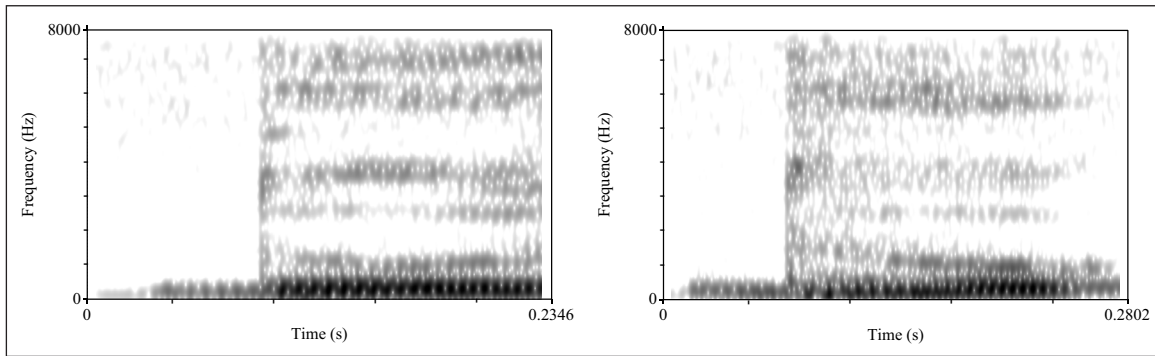
Figure 5: Voiced unaspirated and aspirated stops in Hindi. Left: CV sequence from token of /dut/. Right: CV sequence from token of /dʰup/.
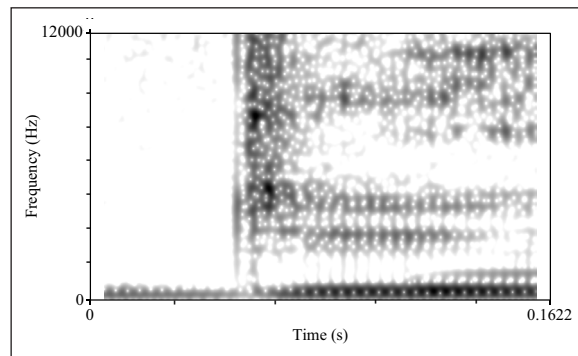

Figure 6: Short lag stop with closure voicing in English.

analysis. The pattern of results does not change (there is still more variation in English relative to Hindi) if the voiced aspirated stops in Hindi are included. **Figure 7** provides a density plot of the closure voicing percentages in both languages, collapsed over speaker and vowel context. In Hindi, the distribution of proportion voiced is skewed as almost all stops are produced with voicing during 100% of the closure. In English, the distribution of voicing is more variable.

In **Figure 8**, I show the same data binned according to voicing category (no prevoicing, partial prevoicing, full prevoicing). Error bars show standard deviation between speakers. In Hindi, almost all voiced stops are produced with full prevoicing (voicing through at least 90% of closure duration). In English, there is more overall variation in degree of prevoicing. Most of the phonologically voiced stops produced in English are partially prevoiced but this varies across speakers.

As in the lag time analysis, I use mixed effects linear regression where within-category variance is the dependent variable. This was calculated by determining the variance in closure voicing within stop category, speaker, and vowel context, which was then used to calculate the coefficient of variation for each condition. This resulted in 126 observations of variance. Language, Place, and Vowel were again included as predictors with random intercepts for speaker. Default dummy coding contrast structure was used with the English coronal _/a/ context as the reference level. The model output is given in **Table 9**. Under Contrast-Dependent Variation, we expect less group-level within-speaker variation in Hindi relative to English. We therefore expect a significant effect of Language in the model, which was observed. The effect size of Language is large (d = 2.01), which is expected given the substantial body of work documenting voicing variation in English.
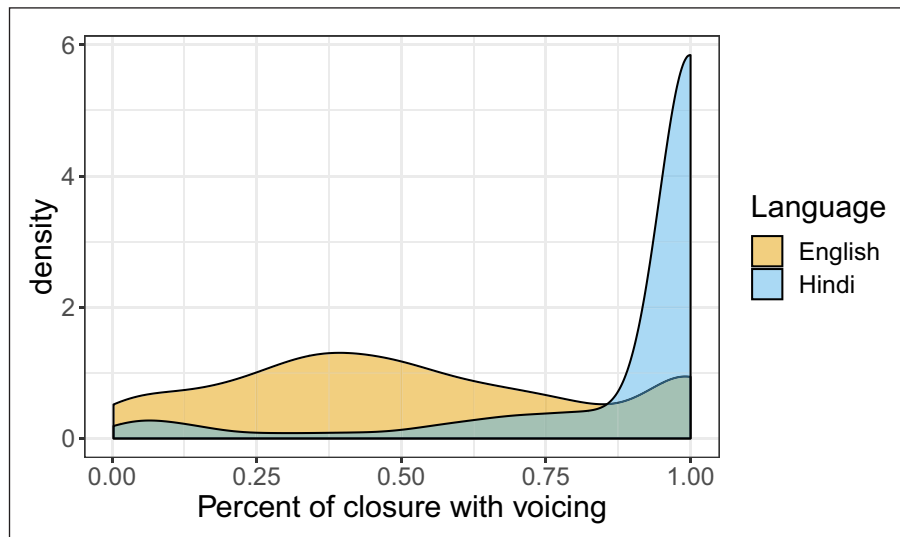
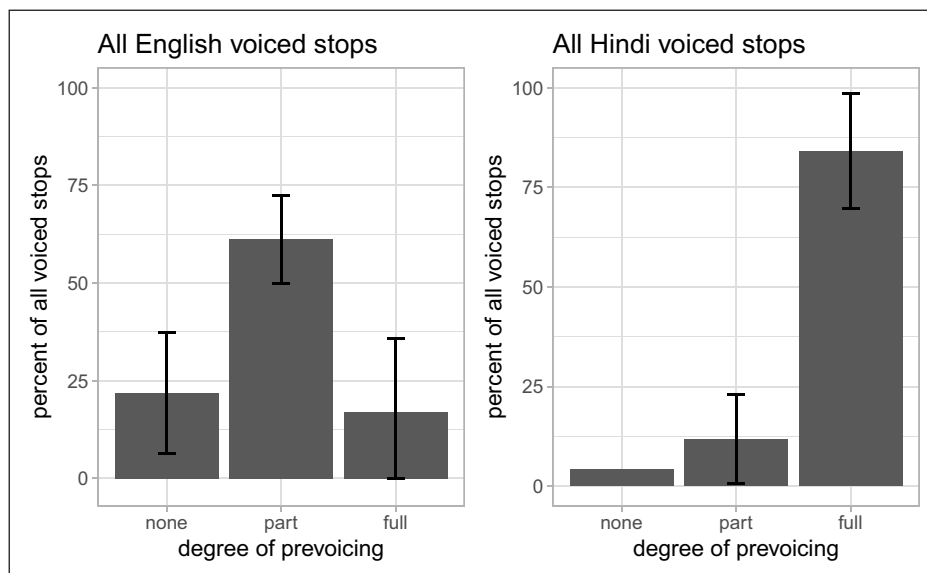**Figure 7:** Density plot of percentage of voicing during stop closures.

Figure 8: Voicing during stop closure in phonologically voiced stops (categorical bins); Error bars show standard deviation between speakers.

Table 9: Fixed effects table for linear mixed effects regression. Dependent variable: within-category within-speaker voicing variation (quantified by coefficient of variation). Predictors: language, place of articulation, V (vowel context), random intercepts for speaker. Model intercept is English coronal _/a/ context.

Fixed effects	Estimate (*SE*)	95% CI	df	t	p
(Intercept)	52.87 (7.50)	38.50 – 67.24	15.05	7.05	<0.001***
Language-Hindi	−37.44 (9.53)	−55.96 – −18.93	9.85	−3.93	0.003**
Place-labial	1.99 (3.97)	−5.72 – 9.69	109.77	0.50	0.618
Place-velar	5.97 (3.97)	−1.73 – 13.67	109.77	1.50	0.135
V-/i/	−6.46 (3.98)	−14.20 – 1.27	109.77	−1.62	0.107
V-/u/	−8.43 (3.98)	−16.17 – −0.70	109.77	−2.12	0.040*

6.3 Results: Structure in voicing variation

6.3.1 Individual differences

There are some individual differences in extent of closure voicing variability in Hindi, but all Hindi speakers consistently fully voice the majority of phonologically voiced stops. **Figure 9** shows distributions for the two Hindi speakers with the most between-speaker difference in amount of voicing. I also provide the data binned in discrete voicing categories for all speakers in **Figure 10**.

   English speakers, however, do not display a consistent pattern in degree of closure voicing. Some English speakers exhibit closure voicing on almost all phonologically voiced stops while others exhibit little closure voicing. **Figure 11** shows the two English speakers with the most between-speaker difference in closure voicing. These two English speakers display near opposite patterns. **Figure 12** provides the same data binned into voicing categories, along with the data from all other English speakers. The English speaker with the most voicing exhibits a pattern which resembles that of the Hindi speakers—the majority of phonologically voiced stops exhibit full prevoicing. The English speaker with the least voicing exhibits the opposite pattern, with about 35% of stops
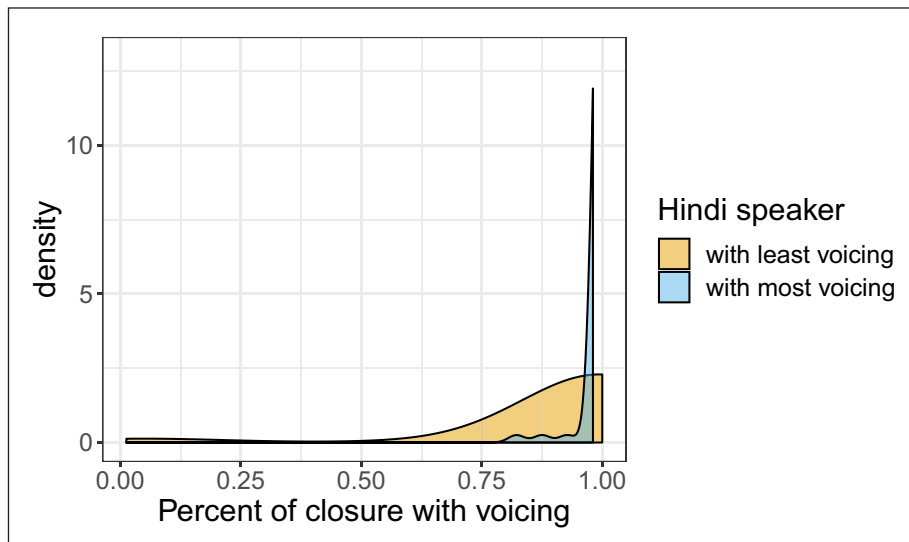
Figure 9: Hindi speakers with greatest difference in voicing.
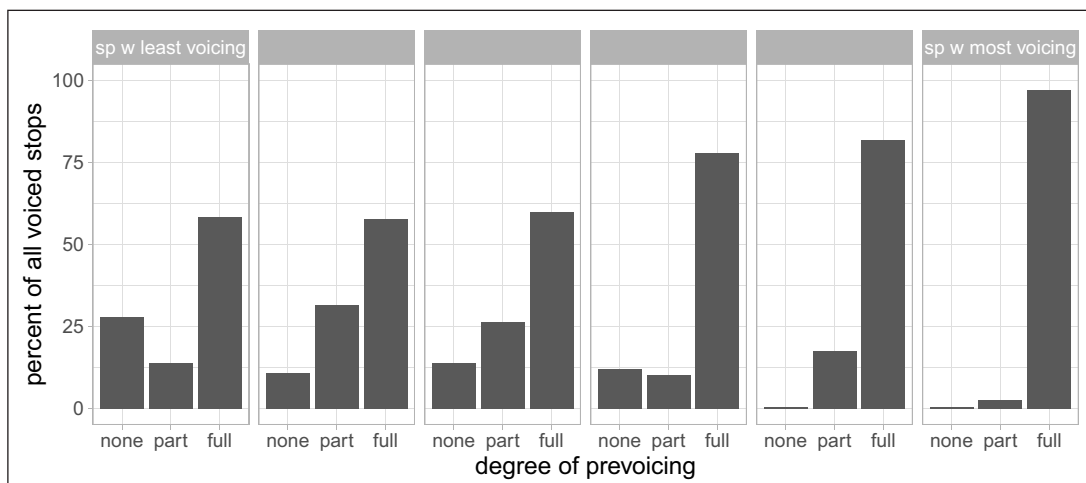
Figure 10: All Hindi speakers in order of amount of voicing.
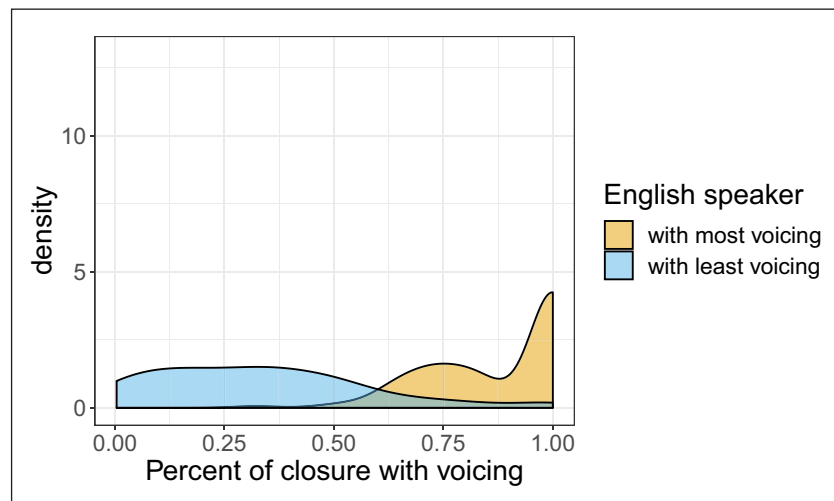
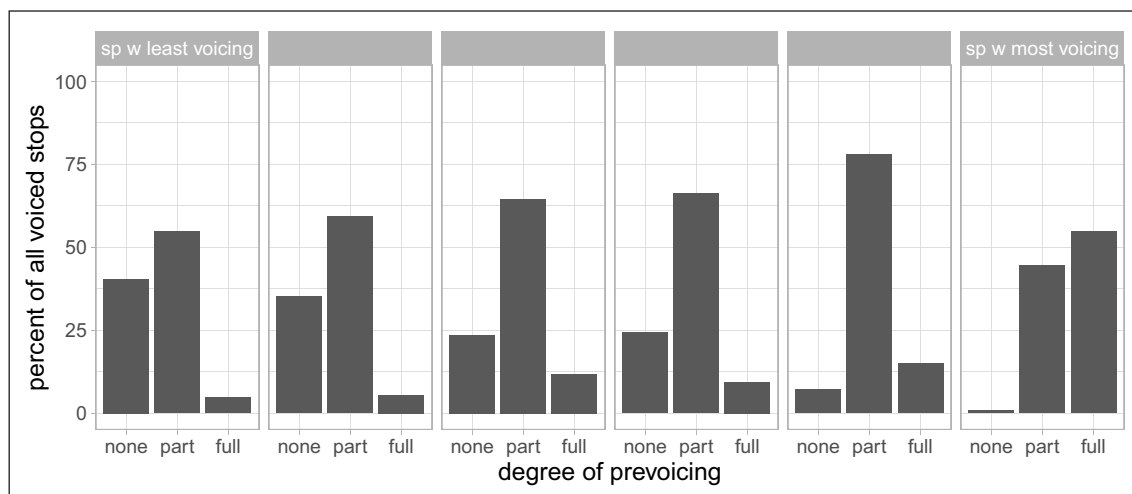**Figure 11:** English speakers with greatest difference in voicing.

Figure 12: All English speakers in order of amount of voicing.

showing no closure voicing and more than half of stops showing partial voicing. These graphs also demonstrate the fact that the 'average' pattern (**Figure 8**) is not particularly representative of the individual English speakers. By contrast, multiple Hindi speakers mirror the 'average' pattern for Hindi.

6.3.2 Variation across vowel contexts

Smith and Westbury (1975) report more prevoicing in English stops before high vowels relative to low vowels. I observe a similar pattern in the English data, but not in Hindi. Just as the pattern of voicing in Hindi is fairly consistent across speakers, the pattern of voicing is also consistent across vowel contexts. The data for both languages are shown in **Figures 13–14**.

6.3.3 Modeling sources of variance

To illustrate the differences in sources and structure of voicing variability, I compare the effects of different factors in accounting for overall voicing variance in both languages. In these models, percent of closure with voicing is the dependent variable, rather than variance in closure voicing as in **Table 9**. Due to the dependent variable being continuous proportion data, I use Beta Regression (Ferrari & Cribari-Neto, 2004), which is intended
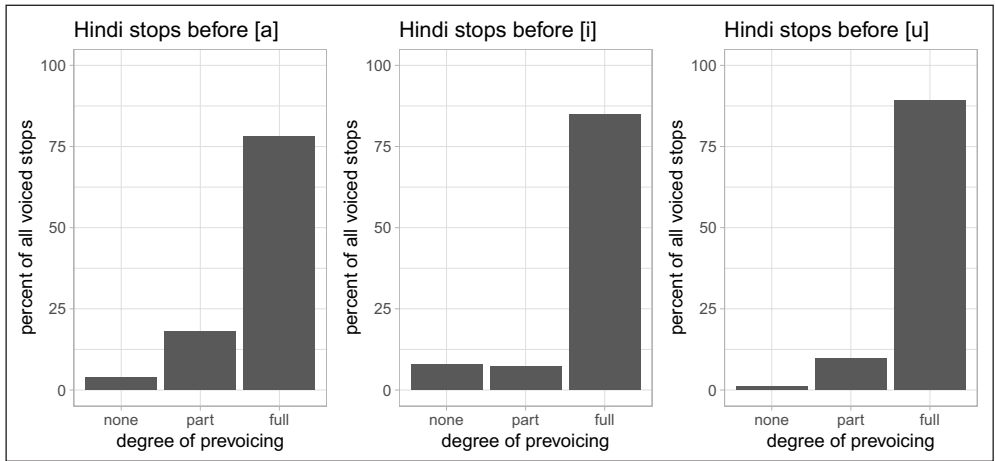
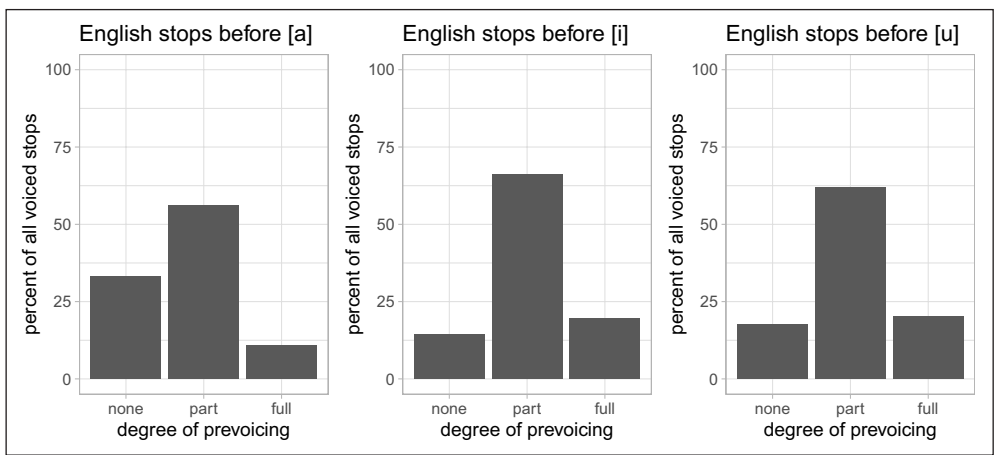**Figure 13:** Prevoicing across vowel contexts in Hindi phonologically voiced stops.

Figure 14: Prevoicing across vowel contexts in English phonologically voiced stops.

for proportion data bounded between (0,1). Unlike a standard linear regression which assumes the data follow Gaussian distributions, the Beta Regression assumes Beta distributions, which tend to be more characteristic of proportion data. As evident from the density plots in the previous section, the proportion data here are not normally distributed and are better approximated with Beta distributions.

Separate regression models were fit for English and Hindi using the following factors as predictors: phonological voicing, place of articulation, speaker, vowel context (V), experimental block, and closure duration, with random intercepts for item. The following interactions were also included in the full models: place × V, place × speaker, and V × speaker. These models included all of the stops elicited, both phonologically voiced and voiceless. Models were fit using the betareg (Cribari-Neto & Zeileis, 2010) and glmmTMB (Brooks et al., 2017) R packages. Best fit models for both languages were determined using variable selection with the Akaike Information Criterion (AIC; Akaike, 1974). Likelihood ratio tests were performed using the lmtest package (Zeileis & Hothorn, 2002) and stepwise selection was performed using the MASS package (Ripley et al., 2013).

While using fixed effects to model factors like vowel context or place of articulation is typical, speaker effects are often modeled using random effects (Allen et al., 2003; Baayen, Davidson, & Bates, 2008). However, the main question for these models is how sources of variance differ in the two languages. This is different from the previous models shown in **Tables 8–9**. The question under investigation there was whether extent of variation differed between languages, for which I did include speaker as a random effect. Including

speaker as a fixed effect allows for quantitative measurement (via the R squared value) of how much variation is accounted for by Speaker relative to the other factors. In the present analysis, significance of main effects is not the focus. Instead, the main question is how much variance is accounted for by each factor and how the best fit models differ between languages. If speaker is included as a random effect, the overall pattern of results about extent of variability remains consistent—there is more voicing variation in English relative to Hindi. The fixed effect analysis allows us to gain more insight into how sources of variation differ between the two languages.

In the full models, there is a significant effect of phonological voicing in both languages, indicating more closure voicing for phonologically voiced stops relative to phonologically voiceless stops. In English, there is a also significant effect of the high vowel /i/ (indicating more voicing relative to /a/), but neither of the vowel effects are significant in Hindi. Many speaker effects are significant in English, while there there are no significant speaker effects in Hindi. In addition, there are several significant interactions between speaker, vowel context, and place of articulation in English, while none of the interactions reach significance in Hindi.

The differences between the full models for the two languages result in different best fit models using the AIC criterion for model selection. The best fit model for the Hindi data (given in **Table 10**) includes only two of the predictors from the full model: phonological voicing and closure duration. With these two factors, this model accounts for 78% of the overall voicing variation in the data. Vowel context, speaker, block, their interactions, or the random effect of item do not significantly improve the model fit, which indicates that they are not significant sources of voicing variation in Hindi. A likelihood ratio test comparing the full model to the best fit model verifies this (given in **Table 11**). The non-significant Chi Square value indicates that there is no significant change in log likelihood when the full model is reduced to the best fit model.[7]

The best fit model for the English data in **Table 12** includes the same predictors as the best fit model in Hindi (voicing and closure duration) as well as vowel context, speaker, the

Table 10: Effect table for best fit model in Hindi. Beta regression with logit link. Dependent variable: closure voicing. Call: voicing percent ~ phonological voicing + closure duration.

Effects	Estimate (*SE*)	z	p
(Intercept)	3.69 (0.09)	43.23	<0.001***
Phonological voicing	−4.92 (0.09)	−55.58	<0.001***
Closure duration	−2.26 (0.45)	−5.05	<0.001***
Pseudo R^2: 0.78			

Table 11: Model comparison: Likelihood ratio test of Hindi restricted model versus full model.

Model	#Df	Log Likelihood	Change in Df	ChiSq	p
1	4	6221.10	−37	36.82	0.48
2	41	6239.60			

Model 1: best fit model (voicing percent ~ phonological voicing + closure duration).

Model 2: full model (voicing percent ~ voicing + V × speaker + place × V + place:speaker + closure duration + block + (1 | item)).

---

[7] This does not fail to reach significance simply because of the large change in degrees of freedom. A comparison of the two models using the English data instead of the Hindi data results in a Chi Square value of 721.3*** with the same change in degrees of freedom.

Table 12: Main effect table for best fit model in English. Beta regression with logit link. Dependent variable: closure voicing. Call: voicing + V × speaker + place × V + place:speaker + closure duration + block. Intercept is speaker e02 voiced coronal /a/ context block 1.

Effects	Estimate (*SE*)	*z*	*p*
(Intercept)	2.41(0.23)	10.34	<0.001***
voicing-voiceless	−1.54 (0.07)	−23.28	<0.001***
V-/i/	0.66 (0.20)	3.24	0.001**
V-/u/	−0.40 (0.21)	−1.86	0.063
speaker-e03	0.55 (0.23)	2.40	0.016*
speaker-e04	−0.66 (0.23)	−2.86	0.004**
speaker-e06	−1.04 (0.23)	−4.56	<0.001***
speaker-e07	−1.94 (0.23)	−8.37	<0.001***
speaker-e09	−1.90 (0.23)	−8.30	<0.001***
place-labial	−0.48 (0.20)	−2.35	0.019*
place-velar	−0.21 (0.20)	−1.04	0.299
closure duration	−9.99 (1.02)	−9.77	<0.001***
block 2	0.05 (0.08)	0.62	0.535
block 3	−0.05 (0.08)	−0.64	0.525
block 4	0.05 (0.08)	0.61	0.541
V-/i/:speaker-e03	−0.44(0.25)	−1.79	0.073
V-/u/:speaker-e03	0.83(0.25)	3.31	<0.001***
V-/i/:speaker-e04	−0.07(0.25)	−0.27	0.788
V-/u/:speaker-e04	1.37(0.26)	5.30	<0.001***
V-/i/:speaker-e06	−0.55(0.25)	−2.20	0.028*
V-/u/:speaker-e06	0.50(0.26)	1.94	0.053
V-/i/:speaker-e07	0.07(0.25)	0.29	0.774
V-/u/:speaker-e07	1.00(0.26)	3.89	<0.001***
V-/i/:speaker-e09	−0.11(0.25)	−0.45	0.656
V-/u/:speaker-e09	0.84(0.26)	3.28	0.001**
V-/i/:place-labial	−0.08(0.17)	−0.49	0.622
V-/u/:place-labial	0.08(0.19)	0.41	0.682
V-/i/:place-velar	0.24(0.18)	1.35	0.177
V-/u/:place-velar	0.24(0.18)	1.34	0.182
speaker-e03:place-labial	0.50(0.25)	2.01	0.045*
speaker-e04:place-labial	0.11(0.26)	0.44	0.660
speaker-e06:place-labial	0.37(0.25)	1.45	0.146
speaker-e07:place-labial	0.70(0.25)	2.77	0.006**
speaker-e09:place-labial	0.51(0.25)	2.01	0.044
speaker-e03:place-velar	−0.24(0.25)	−0.95	0.342
speaker-e04:place-velar	−0.68(0.26)	−2.64	0.008**
speaker-e06:place-velar	−0.50(0.26)	−1.94	0.053*
speaker-e07:place-velar	0.02(0.25)	0.09	0.931
speaker-e09:place-velar	−0.11 (0.25)	−0.42	0.674

Pseudo R²: 0.40

V × speaker interaction, the place × V interaction, the place × speaker interaction, and experimental block. This model accounts for 40% of the overall closure voicing variation in the English data. The only factor from the full model which is not included in the best fit model is the random effect of item. However, including that effect does significantly improve model fit, as is seen in the significant Chi Square value in a likelihood ratio test comparing the English best fit model to the English full model (**Table 13**).

The differences in the best fit models and likelihood ratio tests between the two languages show how the sources of stop voicing variation differ between Hindi and English. The best fit model in Hindi (with only voicing and closure duration as predictors) accounts for 78% of the overall voicing variation. However, the best fit model in English (with almost all predictors included) only accounts for 40% of the overall voicing variation. The variance accounted for by individual predictors also differs between the two languages. In the graphs in **Figure 15**, I show the proportion of total variance accounted for by each individual factor in the full models for both languages. In the Hindi model, almost 78% of the overall voicing variation is accounted for by phonological voicing. All other factors account for less than 1% of overall voicing variation in Hindi. In the English model, only 14% of the overall voicing variation is accounted for by phonological voicing while around 22% of the overall voicing variation is accounted for by speaker.

Although the full model for English still only accounts for 40% of the overall variance, this does not necessarily indicate that the remaining 60% percent of the variance is due to random variation. It could be the case that this variation is also structured by additional factors which are not analyzed in these models. What can be concluded from these models is that (1) the factors analyzed here account for less of the overall variance in the English data relative to the Hindi data, and (2) the strongest predictor of amount of closure voicing in Hindi is phonological voicing while the strongest predictor of voicing in English is individual speaker.

6.4 Interim discussion: Voicing

This section has examined extent and structure of closure voicing variation in Hindi and English stops. The pattern of voicing in Hindi is largely consistent across speakers and vowel contexts. The English speakers vary more in closure voicing both within- and between-speakers. Sources and structure of voicing variation also differ between the two languages. While phonological voicing accounts for ~78% of the total voicing variation in Hindi, it accounts for only 14% of the total voicing variation in English. Speaker, vowel context, and their interactions significantly contribute to the English model, showing that the additional variation in English is structured according to these non-contrastive factors. However, these factors together still account for only 40% of the overall voicing variation in the English data. This suggests that there is either more random variation in

Table 13: Model comparison: Likelihood ratio test of English restricted model versus full model.

Model	#Df	Log Likelihood	Change in Df	ChiSq	p
1	40	1954.0	−1	8.87	0.003[**]
2	41	1958.40			

Model 1: best fit model (voicing percent ~ voicing + V × speaker + place × V + place:speaker + closure duration + block).

Model 2: full model (voicing percent ~ voicing + V × speaker + place × V + place:speaker + closure duration + block + (1 | item)).
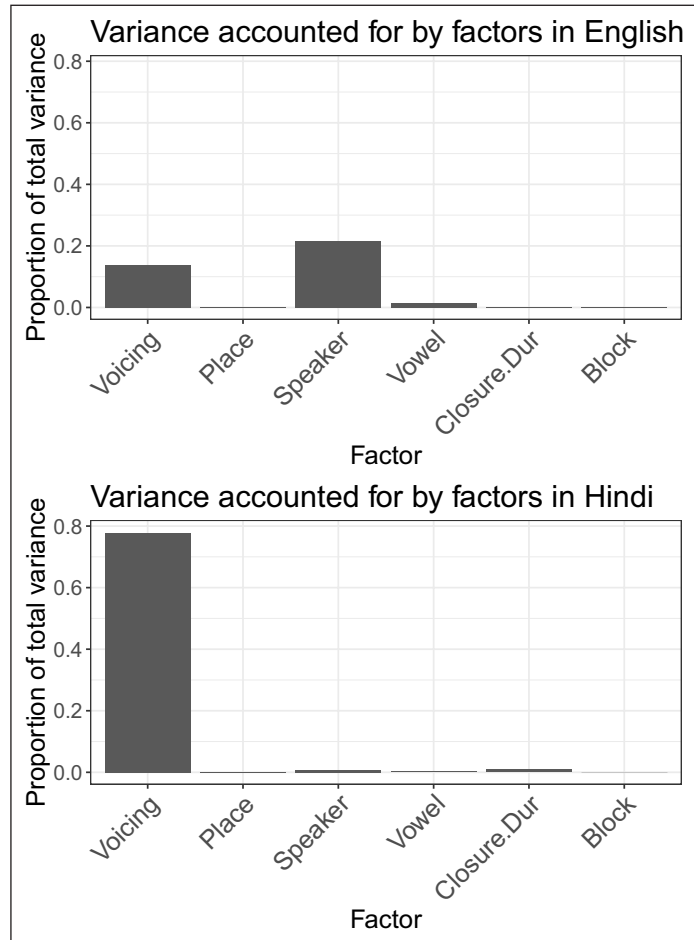
Figure 15: Proportion of total variance accounted for in regression models.

English voicing relative to Hindi, or there are additional factors that structure the English variation which are not considered here. The following subsections discuss these voicing results in light of the literature on prevoicing in English stops (Section 6.4.1), review implications for laryngeal realism and English featural analyses (Section 6.4.2), and review a potential articulatory explanation for voicing differences across vowel contexts (Section 6.4.3).

6.4.1 Prevoicing in English stops

The prevoicing variation in English observed here is in line with recent work on American English documenting prevoicing. Many studies of prevoicing have concentrated on Southern varieties, sometimes reporting prevoicing with higher incidence among male and African-American speakers (Jacewicz et al., 2009; Elston et al., 2016; Herd et al., 2016; Hunnicutt & Morris, 2016). I observed high incidence of prevoicing for some speakers, yet none of the speakers in this study were speakers of a Southern variety.[8] All speakers were female so gender effects could not be tested, but there were female speakers with substantial closure voicing despite previously documented higher incidence among males. These results suggest that prevoicing in American English may be more widespread than previously documented. The results here differ from some of the previous studies on English prevoicing in that the stops were elicited word-initially and utterance-medially,

---

[8] We assume this is the case based on participant answers to a demographic questionnaire. None listed any Southern states as places where they or their parents learned to speak English.

rather than utterance initially. Further work will need to be done with non-Southern populations to determine whether similar patterns are present on utterance-initial stops.

It might seem that the degree of closure voicing observed here could be the result of hyperarticulation in a lab setting. Under this explanation, the between-speaker variation could be due to speaker-specific preference for different hyperarticulation strategies. Speakers that produced mostly prevoiced stops could be using prevoicing as a hyperarticulation strategy, and speakers who produced little voicing could be using other strategies (more salient release bursts, increase in lag time difference, etc.). However, there are multiple reasons to not solely attribute these findings to hyperarticulation effects. First, multiple studies of clear/careful speech have shown that English speakers do not generally use prevoicing as a hyperarticulation strategy, but instead produce more salient release bursts (Keating, 1984; Picheny et al., 1986; J. J. Ohala, 1994b; Hazan & Simpson, 2000). In addition, existing literature documenting prevoicing (summarized above) suggests these findings are typical for American English speakers.

More importantly, if more closure voicing were an indication of hyperarticulation, we would also expect to see other evidence of hyperarticulation such as extended lag time on voiceless stops. However, the speakers who produced the most prevoicing did not also produce the longest lag times on phonologically voiceless stops. In fact, these speakers showed a general preference for more closure voicing across all stops, even phonologically voiceless stops. If the speakers who typically exhibit closure voicing during phonologically voiced stop closures were doing so to hyperarticulate those stops, we would not expect the same speakers to produce voicing during the closures of phonologically voiceless stops. This seems to indicate that these speakers have a more general preference for voicing which cannot be solely attributed to hyperarticulation or clear speech effects. Lastly, if closure voicing were the result of hyperarticulation we may also expect block effects, with hyperarticulation decreasing throughout the experiment, but this was not observed.

6.4.2 Laryngeal realism and English featural analyses

The fact that English speakers use prevoicing on stops (at least sometimes) is potentially compatible with either a laryngeal realist or relativist analysis and these data could be interpreted within either framework. Under a relativist hypothesis, phonetic implementation of [±voice] contrasts can be language-specific, so variation in English prevoicing is not problematic.

Under a realist hypothesis (e.g., Honeybone, 2005; Beckman et al., 2013), the feature system should reflect the phonetic reality of production. English is frequently analyzed by laryngeal realists as a language which does not use the feature [voice], but instead uses [spread glottis], because voiceless lag time is the primary cue. English prevoicing variation has been analyzed with the realist framework; Hunnicutt and Morris (2016) offer a realist analysis of prevoicing in Southern American English. However, the current results showing lack of major structure in voicing variability (aside from individual differences, which together with other factors examined only account for 40% of voicing variation) might be interpreted to indicate that [voice] is not actively controlled by speakers, in keeping with traditional realist analyses of English.

The individual differences observed in English prevoicing patterns might also suggest the need for different feature specifications on the individual level under a realist analysis. For example, the English speaker who consistently produced voicing throughout the closure duration could be described as utilizing both [voice] and [spread glottis] (as in Hunnicutt & Morris, 2016), while the speaker who produced very little closure voicing could be described as utilizing only the [spread glottis] feature. Ultimately, the data could

be potentially interpreted in both realist and relativist frameworks, and do not provide direct support for either approach.

6.4.3 Variation across vowel contexts

As in Smith and Westbury (1975), I observed more prevoicing before high vowels in English. Smith and Westbury propose a possible articulatory explanation for this: Moving the tongue root to produce a high vowel puts additional tension on the vocal folds, making it easier to sustain voicing through the closure. However, the Hindi speakers are consistent in voicing across vowel contexts and do not prevoice less before low vowels relative to high vowels. The lack of even a small effect of this kind in Hindi suggests two explanations. (1) It could be that the pattern observed in English does not actually have a physiological basis and is a learned non-contrastive pattern or (2) the Hindi speakers are able to overcome the physiological challenges to maintain the contrasts of their language.

7 Discussion

7.1 Inventory size and phonetic variability

Lindblom's (1986) hypothesis "that phonetic values of vowel phonemes should exhibit less variation in small systems than in large systems" is intuitive, but has scant and conflicting evidence in the literature. Previous work has shown it to be inadequate for predicting patterns of variability in vowel inventories, demonstrating the need for a more nuanced approach (see Section 2.3.2 for a review). Similarly, the results here show that it is not the case that phonetic values in larger 'systems' are always less variable—Hindi speakers showed just as much variation as English speakers in voiceless lag time, despite having twice the number of stop phonemes.

In this paper, I have proposed an alternative hypothesis which makes predictions about phonetic variability according to cue weight of individual phonetic dimensions rather than size of phonemic inventories. This accounts for the differential behavior of individual cues with respect to variability patterns (as demonstrated here and in e.g., Recasens & Espinosa, 2006 for vowels) as well as importance of phonological context, as demonstrated by Renwick (2012). Contrast-Dependent Variation incorporates these notions by making separate predictions about individual phonetic dimensions through the use of cue weight, which differs across phonological contexts. For example, the predictions of Contrast-Dependent Variation would change for Hindi versus English stops in syllable-final position (as opposed to utterance-medial word-initial position, examined here), as cue status and relative weights differ across positions. As the predictions are tied to cue weights in particular contexts, this hypothesis only makes predictions about token-by-token variability and is not intended to account for variation across phonological contexts or contextual allophonic variation. Testing the generalizability of Contrast-Dependent Variation by directly examining extent of variation across contexts is an area for future work.

While Contrast-Dependent Variation offers an alternative to DT predictions, it is in some ways consistent with the original intuition behind Lindblom (1986). Lindblom's hypothesis (and work in DT more generally) assumes that production is optimized for ease of perception through sufficient dispersion of phonological categories in acoustic space. Contrast-Dependent Variation still carries an implicit assumption about the relevance of perceptual distinction by incorporating cue weight and assuming that speech sound contrasts must somehow be sufficiently distinct. However, by comparing cue weights of individual dimensions rather than inventory size, this approach acknowledges the multidimensional context-sensitive nature of phonological contrast. Therefore, no

differences in patterns of variability are predicted based solely on phoneme inventory size. Rather, for a single phonetic dimension, we expect less within-category variability in languages for which that dimension is employed as a primary cue relative to languages in which that dimension is employed as a secondary cue.

7.2 Perception and cue weighting

Modeling work on cue weighting has shown that algorithmically weighting cues based on how reliably they distinguish phonological contrasts mirrors the cue weighting patterns observed in perceptual data (Toscano & McMurray, 2010). The model employed by Toscano and McMurray (2010: 438) estimates the reliability of a phonetic dimension with a ratio of mean values to within-category variances, mirroring the Dispersion Theoretic assumption that distributions must be tightened in a crowded space to avoid overlap and preserve perceptual distinction. This type of model is supported by empirical work on the relationship between within-category variability and cue weighting in perception. For example, Clayards et al. (2008) show that perceptual uncertainty increases with within-category phonetic variability and Holt and Lotto (2006) show that cue weighting strategies are affected by changes in input variability.

The present results provide empirical support from production for the inclusion of within-category variance in cue-weighting models. A prediction that arises from the reliability definition in Toscano and McMurray (2010) is that strength of cue and relative amount of within-category variation should be inversely correlated, which is supported here through between-language and within-language comparison. English speakers exhibited more variation in extent of closure voicing (a secondary perceptual cue) relative to Hindi speakers, for whom voicing provides a primary perceptual cue. Within English, speakers also display less variation in the primary cue of lag time relative to the secondary cue of voicing.

8. Conclusion

In this paper, I have compared within-category acoustic-phonetic variation of stops in Hindi and English. Hindi and English speakers produced similar amounts of group-level within-speaker within-category variation in voiceless lag time of phonologically voiceless stops, but English speakers produced significantly more within- and between-speaker variation in closure voicing. This is consistent with Contrast-Dependent Variation (Hauser, 2019), the proposed revision of Lindblom's (1986) hypothesis: There should be less variation along a phonetic dimension in languages that use that dimension as a primary cue relative to languages that use the same dimension as a secondary cue. While it is well-established that production is variable in every language, these results show that extent and sources of variation, including individual differences, are language-specific and sensitive to differences in phonological contrast implementation.

Data Accessibility Statement

Data, code, and other materials used in this project have been made openly available through the Open Science Foundation. They can be found here: https://osf.io/7cxhk/.

Additional File

The additional file for this article can be found as follows:

- Appendix. Null lexical effects on lag time (Table 1) and closure voicing (Table 2). DOI: https://doi.org/10.16995/labphon.6465.s1

Ethics and Consent

Acknowledgements

Funding Information

Competing Interests

The author has no competing interests to declare.

References

Abramson, A. S., & Lisker, L. (1967). Laryngeal behavior, the speech signal and phonological simplicity. In A. Graur (Ed.), *Proceedings of the Tenth International Congress of Linguistics, Bucharest* (Vol. 4, p. 123–129).

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716–723. DOI: https://doi.org/10.1007/978-1-4612-1694-0_16

Alexander, J. A. (2010). *The theory of adaptive dispersion and acoustic-phonetic properties of cross-language lexical-tone systems* (Unpublished doctoral dissertation). Northwestern University, Evanston, IL.

Allen, J. S., Miller, J. L., & DeSteno, D. (2003). Individual talker differences in voice-onset-time. *The Journal of the Acoustical Society of America, 113*(1), 544–552. DOI: https://doi.org/10.1121/1.1528172

Audacity Team. (1999–2021). Audacity(r): Free Audio Editor and Recorder [Computer software manual]. http://audacity.sourceforge.net/.

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

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics, 40*(1), 177–189. DOI: https://doi.org/10.1016/j.wocn.2011.09.001

Baese-Berk, M. M., & Morrill, T. H. (2015). Speaking rate consistency in native and nonnative speakers of English. *The Journal of the Acoustical Society of America, 138*(3), EL223–EL228. DOI: https://doi.org/10.21437/SpeechProsody.2016-230

Baker, W., & Trofimovich, P. (2006). Perceptual paths to accurate production of L2 vowels: The role of individual differences. *IRAL–International Review of Applied*

Linguistics in Language Teaching, 44(3), 231–250. DOI: https://doi.org/10.1515/IRAL.2006.010

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., … Treiman, R. (2007). The English lexicon project. *Behavior research methods, 39*(3), 445–459. DOI: https://doi.org/10.3758/BF03193014

Bang, H.-Y., & Clayards, M. (2016). Structured variation across sound contrasts, talkers, and speech styles. *Poster presented at LabPhon15: Speech Dynamics and Phonological Representation. Ithaca, NY*.

Bang, H.-Y., Sonderegger, M., Kang, Y., Clayards, M., & Yoon, T.-J. (2018). The emergence, progress, and impact of sound change in progress in Seoul Korean: Implications for mechanisms of tonogenesis. *Journal of Phonetics, 66*, 120–144. DOI: https://doi.org/10.1016/j.wocn.2017.09.005

Bates, D., Sarkar, D., Bates, M. D., & Matrix, L. (2007). The lme4 package. *R package version, 2*(1), 74.

Becker-Kristal, R. (2010). *Acoustic typology of vowel inventories and dispersion theory: Insights from a large cross-linguistic corpus* (Unpublished doctoral dissertation). University of California, Los Angeles, Los Angeles, CA.

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

Benguerel, A.-P., & Bhatia, T. K. (1980). Hindi stop consonants: an acoustic and fiberscopic study. *Phonetica, 37*(3), 134–148. DOI: https://doi.org/10.1159/000259987

Blake, K. (2019). *Effects of contrast and articulatory precision on the realization of sibilants.* Ithaca, NY. (Manuscript, Cornell University)

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5:9/10*, 341–345.

Boersma, P., & Hamann, S. (2008). The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology, 25*(02), 217–270. DOI: https://doi.org/10.1017/S0952675708001474

Bradlow, A. R. (1995). A comparative acoustic study of English and Spanish vowels. *The Journal of the Acoustical Society of America, 97*(3), 1916–1924. DOI: https://doi.org/10.1121/1.412064

Bradlow, A. R., & Bent, T. (2002). The clear speech effect for non-native listeners. *The Journal of the Acoustical Society of America, 112*(1), 272–284. DOI: https://doi.org/10.1121/1.1487837

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., … Bolker, B. M. (2017). Modeling zero-inflated count data with glmmtmb. *bioRxiv*. DOI: https://doi.org/10.1101/132753

Chandrasekaran, B., Sampath, P. D., & Wong, P. C. (2010). Individual variability in cueweighting and lexical tone learning. *The Journal of the Acoustical Society of America, 128*(1), 456–465. DOI: https://doi.org/10.1121/1.3445785

Chen, F. R. (1980). *Acoustic characteristics and intelligibility of clear and conversational speech at the segmental level* (Unpublished doctoral dissertation). Massachusetts Institute of Technology, Cambridge, MA.

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

Chodroff, E. (2018). *Corpus phonetics tutorial.* Retrieved from https://arxiv.org/abs/1811.05553

Chodroff, E., Godfrey, J., Khudanpur, S., & Wilson, C. (2015). Structured variability in acoustic realization: A corpus study of voice onset time in American English stops. In T. S. C. for ICPhS 2015 (Ed.), *Proceedings of the 18th international congress of phonetic sciences.* Glasgow, UK: The University of Glasgow.

Chodroff, E., & Wilson, C. (2017). Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *Journal of Phonetics*, *61*, 30–47. DOI: https://doi.org/10.1016/j.wocn.2017.01.001

Clayards, M. (2018). Differences in cue weights for speech perception are correlated for individuals within and across contrasts. *The Journal of the Acoustical Society of America*, *144*(3), EL172–EL177. DOI: https://doi.org/10.1121/1.5052025

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition, 108*(3), 804–809. DOI: https://doi.org/10.1016/j.cognition.2008.04.004

Clements, G. N. (2003). Feature economy in sound systems. *Phonology*, *20*(03), 287–333. DOI: https://doi.org/10.1017/S095267570400003X

Clopper, C. G., Pisoni, D. B., & De Jong, K. (2005). Acoustic characteristics of the vowel systems of six regional varieties of American English. *The Journal of the Acoustical Society of America, 118*(3), 1661–1676. DOI: https://doi.org/10.1121/1.2000774

Coetzee, A. W., Beddor, P. S., Shedden, K., Styler, W., & Wissing, D. (2018). Plosive voicing in Afrikaans: Differential cue weighting and tonogenesis. *Journal of Phonetics, 66*, 185–216. DOI: https://doi.org/10.1016/j.wocn.2017.09.009

Cribari-Neto, F., & Zeileis, A. (2010). Beta regression in R. *Journal of Statistical Software, 34*(2), 1–24. Retrieved from http://www.jstatsoft.org/v34/i02/. DOI: https://doi.org/10.18637/jss.v034.i02

Cyran, E. (2011). Laryngeal realism and laryngeal relativism: Two voicing systems in Polish? *Studies in Polish Linguistics, 6*(1), 45–80.

Davidson, L. (2016). Variability in the implementation of voicing in American English obstruents. *Journal of Phonetics, 54*, 35–50. DOI: https://doi.org/10.1016/j.wocn.2015.09.003

Deterding, D., & Nolan, F. (2007). Aspiration and voicing of Chinese and English plosives. In *Proceedings of the 16th international congress of phonetic sciences* (pp. 385–388). Saarbrücken, Germany.

DiCanio, C., Nam, H., Amith, J. D., García, R. C., & Whalen, D. H. (2015). Vowel variability in elicited versus spontaneous speech: Evidence from Mixtec. *Journal of Phonetics, 48*, 45–59. DOI: https://doi.org/10.1016/j.wocn.2014.10.003

Dixit, R. P. (1989). Glottal gestures in Hindi plosives. *Journal of Phonetics, 17*, 213–237. DOI: https://doi.org/10.1016/S0095-4470(19)30431-0

Docherty, G. J. (1992). *The timing of voicing in British English obstruents*. Berlin, Germany: Foris. DOI: https://doi.org/10.1515/9783110872637.1

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

Eckert, P. (2000). *Linguistic variation as social practice*. Malden, MA: Blackwell.

Elston, A. H., Blake, K., Berkson, K., Herd, W., Cariño, J., Nelson, M., … Torrence, D. (2016). Region, gender, and within-category variation in American English voiced stops. *The Journal of the Acoustical Society of America, 139*(4), 2123–2123. DOI: https://doi.org/10.1121/1.4950320

Engstrand, O., & Krull, D. (1994). Durational correlates of quantity in Swedish, Finnish and Estonian: Cross-language evidence for a theory of adaptive dispersion. *Phonetica, 51*(1–3), 80–91. DOI: https://doi.org/10.1159/000261960

Ferguson, S. H., & Kewley-Port, D. (2007). Talker differences in clear and conversational speech: Acoustic characteristics of vowels. *Journal of Speech, Language, and Hearing Research, 50*(5), 1241–1255. DOI: https://doi.org/10.1044/1092-4388(2007/087)

Ferrari, S., & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics, 31*(7), 799–815. DOI: https://doi.org/10.1080/0266476042000214501

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

Flemming, E. (1996). Evidence for constraints on contrast: The dispersion theory of contrast. *UCLA Working Papers in Phonology, 1*, 86–106.

Flemming, E. (2004). Contrast and perceptual distinctiveness. *Phonetically based phonology*, 232–276. DOI: https://doi.org/10.1017/CBO9780511486401.008

Gaeta, L., & Brydges, C. R. (2020). An examination of effect sizes and statistical power in speech, language, and hearing research. *Journal of Speech, Language, and Hearing Research, 63*(5), 1572–1580. DOI: https://doi.org/10.1044/2020_JSLHR-19-00299

Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language, 84*(3), 474–496. DOI: https://doi.org/10.1353/lan.0.0035

Garellek, M., & White, J. (2015). Phonetics of Tongan stress. *Journal of the International Phonetic Association, 45*(1), 13–34. DOI: https://doi.org/10.1017/S0025100314000206

Gendrot, C., & Adda-Decker, M. (2007). Impact of duration and vowel inventory size on formant values of oral vowels: An automated formant analysis from eight languages. In *Proceedings of the 16th international congress of phonetic sciences* (pp. 1417–1420). Saarbrücken, Germany.

Gimenes, M., & New, B. (2016). Worldlex: Twitter and blog word frequencies for 66 languages. *Behavior research methods, 48*(3), 963–972. DOI: https://doi.org/10.3758/s13428-015-0621-0

Green, P., & MacLeod, C. J. (2016). Simr: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution, 7*(4), 493–498. DOI: https://doi.org/10.1111/2041-210X.12504

Hall, D. C. (2011). Phonological contrast and its phonetic enhancement: Dispersedness without dispersion. *Phonology, 28*(1), 1–54. DOI: https://doi.org/10.1017/S0952675711000029

Harnsberger, J. D., Wright, R., & Pisoni, D. B. (2008). A new method for eliciting three speaking styles in the laboratory. *Speech Communication, 50*(4), 323–336. DOI: https://doi.org/10.1016/j.specom.2007.11.001

Hauser, I. (2019). *Effects of phonological contrast on within-category phonetic variation* (Unpublished doctoral dissertation). University of Massachusetts Amherst, Amherst, MA.

Hazan, V., & Simpson, A. (2000). The effect of cue-enhancement on consonant intelligibility in noise: Speaker and listener effects. *Language and Speech, 43*(3), 273–294. DOI: https://doi.org/10.1177/00238309000430030301

Herd, W., Torrence, D., & Carino, J. (2016). Prevoicing differences in Southern English: Gender and ethnicity effects. *The Journal of the Acoustical Society of America, 139*(4), 2217–2217. DOI: https://doi.org/10.1121/1.4950635

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America, 97*(5), 3099–3111. DOI: https://doi.org/10.1121/1.409456

Hillenbrand, J., Ingrisano, D. R., Smith, B. L., & Flege, J. E. (1984). Perception of the voiced–voiceless contrast in syllable-final stops. *The Journal of the Acoustical Society of America, 76*(1), 18–26. DOI: https://doi.org/10.1121/1.391094

Hogan, J. T., & Rozsypal, A. J. (1980). Evaluation of vowel duration as a cue for the voicing distinction in the following word-final consonant. *The Journal of the Acoustical Society of America, 67*(5), 1764–1771. DOI: https://doi.org/10.1121/1.384304

Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *The Journal of the Acoustical Society of America, 119*(5), 3059–3071. DOI: https://doi.org/10.1121/1.2188377

Honeybone, P. (2005). Diachronic evidence in segmental phonology: The case of obstruent laryngeal specifications. In J. v. d. W. M. van Oostendorp (Ed.), *The internal organization of phonological segments* (p. 319–354). Berlin, Germany: Mouton de Gruyter. DOI: https://doi.org/10.1515/9783110890402.317

Hunnicutt, L., & Morris, P. A. (2016). Prevoicing and aspiration in Southern American English. *University of Pennsylvania Working Papers in Linguistics, 22*(1), 24.

Idemaru, K., Holt, L. L., & Seltman, H. (2012). Individual differences in cue weights are stable across time: The case of Japanese stop lengths. *The Journal of the Acoustical Society of America, 132*(6), 3950–3964. DOI: https://doi.org/10.1121/1.4765076

Iskarous, K., Shadle, C. H., & Proctor, M. I. (2011). Articulatory–acoustic kinematics: The production of American English /s/. *The Journal of the Acoustical Society of America, 129*, 944–954. DOI: https://doi.org/10.1121/1.3514537

Jacewicz, E., Fox, R. A., & Lyle, S. (2009). Variation in stop consonant voicing in two regional varieties of American English. *Journal of the International Phonetic Association, 39*(3), 313–334. DOI: https://doi.org/10.1121/1.4783053

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

Johnson, D. E. (2007). *Stability and change along a dialect boundary: The low vowels of southeastern New England* (Unpublished doctoral dissertation). University of Pennsylvania, Philadelphia, PA.

Johnson, K., Ladefoged, P., & Lindau, M. (1993). Individual differences in vowel production. *The Journal of the Acoustical Society of America, 94*(2), 701–714. DOI: https://doi.org/10.1121/1.406887

Jongman, A., Fourakis, M., & Sereno, J. A. (1989). The acoustic vowel space of Modern Greek and German. *Language and Speech, 32*(3), 221–248. DOI: https://doi.org/10.1177/002383098903200303

Keating, P. A. (1983). Comments on the jaw and syllable structure. *Journal of Phonetics, 11*(4), 401–406. DOI: https://doi.org/10.1016/S0095-4470(19)30839-3

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language, 60*, 286–319. DOI: https://doi.org/10.2307/413642

Keshet, J., Sonderegger, M., & Knowles, T. (2014). AutoVOT: A tool for automatic measurement of voice onset time using discriminative structured prediction [computer program]. Version 0.91.

Kim, D., & Clayards, M. (2019). Individual differences in the link between perception and production and the mechanisms of phonetic imitation. *Language, Cognition and Neuroscience, 31*, 1–18. DOI: https://doi.org/10.1080/23273798.2019.1582787

Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language, 70*(3), 419–454. DOI: https://doi.org/10.2307/416481

Kong, E. J., & Edwards, J. (2016). Individual differences in categorical perception of speech: Cue weighting and executive function. *Journal of Phonetics, 59*, 40–57. DOI: https://doi.org/10.1016/j.wocn.2016.08.006

Kong, E. J., & Yoon, I. H. (2013). L2 proficiency effect on the acoustic cue-weighting pattern by Korean L2 learners of English: Production and perception of English stops. *Phonetics and Speech Sciences, 5*(4), 81–90. DOI: https://doi.org/10.13064/KSSS.2013.5.4.081

Krause, J. C., & Braida, L. D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *The Journal of the Acoustical Society of America, 115*(1), 362–378. DOI: https://doi.org/10.1121/1.416659

Kuang, J., & Cui, A. (2018). Relative cue weighting in production and perception of an ongoing sound change in Southern Yi. *Journal of Phonetics, 71*, 194–214. DOI: https://doi.org/10.1016/j.wocn.2018.09.002

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmertest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13). DOI: https://doi.org/10.18637/jss.v082.i13

Lee, H., & Jongman, A. (2019). Effects of sound change on the weighting of acoustic cues to the three-way laryngeal stop contrast in Korean: Diachronic and dialectal comparisons. *Language and Speech, 62*(3), 509–530. DOI: https://doi.org/10.1177/0023830918786305

Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language, 48*, 839–862. DOI: https://doi.org/10.2307/411991

Lindblom, B. (1986). Phonetic universals in vowel systems. In J. Ohala & J. Jaeger (Eds.), *Experimental phonology* (pp. 13–44). New York, NY: Academic Press.

Lindblom, B., & Maddieson, I. (1988). Phonetic universals in consonant systems. In L. Hyman & C. Li (Eds.), *Language, speech, and mind: Studies in honor of Victoria Fromkin* (p. 62–80). New York, NY: Routledge.

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

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20*(3), 384–422. DOI: https://doi.org/10.1080/00437956.1964.11659830

Lisker, L., & Abramson, A. S. (1967). Some effects of context on voice onset time in English stops. *Language and Speech, 10*(1), 1–28. DOI: https://doi.org/10.1177/002383096701000101

Livijn, P. (2000). Acoustic distribution of vowels in differently sized inventories–hot spots or adaptive dispersion. In *Proceedings of the XIIIth Swedish Phonetics Conference* (p. 193–96). Skövde, Sweden.

Lombardi, L. (1994). *Laryngeal features and laryngeal neutralization.* New York, NY: Routledge. DOI: https://doi.org/10.4324/9780429454929

Lorge, B. (1967). A study of the relationship between production and perception of initial and intervocalic /t/ and /d/ in individual English speaking adults. *Haskins Labs Status Report Speech Research, SR-9*, 3.1-1.18.

Mackie, S., & Mielke, J. (2011). Feature economy in natural, random, and synthetic inventories. In R. Ridouane & G. N. Clements (Eds.), *Where do phonological features come from? Cognitive, physical, and developmental bases of distinctive speech categories* (pp. 43–63). Amsterdam, Netherlands: John Benjamins. DOI: https://doi.org/10.1075/lfab.6.03mac

MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences, 21*(04), 499–511. DOI: https://doi.org/10.1017/S0140525X98001265

Maddieson, I. (1977). Tone loans: A question concerning tone spacing and a method of answering it. *UCLA Working Papers on Phonetics: Studies on Tone, 36*, 49–83.

Maddieson, I., & Disner, S. F. (1984). *Patterns of sound*. Cambridge, UK: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511753459

Manuel, S. Y. (1990). The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *The Journal of the Acoustical Society of America*, *88*(3), 1286–1298. DOI: https://doi.org/10.1121/1.399705

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal forced aligner [computer program] version 0.9.0, retrieved from http://montrealcorpustools.github.io/montreal-forced-aligner/.

Mikuteit, S., & Reetz, H. (2007). Caught in the ACT: The timing of aspiration and voicing in East Bengali. *Language and Speech*, *50*(2), 247–277. DOI: https://doi.org/10.1177/0023830907050002040101

Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *The Journal of the Acoustical Society of America*, *85*(5), 2088–2113. DOI: https://doi.org/10.1121/1.397861

Newman, R. S., Clouse, S. A., & Burnham, J. L. (2001). The perceptual consequences of within-talker variability in fricative production. *The Journal of the Acoustical Society of America*, *109*(3), 1181–1196. DOI: https://doi.org/10.1121/1.1348009

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics*, *39*(2), 132–142. DOI: https://doi.org/10.1016/j.wocn.2010.12.007

Ohala, J. (1979). Chairman's introduction to symposium on phonetic universals in phonological systems and their explanations. In *Proceedings of the Ninth International Congress of Phonetic Sciences*. University of Copenhagen.

Ohala, J. J. (1994a). Acoustic study of clear speech: A test of the contrastive hypothesis. In *Proceedings of the international symposium on prosody* (pp. 75–89). Yokohama, Japan.

Ohala, J. J. (1994b). Clear speech does not exaggerate phonemic contrast. *The Journal of the Acoustical Society of America*, *96*(5), 3227–3227. DOI: https://doi.org/10.1121/1.411157

Ohala, M. (1983). *Aspects of Hindi phonology*. Delhi, India: Motilal Banarsidass Publishers.

Padgett, J. (2001). Contrast dispersion and Russian palatalization. In E. Hume & K. Johnson (Eds.), *The role of speech perception in phonology* (pp. 187–218). Boston, MA: Brill.

Padgett, J., & Tabain, M. (2005). Adaptive dispersion theory and phonological vowel reduction in Russian. *Phonetica*, *62*(1), 14–54. DOI: https://doi.org/10.1159/000087223

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech, Language, and Hearing Research*, *29*(4), 434–446. DOI: https://doi.org/10.1044/jshr.2904.434

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

Poon, P. G., & Mateer, C. A. (1985). A study of VOT in Nepali stop consonants. *Phonetica*, *42*(1), 39–47. DOI: https://doi.org/10.1159/000261736

Prince, A., & Smolensky, P. (1993/2004). *Optimality theory: Constraint interaction in generative grammar*. Malden, MA: Blackwell. DOI: https://doi.org/10.1002/9780470759400

Quirk, R., Greenbaum, S., Leech, G. N., & Svartvik, J. (1972). *A grammar of contemporary English*. London, UK: Longman.

R Core Team. (2013). R: A Language and Environment for Statistical Computing [Computer software manual]. Vienna, Austria.

Recasens, D., & Espinosa, A. (2006). Dispersion and variability of Catalan vowels. *Speech Communication*, *48*(6), 645–666. DOI: https://doi.org/10.1016/j.specom.2005.09.011

Renwick, M. (2012). *Vowels of Romanian: Historical, phonological and phonetic studies* (Unpublished doctoral dissertation). Cornell University, Ithaca, NY.

Ripley, B., Venables, B., Bates, D. M., Hornik, K., Gebhardt, A., Firth, D., & Ripley, M. B. (2013). Package 'mass'. *CRAN Repository*. Retrieved from https://cran.r-project.org/web/packages/MASS/

Rose, P. (2010). The effect of correlation on strength of evidence estimates in Forensic Voice Comparison: Uni-and multivariate Likelihood Ratio-based discrimination with Australian English vowel acoustics. *International Journal of Biometrics*, *2*(4), 316–329. DOI: https://doi.org/10.1504/IJBM.2010.035447

Schertz, J., Cho, T., Lotto, A., & Warner, N. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics*, *52*, 183–204. DOI: https://doi.org/10.1016/j.wocn.2015.07.003

Schertz, J., & Clare, E. J. (2020). Phonetic cue weighting in perception and production. *Wiley Interdisciplinary Reviews: Cognitive Science*, *11*(2). DOI: https://doi.org/10.1002/wcs.1521

Schiefer, L. (1986). F0 in the production and perception of breathy stops: Evidence from Hindi. *Phonetica*, *43*(1–3), 43–69. DOI: https://doi.org/10.1159/000261760

Schwartz, J.-L., Boë, L.-J., Badin, P., & Sawallis, T. R. (2012). Grounding stop place systems in the perceptuo-motor substance of speech: On the universality of the labial–coronal–velar stop series. *Journal of Phonetics*, *40*(1), 20–36. DOI: https://doi.org/10.1016/j.wocn.2011.10.004

Schwartz, J.-L., Boë, L.-J., Vallée, N., & Abry, C. (1997). The dispersion-focalization theory of vowel systems. *Journal of Phonetics*, *25*(3), 255–286. DOI: https://doi.org/10.1006/jpho.1997.0043

Schwarz, M., Sonderegger, M., & Goad, H. (2019). Realization and representation of Nepali laryngeal contrasts: Voiced aspirates and laryngeal realism. *Journal of Phonetics*, *73*, 113–127. DOI: https://doi.org/10.1016/j.wocn.2018.12.007

Scobbie, J. M. (2006). Flexibility in the face of incompatible English VOT systems. In L. Goldstein, D. H. Whalen, & C. T. Best (Eds.), *Laboratory Phonology 8: Varieties of phonological competence* (pp. 367–392). Boston, MA: Mouton de Gruyter. DOI: https://doi.org/10.1515/9783110197211.2.367

Shultz, A. A., Francis, A. L., & Llanos, F. (2012). Differential cue weighting in perception and production of consonant voicing. *The Journal of the Acoustical Society of America*, *132*(2), EL95–EL101. DOI: https://doi.org/10.1121/1.4736711

Smiljanić, R., & Bradlow, A. R. (2005). Production and perception of clear speech in Croatian and English. *The Journal of the Acoustical Society of America*, *118*(3), 1677–1688. DOI: https://doi.org/10.1121/1.2000788

Smith, B. L. (1978). Temporal aspects of English speech production: A developmental perspective. *Journal of Phonetics*, *6*(1), 37–67. DOI: https://doi.org/10.1016/S0095-4470(19)31084-8

Smith, B. L., & Westbury, J. R. (1975). Temporal control of voicing during occlusion in plosives. *The Journal of the Acoustical Society of America*, *57*(S1), S71–S71. DOI: https://doi.org/10.1121/1.1995394

Sonderegger, M., Bane, M., & Graff, P. (2017). The medium-term dynamics of accents on reality television. *Language*, *93*, 598–640. DOI: https://doi.org/10.1353/lan.2017.0054

Stevens, K. N., & Keyser, S. J. (2010). Quantal theory, enhancement and overlap. *Journal of Phonetics*, *38*, 10–19. DOI: https://doi.org/10.1016/j.wocn.2008.10.004

Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., & Stokes, M. A. (1988). Effects of noise on speech production: Acoustic and perceptual analyses. *The Journal of the Acoustical Society of America, 84*(3), 917–928. DOI: https://doi.org/10.1121/1.396660

Tabain, M. (2001). Variability in fricative production and spectra: Implications for the hyper-and hypo-and quantal theories of speech production. *Language and speech, 44*(1), 57–93. DOI: https://doi.org/10.1177/00238309010440010301

Tanner, J., Sonderegger, M., & Stuart-Smith, J. (2020). Structured speaker variabiltiy in Japanese stops: Relationships within versus across cues to stop voicing. *Journal of the Acoustical Society of America, 148*(793), 1. DOI: https://doi.org/10.1121/10.0001734

Theodore, R. M., Miller, J. L., & DeSteno, D. (2009). Individual talker differences in voice onset time: Contextual influences. *The Journal of the Acoustical Society of America, 125*(6), 3974–3982. DOI: https://doi.org/10.1121/1.3106131

Toscano, J. C., & McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive Science, 34*(3), 434–464. DOI: https://doi.org/10.1111/j.1551-6709.2009.01077.x

Turk, A., Nakai, S., & Sugahara, M. (2006). Acoustic segment durations in prosodic research: A practical guide. In S. Sudhoff et al. (Eds.), *Methods in empirical prosody research* (Vol. 3, p. 1–28). Berlin, Germany: Walter de Gruyter. DOI: https://doi.org/10.1515/9783110914641.1

Vaughn, C., Baese-Berk, M., & Idemaru, K. (2018). Re-examining phonetic variability in native and non-native speech. *Phonetica, 76*, 1–32. DOI: https://doi.org/10.1159/000487269

Wardrip-Fruin, C., & Peach, S. (1984). Developmental aspects of the perception of acoustic cues in determining the voicing feature of final stop consonants. *Language and Speech, 27*(4), 367–379. DOI: https://doi.org/10.1177/002383098402700407

Warner, N., & Tucker, B. V. (2011). Phonetic variability of stops and flaps in spontaneous and careful speech. *The Journal of the Acoustical Society of America, 130*(3), 1606–1617. DOI: https://doi.org/10.1121/1.3621306

Well, A. D., Myers, J. L., & Lorch, R. F. (2010). *Research design & statistical analysis* (3rd ed.). New York, NY: Routledge. DOI: https://doi.org/10.4324/9781410607034

Westbury, J., & Niimi, S. (1979). An effect of phonetic environment on voicing control mechanisms during stop consonants. *The Journal of the Acoustical Society of America, 65*(S1), S23–S23. DOI: https://doi.org/10.1121/1.2017165

Wright, R. (2004). Factors of lexical competition in vowel articulation. In J. Local, R. Ogden, & R. Temple (Eds.), *Papers in Laboratory Phonology VI* (pp. 75–87). Cambridge, UK: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511486425.005

Xu, Y. (2010). In defense of lab speech. *Journal of Phonetics, 38*(3), 329–336. DOI: https://doi.org/10.1016/j.wocn.2010.04.003

Zeileis, A., & Hothorn, T. (2002). Diagnostic checking in regression relationships. *R News, 2*(3), 7–10. Retrieved from https://CRAN.R-project.org/doc/Rnews/

Zlatin, M. A. (1974). Voicing contrast: Perceptual and productive voice onset time characteristics of adults. *The Journal of the Acoustical Society of America, 56*(3), 981–994. DOI: https://doi.org/10.1121/1.1903359