

**SPEECH RATE EFFECTS ON VOT IN A THREE-CATEGORY LANGUAGE: EVIDENCE
FROM HAKHA CHIN**

July, 2019

Seung Suk Lee¹, Kelly Harper Berkson¹

¹Indiana University Department of Linguistics

ABSTRACT

This study probes the claim made under Laryngeal Realism (e.g., Beckman *et al.*, 2011/2013), by investigating the effect of speech rate on VOT in Hakha Chin. The present study uses the diagnostics of changing the speech rate (Beckman *et al.*, 2011) and examines whether it can be used to find the specified phonological features of a language with a three-way contrast, Hakha Chin. Laryngeal Realism states that the phonological features are privative and that the aspirating language is specified with the feature [spread glottis], while the true voiced language is specified with the feature of [voice]. It has been widely known that the speech rate affects laryngeal stops asymmetrically, and LR authors argue this is because the phonological features are privative rather than binary (e.g. Kessinger & Blumstein, 1997). Methodologically, it attempts to experimentally control the rate variation with the help of a metronome (de Jong, 2001). The present study observes that in Hakha Chin, at a slower rate, the VOT of the prevoiced stop and the aspirated stop increase, while that of the voiceless unaspirated does not, which supports the claims of the LR, but with caveats due to speaker variations.

Key Terms — Laryngeal Realism, Speech Rate Effect on VOT, Hakha Chin, Laryngeal Feature Specifications, Tibeto-Burman Language

1. INTRODUCTION

The present study is concerned with three primary topics: previous theories about the descriptions of laryngeal contrast in comparison with the theory of Laryngeal Realism (Beckman *et al.*, 2011/2013; Brown, 2016; Harris, 1994; Honeybone, 2002/2005; Iverson & Salmons, 1995/2003; Jessen & Ringen, 2002; Vaux & Samuels, 2005), the effect of altering speech rate on the production of VOT, and an under-resourced language called Hakha Chin.

Laryngeal contrast is generally defined as the contrast found in oral stop consonants. For example, English has three places of articulation for oral stops (bilabial, alveolar, and velar) and a two-category contrast (voiced vs. voiceless); in other words, English has /b, p/, /d, t/, and /g, k/. There are languages that have more than two-way contrasts. Some languages have a three-way contrast (e.g. Thai, with the voiced /b/, plain voiceless /p/, and voiceless aspirated /p^h/), and some have a four-way contrast (e.g. Hindi, with the three stop types noted for Thai along with breathy voiced [b^h]). There has been robust debate about how to describe these laryngeal contrasts phonologically. While the traditional way

of describing laryngeal contrast is using the feature [voice] as a binary, meaning that sounds are specified as either [+voice] or [-voice], there is a body of literature referred to as Laryngeal Realism which argues that using [\pm voice] is problematic for some languages (Beckman *et al.*, 2011/2013; Brown, 2016; Harris, 1994; Honeybone, 2002/2005; Iverson & Salmons, 1995/2003; Jessen & Ringen, 2002; Vaux & Samuels, 2005).

The authors who make this argument suggest that laryngeal features are in fact privative rather than binary, and they suggest using another feature [spread glottis] to account for aspiration (Beckman *et al.*, 2011/2013; Brown, 2016; Harris, 1994; Honeybone, 2002/2005; Iverson & Salmons, 1995/2003; Jessen & Ringen, 2002; Vaux & Samuels, 2005). In support of their claim, they discuss various types of evidence such as historical development of Germanic languages and passive voicing effects in German (Beckman *et al.*, 2013; Honeybone, 2002/2005; Jessen & Ringen, 2002). The present study, however, focuses exclusively on what has been referred to as the speech rate effect (Beckman *et al.*, 2011; Schwarz *et al.*, 2018; Islam, 2017).

The speech rate effect refers to the phenomenon where speakers intensify Voice Onset Time (VOT) in slower speech for some consonants but not others (Beckman *et al.*, 2011; Kessinger & Blumstein, 1997; Magloire & Green, 1999; Miller *et al.*, 1986; Pind, 1995). Laryngeal Realists argue that this type of asymmetry, where only some consonants are affected by speech rate, is because the consonants that are affected are marked with phonological feature(s) while the others are not (Beckman *et al.*, 2011; Schwarz *et al.*, 2018; Islam, 2017).

There are three objectives of the present study. First, this study investigates a language with a three-category contrast, Hakha Chin. The previous literatures on both Laryngeal Realism and the speech rate effect on VOT mainly pay attention to two-way contrast languages, meaning that investigation of languages with three- or four-way contrasts are scarce. The present study probes whether speech rate control diagnostics prove useful for languages with a more complicated laryngeal contrast.

Second, the previous literature on the speech rate effect on VOT has employed a methodology whereby speech rate differences were elicited using solely verbal instructions (Beckman *et al.*, 2011; Kessinger & Blumstein, 1997; Magloire & Green, 1999; Miller *et al.*, 1986; Pind, 1995). Such instructions are not replicable. The current study attempts to improve on these methods by controlling speech rate using a metronome (de Jong, 2001).

Third, this study investigates VOT in Hakha Chin, which is an under-documented language. It is a Tibeto-Burman language with 125,000 L1 speakers and 40,000 L2 speakers, spoken primarily in Myanmar (Simons & Fennig, 2018), but also by a large proportion of the >17,000 Burmese refugees living in Indianapolis (Vahnie, p.c.). This present study attempts to shed light on the phonetics and phonology of Hakha Chin by investigating its VOT.

Section 2 presents background research on laryngeal contrast and VOT and introduces both the theory of Laryngeal Realism and the concept of speech rate effects on VOT. It also introduces Hakha Chin in more detail. The methods are outlined in Section 3 and the results are presented in Section 4. Section 5 has discussion of the results and suggests future directions.

2. BACKGROUND

This section discusses laryngeal contrast and VOT, the two types of laryngeal contrast that are standardly found in two-way contrast languages and presents a review of the traditional conventions used when discussing laryngeal specifications. Section 2.4 introduces the theory of Laryngeal Realism which is central to the present study, and Section 2.5 introduces one of the diagnostics that the theory adopts,

which is the speech rate control analysis. Section 2.6 discusses studies of four-way contrast languages that adopt the speech rate control. Section 2.7 introduces Hakha Chin, the language that the present study discusses.

2.1 Laryngeal contrast and VOT

The question of how to describe underlying contrasts in stop consonants has long been debated. For example, there are three pairs of oral stop consonants in English—bilabials (b/p), alveolars (d/t), and velars (g/k). English oral stops are generally thought of as having a voicing contrast, which means that whether or not vibration of the vocal cord occurs marks the difference between the two consonants. One of the most common acoustic measures used to assess these stop consonants is Voice Onset Timing (VOT), which is the temporal measure between the release of the stop closure and the beginning of voicing, with the release of the stop closure serving as the starting point (Lisker & Abramson, 1964, Beckman *et al.*, 2013). VOT can be either negative or positive, depending on where the voicing starts. When a consonant is prevoiced, meaning that there is voicing during the closure prior to the closure release, the result is a negative VOT value. When the consonant is voiceless and aspirated, meaning that vocal fold vibration begins after the closure is released, the result is a positive VOT value.

2.2 Voicing and aspirating languages

Lisker & Abramson (1964) studied VOT in utterance initial position in 11 languages. These languages had either two-way stop contrasts (e.g. [p/b], as in Puerto Rican Spanish and English), three-way contrasts (e.g. [p, p^h, b], as in Eastern Armenian and Thai), or four-way contrasts (e.g. [p, p^h, b, b^h], as in Hindi and Marathi). For the 2-way contrast languages, Lisker and Abramson identified two subgroups: in voicing languages, the voiced consonants have a negative VOT; in aspirating languages, the aspirated consonants have a positive VOT. The consonants that do not have the specified feature—that is, voiceless stops in voicing languages and voiced stops in aspirating languages—have a small positive VOT. Lisker & Abramson (1964) found that Dutch, Hungarian, Puerto Rican Spanish and Tamil have a voice contrast, while English and Cantonese have an aspiration contrast. The observation that languages could have a similar two-way contrast in stops (e.g., /p/ vs. /b/) but show such different patterns of acoustic behavior has fueled a great deal of subsequent phonological investigation.

2.3 Traditional conventions for laryngeal specifications

The distinction between voicing and aspirating languages has contributed to a phonological debate about whether the acoustic differences observed need to be addressed within phonological theory. In other words, does the phonetic difference indicate that there should also be a difference in the phonological features used to specify these consonants? One traditional way to describe the contrast is by using the terms “fortis” and “lenis” (or “tense” and “lax”), which are still widely used in the Germanic languages literature (Honeybone, 2002). Whichever consonant has a contrastive feature would be called “fortis”, and the other “lenis”. In German, for example, since the contrast is one of aspiration (like the English and Cantonese results in Lisker and Abramson (1964)), /p/ would be deemed fortis and /b/ lenis. However, it is difficult to know how this designation scheme could be used cross-linguistically because it implies that the contrast of the voicing languages and that of the aspirating languages are underlyingly

the same, despite their differing acoustic behavior (Honeybone, 2005). On the other end of the spectrum, there have also been efforts to capture the contrasts more thoroughly. Lombardi (1991), for example, proposes utilizing not only [voice] and [aspiration] but also an additional feature, [glottalization].

However, the most widely accepted approach before Laryngeal Realism (Beckman *et al.*, 2011/2013; Brown, 2016; Harris, 1994; Honeybone, 2002/2005; Iverson & Salmons, 1995/2003; Jessen & Ringen, 2002; Vaux & Samuels, 2005) has been binary marking with a voice feature, henceforth referred to as the traditional approach (Chomsky & Halle, 1968; Keating, 1984; Rubach, 1990; Wiese, 1996; among many others). Although this approach marks laryngeal contrast with the feature [\pm voice], which is more informative than the fortis/lenis approach, it still does not capture the phonetic difference between voicing languages and aspirating languages. For example, consider comparing the stop consonants of English and Spanish. Despite the fact that they pattern quite differently, English /b/ and Spanish /b/ are marked with the same feature. These approaches are summarized in Table 1.

Table 1: *Different conventions for laryngeal specification compared*

	English (aspiration contrast)		Spanish (voicing contrast)	
Conventions	p	b	p	b
VOT	positive	Small positive	Small positive	negative
Fortis/Lenis	fortis	lenis	lenis	fortis
[\pm voice]	[-voice]	[+voice]	[-voice]	[+voice]

It is notable that with the fortis/lenis approach, it is potentially misleading that the same term “fortis” is used to describe an aspirated consonant in one language while it is used to describe a pre-voiced consonant in another. It is also apparent that the difference between the two contrast types of English and Spanish is not represented with the traditional approach of [\pm voice].

2.4 Laryngeal Realism

On the other side of the debate, there has been a body of literature developed which is generally referred to as “Laryngeal Realism” (LR) (Beckman *et al.*, 2011/2013; Brown, 2016; Harris, 1994; Honeybone, 2002/2005; Iverson & Salmons, 1995/2003; Jessen & Ringen, 2002; Vaux & Samuels, 2005). Honeybone (2002), who was the first to coin the name Laryngeal Realism, explains that LR “seeks to represent the situation [for laryngeal contrast] in all languages correctly” (p. 131). The LR approach differs in two main ways from the traditional approach in that it argues:

- (1) A [voice] feature alone is not sufficient to account for the difference between languages, and another feature [spread glottis] is required for specifying the aspirated languages;
- (2) Laryngeal features are privative rather than binary.

The term privative is used here in contrast with the term “binary”, meaning that the feature is either present or not. If a feature is represented in binary fashion, it means that the existence of the

feature is marked as positive while the absence is marked as negative. For example, in the case of Spanish, the laryngeal feature [voice] has been traditionally used to mark the contrast between /b/ and /p/. /b/ is marked as [+voice] and /p/ is marked as [-voice]. On the other hand, if the feature is privative, the presence of the feature is marked as [voice], while the absence of the feature is marked as null ([∅]).

One immediate implication of this approach is that it can account for the asymmetric contrast patterns of the two-way contrast languages, as presented in Table 2. Taking the LR approach, /p/ in English gets specified with the feature [spread glottis] (or [sg]) while /b/ does not get specified because LR states that the laryngeal feature is privative. On the other hand, /b/ in Spanish gets specified with the feature [voice], while /p/ remains unspecified.

Table 2: LR compared with the traditional approach

	English (aspiration contrast)		Spanish (voicing contrast)	
Conventions	p	b	p	b
VOT	positive	Small positive	Small positive	negative
Fortis/Lenis	fortis	lenis	lenis	fortis
[±voice]	[-voice]	[+voice]	[-voice]	[+voice]
LR	[spread glottis]	[∅]	[∅]	[voice]

It is notable that unlike the traditional approach, LR accounts for the different VOT patterning of aspirating languages like English and voicing languages like Spanish. Honeybone (2005) states that LR allows for “different predictions as to what is a possible phonological innovation”, as opposed to regarding the two patterns as underlyingly identical (330). There have been many arguments that proponents of LR have used to support the theory in the realms of historical linguistics and German passive voicing. In the following section, we focus on one of the diagnostics that the theory adopts, which is the speech rate control analysis (Beckman *et al.*, 2011).

2.5 Speech rate effect on VOT and Laryngeal Realism

As was previously mentioned, LR makes two main claims:

- (1) A [voice] feature alone is not sufficient to account for the difference between languages, and another feature [spread glottis] is required for specifying the aspirated languages;
- (2) Laryngeal features are privative rather than binary.

The reason why LR makes these claims is due to data from German languages that was analyzed by the early authors of LR. This section focuses on one of the many grounds on which LR authors have based their arguments: the speech rate control analysis (Beckman *et al.*, 2011)

The speech rate control analysis draws from the cross-linguistically reported phenomenon that the production of VOT is affected by speech rate (Kessinger & Blumstein, 1997; Magloire & Green,

1999; Pind, 1995, among others). It has been found that at slower rates of speech, VOT of only certain consonants gets intensified while VOT of the other consonants remains the same (Kessinger & Blumstein, 1997; Magloire & Green, 1999; Pind, 1995). The question of which consonants show the VOT intensification depends on the type of laryngeal contrast that the language exhibits.

It has been reported that for aspirating languages—e.g. English (Miller *et al.*, 1986; Summerfield, 1981) and Icelandic (Pind, 1995)—only the VOT of the aspirated consonants increases at slower rates of speech. VOT values for the unaspirated consonants in these languages are unaffected by changes in speech rate.

This finding is corroborated and expanded by the work of Kessinger & Blumstein (1997), where they investigate the effect of speech rate on three different languages: English, French and Thai. They find that speech rate affects languages differently, depending on the type of contrast exhibited by the language. In particular, it has been found that only the VOT of certain consonant(s) increases while the VOT of the other consonant(s) does not (Kessinger & Blumstein, 1997). The contrast type and the result for each of the three languages are presented in Table 3.

Table 3: *Result of Kessinger & Blumstein (1997)*

	Contrast	Category where VOT increases as speaking rate decreases:	Category where VOT is unaffected :
English	/b, p ^h /	/p ^h /	/b/
French	/b, p/	/b/	/p/
Thai	/b, p, p ^h /	/b, p ^h /	/p/

The findings presented in Kessinger & Blumstein (1997) demonstrate that while certain consonants are affected by changes in speech rate, others are not. Further, the specific consonants which are affected differ across languages. The authors interpret this as evidence that speakers have selective and asymmetrical control over the duration of VOT.

Further speech research focused on bilingual speakers of Spanish and English (Magloire & Green, 1999) has provided additional support for the proposal that speakers have intentional control over VOT. Magloire & Green (1999) report that when speaking at a slower rate, the bilingual speakers of Spanish and English lengthen the VOT of different consonants depending on what language they are speaking. They lengthened the (negative) VOT of /b/ when speaking in Spanish, whereas they lengthened the (positive) VOT of /p^h/ when speaking in English. In parallel with the speech rate effect, studies on clear speech/speech enhancement also show that the speakers intentionally choose to intensify VOT of certain consonants (see Kang & Guion, 2008; Cho & Jun, 2000, for example).

To sum up, the findings in the previous literature on the speech rate effect shows that they all report asymmetrical results, where some consonants show VOT values which vary according to speech rate, while others are unaffected by changes in speech rate. It is now relatively well established that the consonants which show VOT increases are dependent on the type of contrast exhibited by the language.

Prior to LR, the authors have attributed their findings to different reasons. Kessinger and Blumstein (1997) argue that the reason why speakers alter VOT durations for one type of consonant while maintaining consistent VOT durations for the other is in order to preserve the phonetic contrast

between the two consonants. They claim that the consonant category for which VOT does not change functions as “the phonetic anchor” (145) maintaining the contrast while the VOT of the other varies (Kessinger & Blumstein, 1997). In other words, the reason why the VOT of certain consonants are lengthened is to maximize the contrasts between the consonant categories.

However, Beckman *et al.* (2011) point out that it does not have to be asymmetrical if it is only for maximizing the contrast. For example, a French speaker could lengthen short positive VOT while also increasing negative VOT (that is, closure voicing), which would in fact generate a bigger contrast to serve the purpose (Beckman *et al.*, 2011). On the contrary, Magloire & Green (1999) argue that speakers are very selective in increasing VOT in their data from Spanish-English bilinguals. They find that which consonants get affected depends on the type of laryngeal contrast that the language has.

Beckman *et al.* (2011) connect these asymmetrical results to LR and offer another interpretation with the data found from Central Standard Swedish (CS Swedish). They claim that the previously reported asymmetrical VOT lengthening of consonants occurs because only the featurally marked consonants are affected in a slower rate of speech whereas the unmarked consonants are not (Beckman *et al.*, 2011). Taking English for example, as mentioned before, the aspirated consonants are affected at the slower rate and the unaspirated consonants are not. Beckman *et al.* (2011) argue that this is because English essentially has a contrast related to the aspiration and only the aspirated consonants are marked with the feature [spread glottis] (or [sg]), while the unaspirated consonants are not marked at all.

In their study, Beckman *et al.* (2011) replicate the experiment of varying speech rate with Central Standard Swedish speakers. CS Swedish is typologically rare because it is found to have both prevoiced consonants (e.g. /b/) and aspirated consonants (e.g. /p^h/), but not the voiceless unaspirated consonant (e.g. /p/) (Helgason & Ringen, 2008). This is a finding that is contrary to the existing knowledge of laryngeal typology because no other two-way contrast languages have been reported to have both prevoicing and aspiration (Iverson & Salmons, 1995).

Their result shows that the VOT contrast between prevoiced stops and aspirated stops is clear at the normal rate of speech, and yet the CS Swedish speakers lengthen both VOT of prevoiced consonants and aspirated consonants when they speak more slowly, meaning the contrast becomes even more extreme during slower speech.

Interpreting their data using the arguments made in LR, Beckman *et al.* (2011) claim that the change in VOT does not serve a crucial purpose for listeners because the contrasts are already very clear and easily perceivable even in fast speech (Beckman *et al.*, 2011). Rather, they argue that regardless of the speaker’s intention, the change of VOT is a by-product of the articulatory dynamic (Beckman *et al.*, 2011). They argue that in CS Swedish, contra the traditional expectation that features should be minimal, both [voice] and [sg] features are present and privatively marked in CS Swedish.

While the effect of speech rate on the production of VOT has been reported repeatedly in previous literature, it is Beckman *et al.* (2011) that adopt this phenomenon in the LR literature, providing a useful diagnostic which can be deployed in the testing of additional languages. Germanic languages have been the main focus of a great deal of LR literature, but in order to develop a broader understanding of how generalizable the findings are, data from additional languages is desirable.

By bringing the speech rate effect literature and LR literature together, Beckman *et al.* provide a useful diagnostic which can potentially be used cross-linguistically, especially with non-Germanic languages and even with those that do not feature two-way contrasts.

2.6 Laryngeal Realist analysis of speech rate control on four-way contrast languages

It has been noted very recently that the LR literature lacks discussion of languages with more complicated laryngeal contrasts, and there have been recent attempts to implement speech rate control diagnostics within four-way contrast languages such as Nepali (Schwarz *et al.*, 2018) and Bangla (Islam, 2017).

First, Schwarz *et al.* (2018) adopt the speech rate control diagnostics of Beckman *et al.* (2011) to investigate Nepali. Nepali is an Indic language with a four-way contrast (voiceless aspirated, voiceless unaspirated, voiced and voiced aspirated, e.g. /p^h, p, b, b^h/) across four places of articulation (bilabial, alveolar, retroflex and velar) (Schwarz *et al.*, 2018). It has previously been proposed that the voiced aspirated consonants in the four-way laryngeal contrast are specified for both [voice] and [sg]. They find that prevoicing increases in slower speech for the voiced consonants (e.g. /b/ and /b^h/) but not for the voiceless ones; similarly, the aspiration increases in slower speech for the aspirated consonants (e.g. /p^h/ and /b^h/). These findings seem to suggest that the voiced aspirated consonants are indeed specified for both [voice] and [sg].

Second, Islam (2017) reports data from Bangla, another Indic language with the same four-way contrast as Nepali. In Bangla, while aspiration of the aspirated consonants (e.g. /p^h/ and /b^h/) increases in slower speech, prevoicing in the voiced consonants is not significantly affected by the rate of speech (Islam, 2017). Instead, the findings reveal that speakers actively lengthen breathiness for the voiced aspirates and argues that the language should be marked with an additional feature, [breathy] (Islam, 2017). His data supports LR only partially, as he finds that prevoicing is not affected by speech rate, though it seems to be an existing feature in the language.

The contradictory data from Schwarz *et al.* (2018) and Islam (2017) casts doubt on the diagnostics of speech rate control and calls for further research with other languages. It is with this in mind that the current study is presented. In the following section we turn to a review of Hakha Chin, a Tibeto-Burman language which features a three-way laryngeal contrast between voiced, voiceless unaspirated, and voiceless aspirated stop consonants. The main goal of the present study is to apply speech rate control diagnostics on Hakha Chin to probe whether the diagnostics will prove informative when used with a three-way contrast language.

2.7 Hakha Chin and speech rate control diagnostics

Hakha Chin is a Tibeto-Burman language with 125,000 L1 speakers and 40,000 L2 speakers, spoken primarily in Myanmar (Simons & Fennig, 2018). It is also spoken by a large number of Chin refugees in the US. According to the statistics report published in 2015 by the Bureau of Population, Refugees, and Migration, there were 18,386 refugees from Burma in the year of 2015 alone. However, despite the great need to study the language, Hakha Chin is under-resourced. In particular, the focus of existing research has largely been on the syntax and the semantics of this language (see Kathol, 2003; Bedell, 1996; and Bedell & VanBik, 2000, for example). On the contrary, the phonetics of this language have not yet been researched in detail. Recorded speech data of this language have not yet been collected to any significant degree, and acoustic analysis is almost nonexistent. With these facts in mind, the present study attempts to shed light on the phonetics and phonology of Hakha Chin by investigating its VOT.

Hakha Chin is a language with a stop contrast which can contribute to the discussion of the validity of the speech rate control diagnostics. It has four places of articulation and a partial three-way

contrast, as summarized in Table 4. Note that the symbols used are mainly phonetic, but that the alveolar consonants are written with a double /tt/, in keeping with Hakha Chin orthography.

Table 4: *Hakha Chin stop contrast*

	Bilabial	Dental	Alveolar	Velar
Voiced	/b/	/d/		
Voiceless Unaspirated	/p/	/t/	/tt/	/k/
Voiceless Aspirated	/p ^h /	/t ^h /	/tt ^h /	/k ^h /

At the bilabial and dental places of articulation, the Hakha Chin inventory resembles that of Thai, which has been investigated and discussed in the previous literature (Kessinger & Blumstein, 1997; Iverson & Salmons, 1995; Beckman *et al.*, 2011). Based on the results of the previous work, it has been claimed that Thai has both [voice] and [sg] specified (Iverson & Salmons, 1995), as summarized in the table below.

Table 5: *Feature representation of Thai, proposed by Iverson and Salmons (1995)*

Contrast	Voiced /b/	Voiceless Unaspirated /p/	Voiceless Aspirated
Laryngeal feature	[voice]	[ø]	[sg]

As discussed in the previous section, the prediction that languages with more than two contrasts are specified with both [voice] and [sg] features was corroborated by the Nepali data (Schwarz *et al.*, 2018) but not by the Bangla data (Islam, 2017).

Therefore, investigating speech rate effects on VOT in Hakha Chin would both shed some light on the phonetics of this under-described language and further verify whether the diagnostics of speech rate control is useful for understanding which features are specified or unspecified in the laryngeal contrast. In addition, the current study is also valuable because Thai is the only other three-category language for which data has been collected.

3. METHODS

In this section, hypotheses of the present study will be addressed and the methods for the present study will be discussed. In particular, it will discuss the methodology of using a metronome to control speech rate in a replicable way, which is different from the methodology used in previous literature.

3.1 Hypotheses

Two hypotheses can be made in the present study with regards to the validity of speech rate control diagnostics. First, this study hypothesizes that in slower speech, VOT of the voiced stops will increase for bilabials and dentals, and VOT of the aspirated stops will increase for all four places of articulation. If the speech rate effect is found, then it can be interpreted using Laryngeal Realism that these stop categories are specified with the features [voice] and [sg] respectively. Second, this study also hypothesizes that VOT values for the voiceless unaspirated stops will not be affected by the speech rate. In the LR theoretical framework, such a finding would confirm that both features are privatively marked.

3.2 Difference from the previous literature

The present study adopts the diagnostics of speech rate control (Beckman *et al.*, 2011; Kessinger & Blumstein, 1997; Magloire & Green, 1999; Miller *et al.*, 1986; Pind, 1995) to investigate whether speech rate affects VOT in the stop consonants of Hakha Chin. However, the present study differs from the previous literature in two major ways: first, it uses a metronome to prompt different speech rates; and second, it uses a word list instead of a carrier sentence.

First, it uses a metronome because it is difficult to make comparisons between the results found in previous research because the authors generally used verbal instructions to vary the speech rate, potentially rendering different speech rates for each of the previous studies. For example, Beckman *et al.* (2011) use verbal instructions such as “increase [the] speaking rate without sacrificing accuracy” (8); whereas Kessinger & Blumstein (1997) use different wording: “subjects were encouraged to speak as quickly as possible without forsaking accuracy, as if speaking to a friend who was leaving the room” (147). Unlike these two studies, Magloire & Green (1999), Miller *et al.* (1986) and Pind (1995) attempt to set five distinct speech rates, still using only verbal instructions: normal rate, two times normal rate, four times normal rate, $\frac{1}{2}$ of normal rate, and $\frac{1}{4}$ of normal rate. There is an inherent assumption underlying these instructions that speakers are able to consistently produce speech at distinct rates, and it is not clear that this is true—what is considered “fast” may differ not only across participants but even within the speech of a single participant at different points in time. This issue was even noted by the authors. For example, the technique used by Miller *et al.* (1986) and replicated by Magloire & Green (1999) and Pind (1995) was reported to produce a substantial range of rates across speakers. Pind (1995) also notes this speaker variability and chooses to use syllable durations of the produced data as the independent variable, rather than using the speech rate categories, such as “fastest, fast, normal, slow and slowest”.

While these studies allowed speakers to control their own speech rate, other research—for instance de Jong (2001)—finds this method problematic because speakers tend not to be able to choose more than three distinct speech rates and each speaker tends to choose different rates for the same, which makes it infeasible to compare across speakers (de Jong, 2001). Instead, de Jong (2001) uses a metronome to elicit more coherent speech data from the speakers. In research on speech rate effect on the VOT in Zulu, Midtlyng (2011) also uses a metronome in order to prompt different speech rates. In addition to these studies, in research on stuttering metronomes are standardly used to prompt speech at different rates in an experimentally controlled fashion (see for instance, Davidow *et al.*, 2010; Davidow, 2014).

In order to elicit uniform speech rates, then, the present study uses a metronome to elicit three distinct speech rates from speakers. With the help of one of my Hakha Chin language consultants,

speaker T, the slow rate was set at 50 beats per minute (bpm), the normal rate at 100 bpm, and the fast rate at 150 bpm. These rates were selected after some trials with one of the participants.

Second, while the previous studies have used a carrier sentence, such as “I say [word]” (Midtlying, 2011), or “Say [word] again” (Kessinger & Blumstein, 1997), the current study elicited monosyllabic words where the contrasting consonants are in the initial position. The reason for this is to investigate the contrastive consonants in word-initial position because previous research has suggested that the acoustic features of consonants in Hakha Chin may actually vary when they occur within a sentence instead of in isolation. This phenomenon occurs in other languages as well—for example, Van Bik (2003) finds that in a related language, Burmese, a voiceless obstruent can be realized as voiced depending on the environment via a process called “junctural voicing” or “parasitic voicing.”

3.3 Participants

Three native speakers of Hakha Chin participated in producing the data for this experiment (one female, two male). They are henceforth referred to as P, Z, and T. All three participants are undergraduate students at Indiana University in Bloomington and are members of the Indianapolis Burmese refugee community. They are around 20 years of age and are fluent speakers of Hakha Chin.

It is to be noted that T helped with the initial experimental design, especially when it came to choosing the speech rates used for collecting data in this experiment, and therefore had some ideas about the experiment prior to producing the data.

3.4 Stimuli

A stimulus word with a syllable structure /CV:/ was chosen for each of the ten stop consonants of Hakha Chin. Six words with the same structure were chosen additionally as filler words. The filler words have six different nasal stop consonants instead of the oral stops in the initial position. For most of the items, the vowel following the stop consonant was controlled to be a long /a/ vowel, but for one token for the target items and one token for the filler words, a word with a /CV?/ structure was chosen due to gaps in the lexicon. The complete stimuli word list is presented below in Table 6.

3.5 Procedure

The experiment began with a practice session since participants had to perform a rather difficult task which required them to process three forms of stimuli simultaneously. They were first presented with a word written in Hakha Chin orthography that appeared on a computer monitor along with its English definition. Stimuli were presented in a pseudo-random order using PsychoPy (Peirce, 2007). At the same time that the word was presented, participants heard a steady metronome beat over a set of headphones. Earbud headphones were used, and participants had only one earbud in, so that they heard the metronome using only one ear. This was because when we have headphones on both ears, we cannot hear ourselves speak as clearly, and this can affect how natural our productions are. The metronome was set to one of the three experimental rates. Finally, participants were also aided by visual flashing on the screen of the metronome. On the metronome, there were four boxes that flashed one after another, one flash for each beep. This visual cue helped speakers to keep track of how many words they had produced, because if they started to produce the first word as the first box flashed, they would know that

Table 6: *Stimuli word list and English translations*

Hakha Chin	English
ba	<i>turnip</i>
pa	<i>father</i>
pha	<i>to arrive</i>
dah	<i>as in [zei dah] 'what' or electricity</i>
ta	<i>possession</i>
tha	<i>strength</i>
tta	<i>brother</i>
ttha	<i>good</i>
ka	<i>mouth</i>
kha	<i>bitter</i>
ma (mak)	<i>stop; divorce</i>
na	<i>buffalo</i>
nga	<i>fish</i>
hma	<i>wound</i>
hna	<i>plural marker</i>
hngah	<i>to wait</i>

as the fourth box flashed, they were producing the word for the fourth time. A complete overview of the experimental set-up is presented in Figure 1.

The words that were used in the practice session had a different syllable structure than the words used in the experiment session. The practice words include [lu] *head*, [lung] *feeling*, [sen] *red*, [caw] *cow*, and [hnar] *nose*.

Once the speakers familiarized themselves with the experiment task, they started the experiment at their will. For each word, they were allowed to listen to the metronome beat as long as they wished before starting to produce the word. They were asked to produce each word eight times, as they could easily count the number of the repetition because the number of the boxes on the metronome screen was four. Although participants were asked to repeat each token eight times, they were also told explicitly that they could produce more or fewer than eight repetitions since it was more important for them to produce the words at the planned rate than to meet the requirement for a certain number of repetitions. The speakers were recorded on a Marantz PMD661 solid state audio recorder with an ElectroVoice Cobalt 9 microphone in a WhisperRoom sound-attenuated booth on the Indiana University, Bloomington campus.

Since there were 10 stimulus words, each repeated (appr.) eight times at three different rates of speech, recorded by three speakers, a total of 720 tokens were expected. There are 813 tokens in the actual results because speakers tended to produce more than eight repetitions.

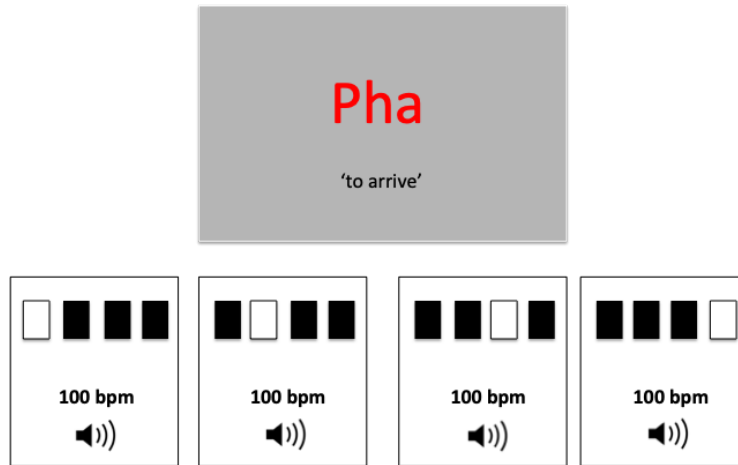


Figure 1: Overview of the experimental set-up. The participants were asked to produce the word on the monitor (e.g. [pha]). They were aided by the auditory and visual cue of the metronome. On the screen of the metronome, there are also four small boxes. Every time the metronome beat starts, the small box flashes from left to right along with the beep sound. The visual cue helps them count the number of times they have produced the word. If they started when the leftmost box on the metronome screen flashed (the leftmost screen image above), they would know that when the final box flashed, they had produced the word four times.

4. RESULTS

This section reports the results of the study. The length of VOT was annotated manually using Praat (Boersma & Weenink, 2018) and duration was measured using a Praat script. All figures included herein were made with Excel. The results of the pilot study have some limitations because there are not enough speakers or number of tokens per speaker at this point in time. The results do show a general trend, though it is subject to speaker variations. In Section 4.1, the general trends are reported and in Section 4.2, the individual speaker data are presented.

Before presenting the graphs, it is necessary to mention that the graphs are made with a unified color scheme. The color blue represents the aspirated consonants (e.g. /p^h/), which are also represented by the name of the feature [sg]. The color green represents the voiced consonants (e.g. /b/), which are also represented by the feature [voice]. Finally, the color red represents the voiceless unaspirated consonants (e.g. /p/), which are considered as unmarked in this study. The numbers 50, 100, and 150 are the number of beats per minute (bpm), representing the slowest, the normal, and the fastest rate of speech, respectively.

4.1 General results

Figure 2 shows the average VOT for all three speakers across four places of articulations divided by three rates of speech (though recall that the voiced consonants only exist for the bilabial and dental places of articulation). As expected, voiced stops are produced with a long negative VOT (meaning that there is voicing during the closure) the aspirated stops have long positive VOT, and the plain voiceless stops have short positive VOT values. These general trends do differ across the different speech rates, however.

Figure 2 can be divided into three sets of bar graphs, each set representing the slow, normal and fast rate of speech respectively. First, it is clear that the VOT for the unmarked category, that is the voiceless unaspirated consonants, does not change across the rate of speech, but instead hovers around 20 ms across all speech rates. On the other hand, the aspirated consonants represented by the blue color are affected by the rate of speech. VOT is the longest in the slowest rate of speech (150ms) and it is the shortest in the fastest rate of speech (50ms). The voiced consonants represented by the green color show a similar pattern where the longest VOT is in the slowest rate of speech (150ms) and the shortest is in the fastest rate of speech (50ms). In Section 3.1 above, it was hypothesized that the aspirated consonants (blue) and the voiced consonants (green) would be affected by the speech rate but not the voiceless unaspirated consonants (red). What we see in Figure 2 seems to suggest that the hypotheses were right. Indeed, there is a visible difference for the aspirated consonants and the voiced consonants between the fastest and the slowest rate of speech, while the voiceless unaspirated consonant seems to stay around the same. According to our LR analysis, it suggests that the aspirated consonants and the voiced consonants are marked with phonological features ([sg] and [voice] respectively), while the voiceless unaspirated is not marked with a feature.

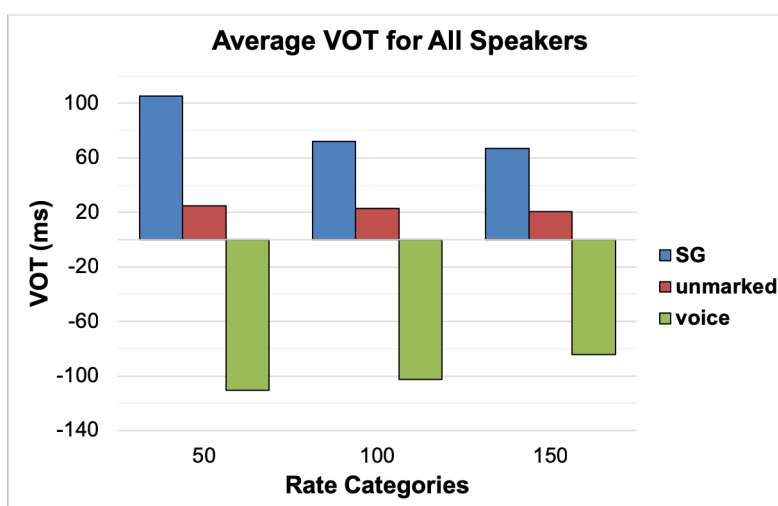


Figure 2: Average VOT for all 3 speakers

Figure 3 below divides the results into four places of articulations. On the left half, the VOT of the bilabials and the dentals are plotted, and on the right half, the VOT of the alveolars and the velars are plotted. The voiced consonants, represented by the green bars, only exist for the bilabials and dentals which are on the left half of the graph. For each place of articulation, again there are three sets of bars representing the three rates of speech.

Figure 3 suggests that the speech rate effect observed from Figure 2 can be found for each of the four places of articulation. For each of the four places of articulation, and for both voiced consonants and aspirated consonants, the longest VOT tends to be in the slowest rate of speech (50) and the shortest VOT in the fastest rate of speech (150). One possible exception is the alveolar stop category, where the VOT of aspirated stops at 100 and 150 bpm is similar. Figure 3 also shows that the red bars which represent the unmarked category are relatively unaffected by the rate of speech for all four places of articulations.

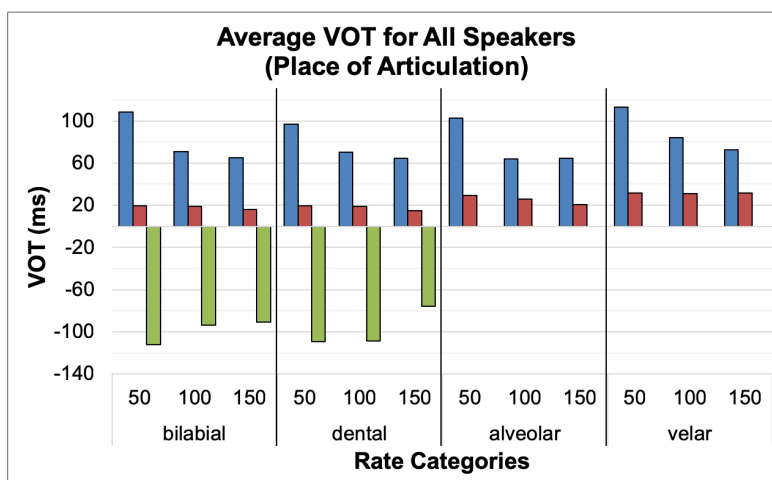


Figure 3: *Average VOT for all three speakers divided by four places of articulations*

Figures 2 and 3 show that there is a speech rate effect in the stop consonants of Hakha Chin. It seems that in slower speech, the VOT for voiced and aspirated consonants is lengthened, while in the fast speech, the VOT for those consonants is shortened. This result is consistent with what has been found for Thai (Kessinger & Blumstein, 1997) and aligns with what would be expected in the LR framework. This trend will be henceforth referred to as the speech rate effect and it can be found in all four places of articulations. It also seems that the VOT for the voiceless unaspirated consonants does not get affected by the rate of speech, also for all four places of articulations.

While it is encouraging that the speech rate effect seems to exist for the stop consonants for Hakha Chin, Figure 2 and Figure 3 show the average of three speakers and therefore does not show the variations between the speakers. In fact, if we look at the result from each of the three speakers, some variation seem to exist, which is contrary to the general trend observed from Figure 2 and 3. The following section discusses this speaker variation in detail.

4.2 Speaker Variations

For the speaker variations, only the left half of Figure 3 is plotted for all three speakers. That is because the present paper mostly focuses on the three-way contrast of Hakha Chin and the voiced consonants only exist for the bilabials and the dentals which are plotted on the left half of Figure 3. Figures 4, 5, and 6 respectively show the results from the speakers P, Z and T.

Figure 4 shows the average VOT for speaker P. The results can be divided into four quadrants. The upper quadrants show the positive VOT and the lower quadrants show the negative VOT. The two quadrants on the left show the VOT of the bilabials and the two on the right show the VOT of the dentals. Several observations can be made about these data.

First, it seems that the VOT of the voiceless unaspirated consonants is not affected by the speech rate. In the left upper, the right upper, and the right lower quadrants, VOT is the longest in the slowest rate of speech and the shortest in the fastest rate of speech. Differences are clearest at the extremes—that is, VOT values at 50 bpm are consistently greater than those at 150 bpm. The left lower quadrant is where the aberration can be observed. The shortest VOT is not in the fastest rate of speech, but in the normal rate of speech.

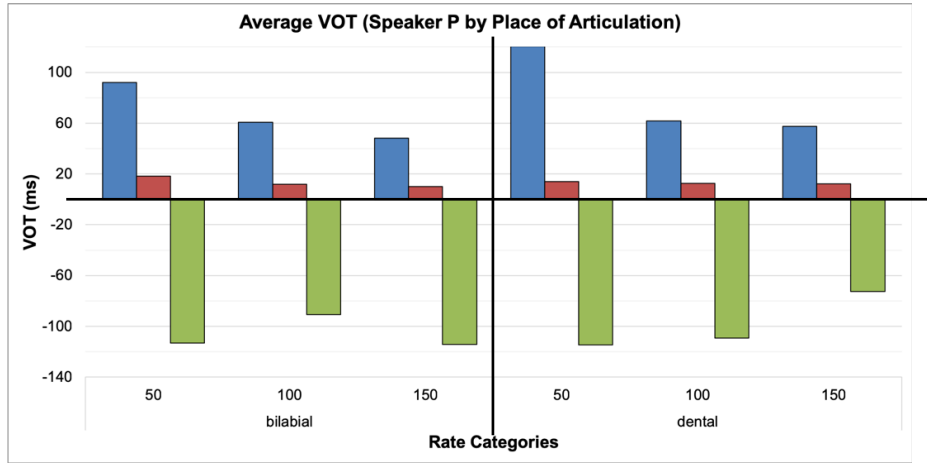


Figure 4: Average VOT for P for bilabials and dentals

Figure 5 shows the average VOT for speaker Z. It is divided into four quadrants in the same fashion. Again, an aberration is observed, though in another quadrant for this speaker.

For speaker Z, the expected speech rate effect can be observed in the left upper, left lower, and right lower quadrants, but the pattern in the right upper quadrant deviates from what was predicted. For speaker Z's aspirated dental consonants, the shortest VOT is not observed at the fastest rate of speech but in the normal rate of speech. While we cannot determine why these unexpected patterns emerge at this point in time, we can say that for each of the speakers we have looked at so far there has been something unusual. Indeed, when we look at the individual data for the third speaker T, we see something similar.

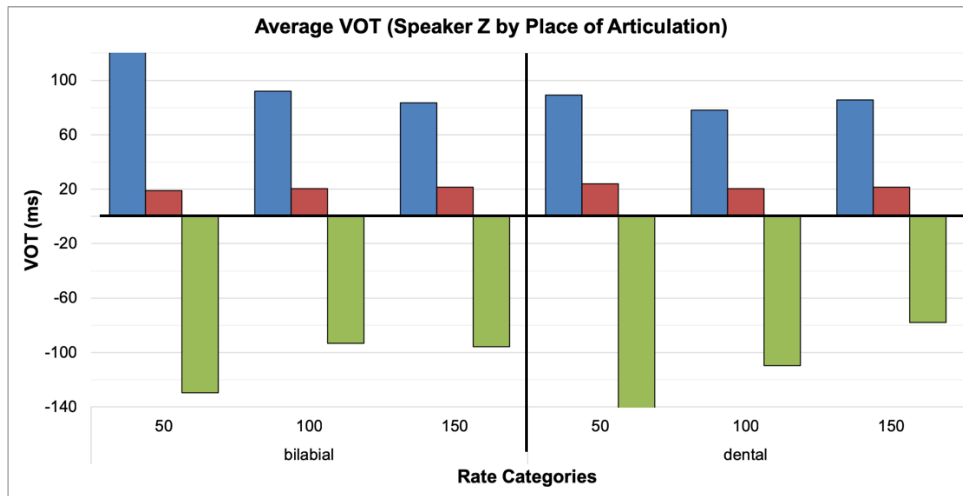


Figure 5: Average VOT for Z for bilabials and dentals

Figure 6 shows the average VOT values for speaker T. Like Figures 4 and 5, it is also divided into four quadrants. The speech rate effect is observed only from two of the four quadrants.

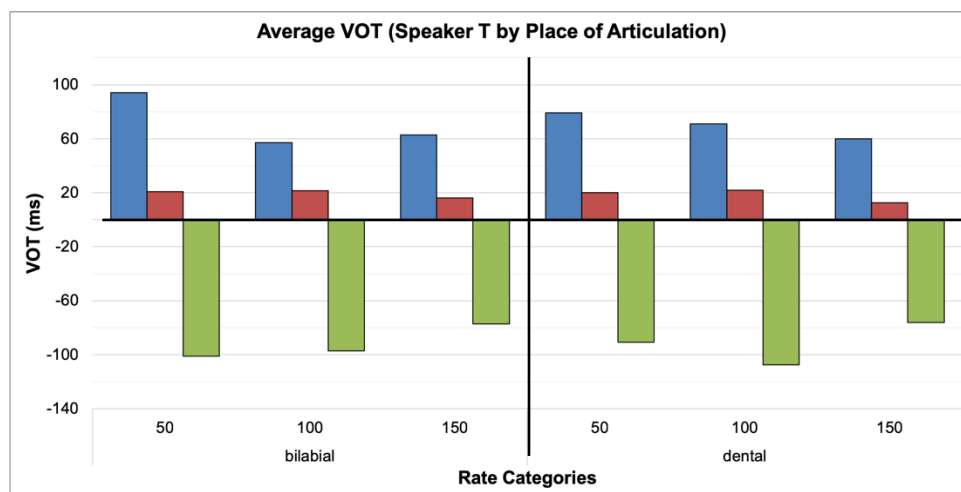


Figure 6: Average VOT for T for bilabials and dentals

For speaker T, the speech rate effect can be observed in the right upper and the left lower quadrants. In the left upper quadrant, however, the aspirated bilabials, the shortest mean VOT is not in the fastest rate of speech but in the normal rate of speech. In the right lower quadrant, the voiced dentals, the longest VOT is not in the slowest rate of speech but in the normal rate of speech.

To sum up Section 4, the results imply that there is a speech rate effect in Hakha Chin stop consonants though there are caveats. As Figure 2 and 3 show, when VOT values are averaged across all three speakers and all four places of articulation, the VOTs of voiced and aspirated consonants are the longest in the slowest speech and the shortest in the fastest speech while the VOT of voiceless unaspirated stops remains unaffected. These findings align with the predictions that were made and support an LR analysis: since the voiced and the aspirated are the ones that are affected by the speech rate, they are each specified with [voice] and [spread glottis] feature; and the voiceless unaspirated stop is not affected, thus suggesting that both of the features are marked privatively.

However, the speech rate effect is not so robust if we look at the result of each speaker separately. For all three speakers, there was at least one category of consonants for which VOT patterns differently across the three rates of speech. For speaker P, it was voiced bilabials; for speaker Z, it was aspirated dentals; and for speaker T, both the aspirated bilabials and the voiced dentals patterned differently than expected. These results are further discussed in Section 5, where limitations and future directions for the study are also discussed.

5. DISCUSSION & CONCLUSION

In this section, the preliminary results from Section 4 are discussed and future directions for the present study are laid out. In Section 5.1, the results are discussed with regards to the hypotheses made in Section 3.1 and Section 5.2 discusses the limitations of the study and the future research questions.

5.1 Discussion

In Section 3.1, two hypotheses were made. First, I hypothesized that VOT of the voiced stops and the aspirated stops would be intensified at slower speech rates. Adopting the LR framework, such a finding would be interpreted as meaning that Hakha Chin has both [voice] and [sg] features that specify the

voiced and aspirated stops respectively. Second, I also hypothesized that VOT of the voiceless unaspirated stops will not be affected by the speech rate, which would confirm that both features are privatively marked.

To start with the first hypothesis, it was found that VOT values of the voiced and the aspirated stops are indeed intensified in the slower speech rates for all four places of articulations, when the average results from all three speakers are considered. In other words, the speech rate effect was found from the data on average which implies that voiced and aspirated consonants are marked with [voice] and [sg] features in Hakha Chin. However, this is perhaps too hasty a conclusion since there are variations among the speaker. Every speaker had some categories that patterned differently from the average results, which is discussed in detail in Section 4.2.

On the other hand, as to the second hypothesis, it is found that VOT of the voiceless unaspirated stops indeed did not change across the three speech rates. This has not been confirmed by statistical analysis yet but at least preliminarily, it can be interpreted as suggesting that voiceless unaspirated stops are not marked by any features, which supports LR's claim that laryngeal features are privatively marked (Beckman *et al.*, 2011/2013; Brown, 2016; Harris, 1994; Honeybone, 2002/2005; Iverson & Salmons, 1995/2003; Jessen & Ringen, 2002; Vaux & Samuels, 2005).

5.2 Limitations and future directions—Korean stop consonants

The objectives of the study, as mentioned in Section 1.2, were to probe whether the speech rate diagnostic proved useful for a three-way contrast language, to improve the methodology so that it can be more replicable, and to shed light on an under-resourced language (Hakha Chin).

As briefly mentioned in the previous section, although the speech rate effect does seem to emerge when the data are considered on average, there are some variations among the speakers as well as across places of articulation. The number of speakers and the number of stimuli produced by each speaker are the obvious limitations of this study. In order to see a more robust effect, an obvious next step would be collecting more data from more speakers. If variation among the speakers remains a theme when more data are collected, it would be worth investigating whether there are any subgroups that pattern similarly. This is because Hakha Chin speakers in Indiana tend to have various sociolinguistic backgrounds, and there are several different dialects represented in Indiana as well.

With the use of a metronome, this study has innovated a replicable method that can be used to investigate other languages. Thus, in addition to collecting more Hakha Chin data, another future direction of this study is replicating this study with Korean. Not only is it the native language of the author of this study, but Korean is also famous for their typologically rare stop consonants. Korean also has a three-way stop consonant contrast, but the contrast does not involve voicing. Instead, it has aspirated, fortis, and lenis stop consonants (Cho & Jun, 2000). It has been reported that VOT difference alone does not mark this three-way contrast (Cho & Jun, 2000), so thinking through how LR could be related to the Korean stops is very interesting. The research that is forthcoming will investigate whether the features that mark the Korean stop contrast will be affected by the rate of speech.

Finally, the present study investigated the phonetics and phonology of Hakha Chin, which are areas that have been relatively neglected and understudied. It provides VOT data for the stop consonants of Hakha Chin and discusses the phonological features of the laryngeal contrast of Hakha Chin, which can be investigated further by future researchers.

ACKNOWLEDGMENTS

This work was supported in part by the Indiana University Office of the Vice Provost for Undergraduate Education, the IU College of Arts and Sciences Ostrom Grants Program, the IU Department of Linguistics, and a summer research grant from Hutton Honors College. I also would like to thank the three Hakha Chin speakers Peng Hlei Thang, Zai Sung and Thomas Thawngza for their precious participation in this project.

REFERENCES

- Beckman, J., Helgason, P., McMurray, B., & Ringen, C. (2011) "Rate effects on Swedish VOT: Evidence for phonological overspecification," *Journal of Phonetics*, **39**(1), 39-49.
- Beckman, J., Jessen, M., & Ringen, C. (2013) "Empirical evidence for laryngeal features: Aspirating vs. true voice languages," *Journal of Linguistics*, **49**(02), 259-284.
- Bedell, G and VanBik, K. (2000) "Lexical and syntactic causatives in Lai," Papers from the 10th annual meeting of the Southeast Asian Linguistics Society (SEALS), 75-86.
- Bedell, G. (1996) "Agreement in Lai," Papers from the 5th annual meeting of the Southeast Asian Linguistics Society (SEALS), 21-32.
- Boersma, P and Weenink, D. (2018) Praat: doing phonetics by computer [Computer program]. Version 6.0.43, retrieved 8 September 2018 from <http://www.praat.org/>
- Brown, J. (2016) "Laryngeal assimilation, markedness and typology," *Phonology*, **33**(03), 393-423.
- Cho, T. and Jun, S. (2000) "Domain-initial strengthening as enhancement of laryngeal features: Aerodynamic evidence from Korean," *University of California Working Papers in Phonetics*, **99**, 57-69.
- Chomsky, N. and Halle, M. (1968) *The Sound Pattern of English*. (Harper & Row, New York).
- Davidow, J. (2014) "Systematic studies of modified vocalization: the effect of speech rate on speech production measures during metronome-paced speech in persons who stutter," *Language and Communication Disorder*, **49**(1), 100-112.
- Davidow, J. H., Bothe, A. K., Richardson, J. D., and Andreatta, R. D. (2010) "Systematic studies of modified vocalization: Effects of speech rate and instatement style during metronome stimulation," *Journal of Speech, Language, and Hearing Research*, **53**, 1579-1594.
- de Jong, K. J. (2001). "Rate-Induced Resyllabification Revisited," *Language and Speech*, **44**(2), 197-216.
- FY15 Refugee Admissions Statistics. (2015, December 31). Retrieved March 06, 2018, from <https://2009-2017.state.gov/j/prm/releases/statistics/251285.htm>
- Harris, J. (1994). *English sound structure* (Blackwell Oxford).
- Helgason, P. and Ringen, C. (2008). Voicing and aspiration in Swedish stops. *Journal of Phonetics*, **36**(4), 607-628.
- Honeybone, P. (2002). "Germanic Obstruent Lenition: some Mutual Implications of Theoretical and Historical Phonology," PhD thesis, University of Newcastle upon Tyne, Newcastle upon Tyne, England, UK.
- Honeybone, P. (2005). "Diachronic evidence in segmental phonology: the case of obstruent laryngeal specifications," in *The Internal Organization of Phonological Segments*, edited by M. van Oostendorp and J. van de Weijer (Mouton de Gruyter, Berlin), pp. 319-354.

- Islam, J. (2017) "Laryngeal Feature of Voiced-Aspirated Stops: Empirical Evidence from Bangla," Poster presented at the 47th Poznań Linguistic Meeting, 2017 Sept 18-20, Poznań, Poland.
- Iverson, G. K. and Salmons, J. C. (1995) "Aspiration and laryngeal representation in Germanic," *Phonology*, **12**(03), 369-396.
- Iverson, G. K. and Salmons, J. C. (2003a) "Laryngeal enhancement in early Germanic," *Phonology*, **20**(01), 43-74.
- Iverson, G. K. and Salmons, J. C. (2003b) "Legacy Specification in the Laryngeal Phonology of Dutch," *Journal of Germanic Linguistics*, **15**(01).
- Jessen, M. and Ringen, C. (2002) "Laryngeal features in German," *Phonology*, **19**(02), 189-218.
- Kang, K. and Guion, S. G. (2008) "Clear speech production of Korean stops: Changing phonetic targets and enhancement strategies" *The Journal of the Acoustical Society of America*, **124**(6), 3909-3917.
- Kathol, A. (2003) "Cooperating Constructions in Lai "Lexical Insertion," in *Proceedings of the 10th International Conference on Head-Driven Phrase Structure Grammar*, pp. 203–221.
- Keating, P. A. (1984) "Phonetic and phonological representation of stop consonant voicing," *Language*, 286–319.
- Kessinger, R. and Blumstein, S. (1997) "Effects of speaking rate on voice-onset time in Thai, French, and English," *Journal of Phonetics*, **25**(2), 143-168.
- Lisker, L. and Abramson, A. S. (1964) "A Cross Language Study of Voicing in Initial Stops: Acoustic Measurements" *Word*, 20, 384-422.
- Lombardi, L. (1991) "Laryngeal features and laryngeal neutralization," PhD dissertation, University of Massachusetts, Amherst.
- Magloire, J. and Green, K. P. (1999) "A cross-language comparison of speaking rate effects on the production of voice onset time in English and Spanish," *Phonetica*, 56, 158-185.
- Midtlyng, P. (2011) "The effects of speech rate on VOT for initial plosives and click accompaniments in Zulu," in *The 40th Annual Conference on African Linguistics*, edited by E.G. Bokamba, R.K. Shosted, and B.T. Ayalew (Cascadilla Proceedings Project, Somerville, MA), pp. 105-118.
- Miller, J., Green, K. and Reeves, A. (1986) "Speaking Rate and Segments: A Look at the Relation between Speech Production and Speech Perception for the Voicing Contrast," *Phonetica*, **43**(1-3), 106-115.
- Peirce, J. W. (2007) "PsychoPy - Psychophysics software in Python," *Journal of Neuroscience Methods*, **162**(12), 8-13.
- Pind, J. (1995) "Speaking rate, voice-onset time, and quantity: The search for higher-order invariants for two Icelandic speech cues," *Perception and Psychophysics*, 57, 291-304.
- Rubach, J. (1990) "Final devoicing and cyclic syllabification in German," *Linguistic Inquiry*, 21, 79-94.
- Schwarz, M., Sonderegger, M., and Goad, H. (2017) "Representing a four-way contrast: Nepali, voiced aspirated and laryngeal realism," in paper presented at the Annual Meeting on Phonology, New York.
- Simons, G. F. and Fennig, C. D. (eds.) (2018) *Ethnologue: Languages of the World*, Twenty-first edition (SIL International, Dallas, Texas).
- Summerfield, Q. (1981) "Articulatory rate and perceptual constancy in phonetic perception," *Journal of Experimental Psychology: Human Perception and Performance*, **7**(5), 1074-1095.
- Van Bik, K. (2003) "Junctural and Parasitic Voicing in Burmese," *Annual Meeting of the Berkeley Linguistics Society*, **29**(1), 473-484.
- Vaux, B., & Samuels, B. (2005) "Laryngeal markedness and aspiration," *Phonology*, **22**(03), 395-436.
- Wiese, R. (1996) *The phonology of German* (Clarendon Press, Oxford).