Kansas Working Papers in Linguistics, Vol. 33 (2012), 27-46

Capturing Breathy Voice: Durational Measures of Oral Stops in Marathi

Kelly Harper Berkson
University of Kansas

1 Introduction

The present study investigates a series of techniques used to capture the durational differences of oral stops in Marathi, an Indic language that exhibits a four-way phonemic distinction among oral stops. Like many of its Indic relatives, Marathi utilizes both voicing and aspiration to achieve oral stop contrasts, yielding aspirated and plain versions of both voiced and voiceless stops. While voice onset time (VOT) is an acoustic feature which can be reliably used to distinguish between plain voiced, plain voiceless, and aspirated voiceless stops, it does not help to differentiate between plain voiced and aspirated voiced—or breathy-voiced—stops. While durational differences may not be sufficient in and of themselves, they do contribute acoustic cues that aid in establishing the four-way contrast.

With this in mind, the present production study investigates the utility of two alternative acoustic measures of duration that have been proposed: Noise Offset Time (NOT), proposed by Davis (1994), is measured from the release burst to the beginning of the second formant in the following vowel. The After Closure Time (ACT), proposed by Mikuteit and Reetz (2007), is measured from the release burst to the beginning of periodicity in the waveform. Six native speakers of Marathi were recorded producing a randomized word-list that contained Marathi oral stops in word-initial position before [a]. The digitized recordings were fed into PRAAT, and waveforms and spectrograms were used to measure the ACT and the NOT of each stop.

While there is some utility to each measurement, this study suggests that neither is completely clear cut; speaker variation means that the hallmarks of the waveform or spectrogram consulted to take each measurement are sometimes absent or open to subjectivity. A new measurement that takes advantage of the consistently present elements of both the ACT and the NOT is proposed. This measure, referred to as the Pre-vocalic Interval (PVI), is measured from the release burst to the jump in intensity which coincides with the appearance of clear formants in the spectrogram of the following vowel.

2 Oral stops in Indic languages

Oral stops in Indic languages such as Bengali, Hindi, and Gujarati customarily exhibit contrasts based on both voicing and aspiration. This pattern is relatively uncommon cross-linguistically, and results in a four-way phonemic contrast consisting of plain voiceless, aspirated voiceless, plain voiced, and aspirated or breathy-voiced stops (henceforth /T, T^h, D, $D^ɦ$/ respectively).

Marathi is the state language of Maharashtra, and is spoken by approximately 72 million people (2001 India Census). In common with its Indic relatives, Marathi evidences the four-way distinction at each of four places of articulation, as seen in Table 1.

BILABIAL	p	pʰ	b	bʱ
DENTAL	t	tʰ	d	dʱ
RETROFLEX	ʈ	ʈʰ	ɖ	ɖʱ
VELAR	k	kʰ	g	gʱ

Table 1: Marathi oral stops

Sounds in the right-most column in Table 1 are alternately referred to as voiced aspirated, murmured, or breathy voiced stops (Dutta 2007, Ladefoged and Maddieson 1996, Schiefer 1987). The current paper utilizes the term *breathy voiced*. Articulatorily, these stops are characterized by vocal fold vibration during and after the stop closure; the sounds are transcribed herein with a raised /ɦ/ to indicate that the breathy release resulting from this continuous voicing differs from voiceless aspiration (Dutta 2007, Ladefoged and Maddieson 1996).

Marathi presents a unique opportunity to investigate the phonetic properties of breathy-voiced sounds because the use of breathiness as a contrastive feature in its phonetic inventory is extended beyond oral stops into additional manners of articulation. Plain and breathy voiced obstruent contrasts have been fairly well studied in languages like Hindi; far less acoustic work has been conducted with Marathi, however, which features plain and breathy voiced sonorants in addition to obstruents. Furthermore, there is no consensus within work that touches on Marathi phonetics as to the exact make-up of its consonant inventory; breathy voiced nasals, fricatives, and approximants are sometimes included (Dhongde and Wali 2007, Masica 1991) and sometimes excluded (Pandharipande 1997). Thus while the current paper focuses rather narrowly on oral stops, ultimately it fits within a larger endeavor: namely, to develop a comprehensive understanding of the acoustic correlates of Marathi consonants in order to make a significant contribution to our overall knowledge of the acoustic correlates of breathy voice.

Recognition of breathy-voiced phonemes in Indic languages dates back to ancient Sanskrit grammarians, and is uncontroversial with regards to obstruents (Lisker & Abramson 1964, Deshpande 1991, Davis 1994, Mikuteit & Reetz 2007).[1] There is ongoing debate about how the difference between stop types is realized acoustically, however. Recent work supports an analysis wherein multiple acoustic cues are utilized to differentiate between the four categories of oral stops; in his analysis of Hindi stops, for example, Dutta (2007) finds that durational and spectral measures jointly contribute to the distinction. His findings are as follows: plain stops have longer closure durations than their aspirated counterparts, and the duration of voicing during the stop closure (Voice Lead Time, or VLT) is consistently shorter for breathy voiced stops than for plain voiced stops. In addition, fundamental frequency is lower after voiced than after voiceless stops, and even lower after breathy voiced stops than after their plain voiced counterpart.

These findings support the idea that the four stop types are characterized at least in part by durational differences. The primary concern of this paper is to establish a durational measure which can be used across different stop types; while durational properties alone do not serve as a sole acoustic cue, it will become clear that they differ systematically from one type to the next, and that a measure of duration can greatly contribute to a thorough acoustic representation of oral stops in Marathi.

[1]The claim that Marathi makes phonemic distinctions based on breathiness in sonorants is less widely accepted, and will be addressed in future work; the present paper, however deals exclusively with obstruents, which are uncontroversial.

3 Durational measures of oral stops

One durational measure often referenced in discussion of oral stops is voice onset time (VOT), a measure of the temporal interval between the release of the stop closure and the onset of vocal fold vibration. In their classic paper, Lisker & Abramson (1964) demonstrate both that VOT may be used to reliably distinguish between two and three categories of oral stops and that languages differ with regards to the specific way in which this parameter is employed. If we consider the closure release to be the zero point on the temporal scale, stops which feature pre-release voicing are considered to have a negative VOT, while those which show some lag between closure release and the onset of vocal fold vibration have a positive VOT value.

Lisker & Abramson's 1964 findings—replicated by many others in the intervening years—reveal that, in the two-way English distinction, voiced stops in absolute initial position are characterized by a short positive VOT for some speakers and a negative VOT for others while voiceless stops in the same position show a long positive VOT value (p. 394). The distinction between voiced and voiceless oral stops in Dutch is realized differently, with a negative VOT for voiced stops and a short positive VOT for voiceless stops (p. 392). VOT measurements can be used to acoustically distinguish stop types in languages which contrast voiced, plain voiceless, and aspirated voiceless stops as well. In Thai, for example, voiced stops have a long negative VOT, plain voiceless stops a short positive VOT, and aspirated voiceless stops a long positive VOT (p. 396).

The usefulness of VOT for differentiating between all stop types falls short in a language with a four-way contrast, however; because both the plain voiced (/D/) and breathy-voiced (/Dʱ/) stops include vocal fold vibration during the stop closure, both have long negative VOTs. Average VOT data for Hindi and Marathi velar stops appear in Table 2.

	/k/	/kʰ/	/g/	/gʱ/
Hindi	18	92	-63	-75
Marathi	24	87	-116	-89

Table 2: Average VOT of velar stops in Hindi and Marathi (in ms)
(data from Lisker & Abramson 1964: 398[2])

At first glance, the VOTs of the plain and breathy voiced stops appear to differ. The problem noted by Lisker & Abramson, however, is the significant overlap in their VOT ranges; Hindi /g/ values range from -95 to -30 ms, for example, while /gʱ/ values range from -160 to -40 ms. Virtually the entire /g/ range, then, falls inside the /gʱ/ range. This extensive overlap renders VOT uninformative in distinguishing/D/ from /Dʱ/.

Several alternative durational measures have been proposed, including "Noise Offset Time" (Davis 1994). Noise Offset Time (or NOT) is measured from the burst associated with the release of the stop closure to the beginning of a clear second formant in the following vowel as illustrated in Figure 1.This image represents the initial portion of the Marathi word [kʰas], *'special,'* as produced by one of the speakers who participated in the present study.

[2]Lisker & Abramson recorded a total of seventeen speakers producing samples from eleven distinct languages; the number of speakers recorded for each language is not reported, however.

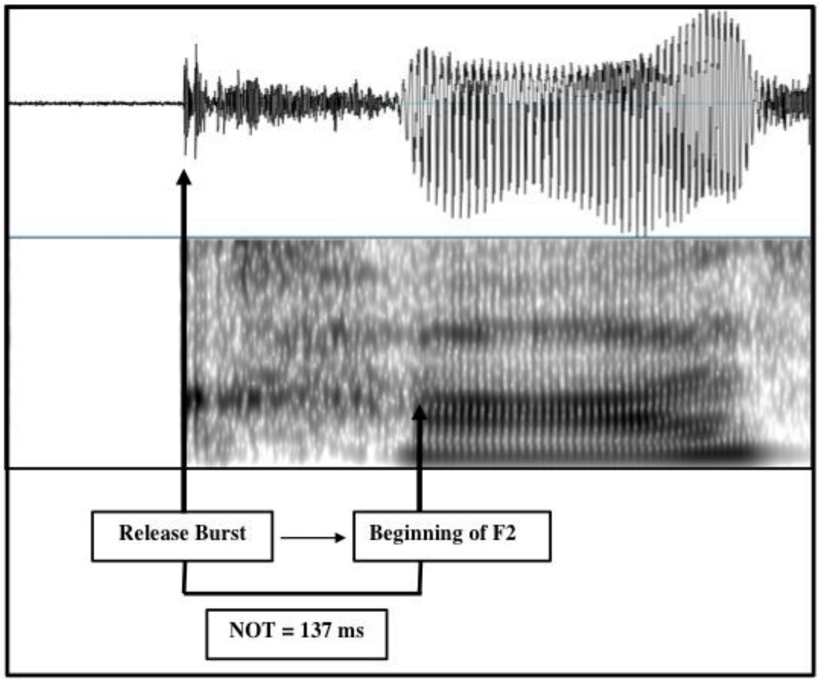

Figure 1: NOT measurement of [k^has], *'special'*

Davis's average NOT values for Hindi velar stops appear in Table3. Pairwise comparisons between each set of phonemes (i.e. /k/ and /k^h/, /k/ and /g/, et cetera) confirm the trend observed numerically: each NOT average differs significantly from the others.

	/k/	/k^h/	/g/	/$g^ɦ$/
NOT	26	74	12	34

Table 3: Average NOT of Velar Stops in Hindi (in ms) (*data from Davis 1994: 184)*

One criticism leveled against the NOT is that the visibility of F2 may vary based on factors such as the intensity the speech signal itself and on the specific settings of the program being used to generate and view a spectrogram of the sound (Mikuteit & Reetz 2007: 253). The potential lack of replicability yielded by this factor is obviated in After Closure Time (ACT), another alternative durational measurement proposed by Mikuteit & Reetz (2007). The authors observe, accurately, that the post-release interval can be divided into two distinct portions. The first of these, the ACT, consists of the somewhat random buzzy noise associated with the closure release and is found in all four stop types. The second portion—Superimposed Aspiration (SA)—retains some jaggedness, but this jaggedness is laid over the clear periodicity associated with the vocal fold vibration of the following vowel. This may be likened to a breathy vowel, and appears in aspirated voiceless and breathy voiced stops.

Consider that the ACT is measured from the closure release to the beginning of clear periodicity; as such, we may think of it as a measure that disregards closure voicing and captures only positive VOT values. Mikuteit & Reetz examined the oral stops at four places of articulation in East Bengali. Average ACT values (reported as least square means) appear in Table 4.

	BILABIAL	DENTAL	RETROFLEX	VELAR
Voiceless Plain	14	15	11	34
Aspirated Voiceless	--	57	48	73
Voiced Plain	8	13	7	23
Breathy Voiced	31	35	36	44

Table 4: Average ACT values of East Bengali stops (least square means, in ms) (*data from Mikuteit & Reetz 2007: 184)*

The above data seem to pattern with the NOT data presented in Figure 4; if we consider the velars, the numerical values differ but the trend remains the same. For both the ACT in East Bengali velars and the NOT in Hindi velars, plain voiced stops have the shortest values, followed by the plain voiceless stops; breathy voiced stops have the second to longest values, and aspirated voiceless stops have the longest values (i.e. /g/ < /k/< /gʱ/ < /kʰ /). It is important to note, however, that statistically significant differences between means for each pair of phonemes are not achieved in all four places of articulation.

In short, ACT values reliably distinguish the plain from the aspirated/breathy voiced phonemes, but—outside the velar category—they are not sufficient on their own to reflect a four-way contrast. ACT fails to distinguish the plain voiced from the plain voiceless stops. This failure is not fatal in and of itself, given that we are not looking for a single acoustic parameter which can differentiate the stop types; in terms of differentiation, the plain voiced and voiceless stops are clearly distinguished based on the presence (or absence) of vocal fold vibration.

In terms of a comparison, ACT and NOT measurements are similar in that both look at the post-release/pre-vocalic section of the segment. They differ, however, in that the NOT is composed of both the random noise associated with the release burst and a portion of the pre-vocalic glottal buzz which may or may not involve some periodicity, while the ACT is composed of only the post-release random noise, and ends the instant any periodicity is visible in the waveform. Periodicity accompanied by “glottal buzz” (à la Lisker & Abramson) in the breathy-voiced stops belongs to the superimposed aspiration portion of the interval, with the periodicity presumed to belong to the following vowel. Mikuteit and Reetz posit that the NOT includes both what they measure as the ACT and some portion of what they designate superimposed aspiration, and when they replicate NOT measurements with their own data they find that this is indeed the case. NOT generally corresponds with ACT measurements for plain stops, and includes some milliseconds of SA for the aspirated and breathy voiced stops.

The experiment reported in the following section assesses the utility of both ACT and NOT measurements in capturing the durational correlates of oral stops in Marathi. As will be seen, both measurement techniques have some strengths and some weaknesses. Division of the post-release interval into separate ACT-like and SA-like components is clearly possible in many cases. The exact division point, however, is often less than clear and sometimes seems to be lacking altogether. Emergence of a clear F2, on the other hand—the hallmark endpoint of the NOT—is consistently present; as mentioned, however, it may vary slightly based on properties related to the intensity of the speech sample as well as program settings dealing with contrast and resolution. Beyond this drawback, a clear but pale F2 band is often visible long before the darkening associated with the cessation of breathiness (and corresponding rise in intensity) in the subsequent vowel. These and other issues will be discussed in more detail following presentation of the results from the current experiment, and a comprehensive solution will be proposed.

4 ACT and NOT measurements in Marathi

To assess the utility of ACT and NOT measurements in capturing durational differences in Marathi oral stop lag times—i.e., the equivalent of positive VOT—six speakers were recorded producing tokens in isolation. ACT and NOT values were then measured.

4.1 Predictions

Marathi stops are expected to be in keeping with both Davis's Hindi findings and Mikuteit and Reetz's Bengali findings. The predicted pattern for Marathi ACT and NOT values is as follows: /T^h/ > /D^ɦ/ > /T/ > /D/. Aspirated/breathy voiced sounds will show longer average values; plain sounds will show shorter average values. Aspirated voiceless stops should have the longest average NOT and ACT durations; plain voiced stops should have the shortest.

4.2 Methods

4.2.1 Subjects

Stimuli were produced by six native speakers of Marathi, two male and four female. With the exception of one female subject, who is older, all speakers are in their twenties and thirties, are currently living in the United States (in Kansas), and are either currently attending English-medium schools for advanced degrees or have graduated from English-medium schools with advanced degrees. All participants are trilingual, and speak Marathi, Hindi, and English fluently and on a daily basis. All grew up in India, in various parts of Maharashtra, and attended schools that used the three above-mentioned languages in various contexts.

4.2.2 Stimuli

The stimuli consisted of Marathi words with each of the 16 oral stops appearing in word-initial position before the vowel /a/ (see Appendix A).Four of the speakers (2 male, 2 female) recorded the full list in (8); the remaining two recorded only the labials. Results presented in this paper focus primarily on the bilabial stops, but occasional reference will be made to the other places of articulation. When this is done, please note that findings are for four speakers only. Three repetitions of each token were recorded and analyzed; stimuli were randomized, and were presented visually in Devanagari script. Participants were asked to read the words at a normal speaking rate in a normal tone of voice.

4.2.3 Recording

Speakers were recorded onto a flash disc using an Electro-Voice RE20 cardioid microphone and a Marantz Portable Solid State Recorder (PMD 671) at a sampling rate of 22 kHz in the anechoic chamber at the University of Kansas, Lawrence.

4.2.4 Acoustic analysis

The digitized recordings were imported into PRAAT, freeware used for acoustic analysis which allows simultaneous examination of the waveform and (wide-band) spectrogram of each token. As previously noted, acoustic measurements taken included both NOT and ACT. All measurements were taken from the nearest relevant zero-crossing on the waveform.

NOT duration was measured from the release burst to the beginning of a clear F2, as shown in Figure 1. ACT duration was measured from the release burst to the upward-going zero crossing immediately preceding the onset of periodicity, as shown in Figure 2.

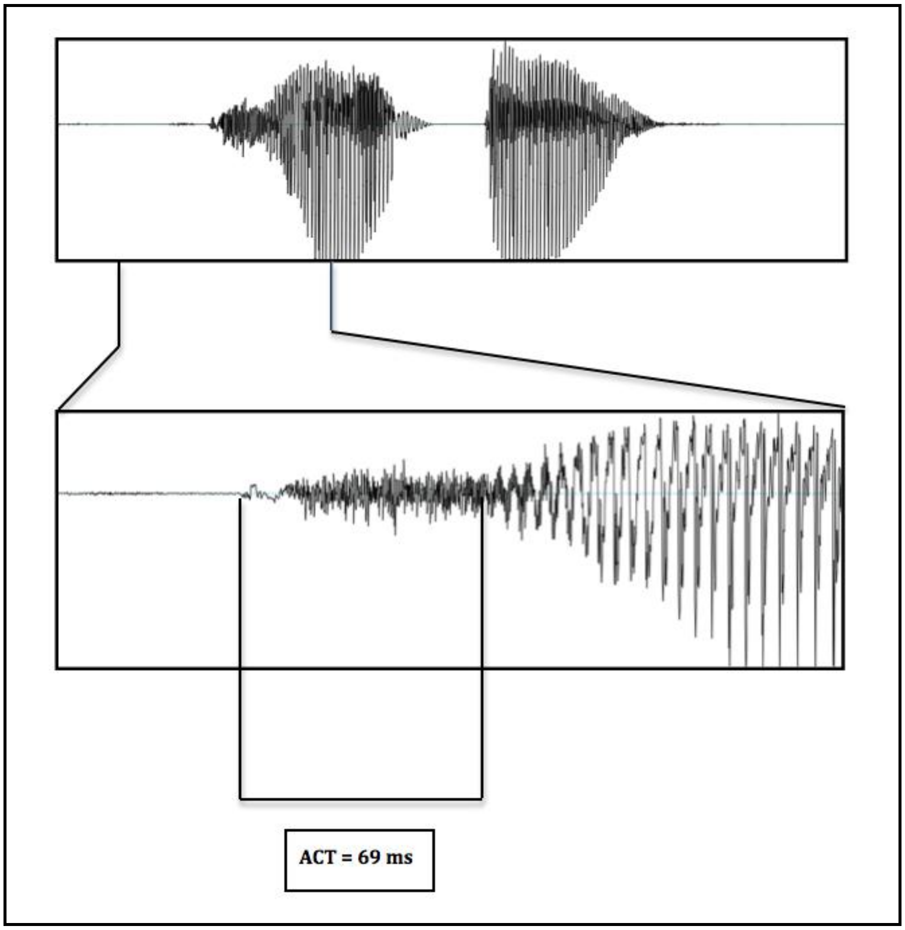

Figure 2: ACT measurement of [pʰaʈa], *'fork (in the road)'*

5 Results

NOT results are reported in section 5.1; ACT results follow in 5.2. One comment must be made at the outset, however. Although Marathi is said to contain all four bilabial stops (Dhongde & Wali 2007, Pandharipande 1997), five of the six speakers recorded herein produced [f] instead of [pʰ] 100% of the time. This cannot be explained by English spelling, as the stimuli were presented in Devanagari, in which each of the four bilabial stops is represented by its own character as shown in Table 5.

प	/pə/	फ	/p^{h}ə/	ब	/bə/	भ	/b^{ɦ}ə/

Table 5: Bilabial Stops in Devanagari[3]

Regardless, the symbol for [p^{h}] was systematically pronounced [f]. Anecdotally, this phenomenon seems to be occurring with increasing regularity for an increasing number of speakers. The speaker who did pronounce the aspirated voiceless bilabial stop did so for only one word, although the sound appeared in several stimuli. She pronounced [p^{h}aʈa], *'fork (in the road),'* each time it appeared in the stimulus list. The other speakers pronounced [faʈa].

Investigation of potential causes will be saved for future work, but for now we must note that this distinction is being lost and results must be read accordingly. While [p, b, b^{ɦ}] results are from all six speakers, [p^{h}] results are from the one speaker who produced the sound preceding an [a] vowel. Thus they give us a reference point for the voiceless aspirated stop, but cannot be interpreted as being on equal footing with the grouped data.

5.1 NOT

NOT measurements included both the post-burst random noise of the ACT measurement plus the portion termed "superimposed aspiration" by Mikuteit and Reetz. Average NOT durations for each of the four bilabial stops—measured in ms—appear in Figure 3.

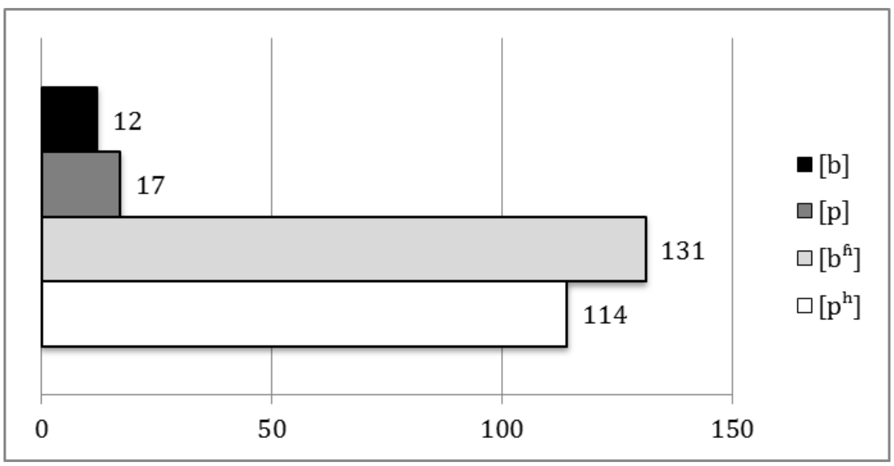

Figure 3: NOT averages (in ms) – bilabials

A one-way analysis of variance was conducted to determine whether NOT values differed significantly across stop types. The test was significant overall (F (3, 127) = 1009.82, p < 0.0001). Further, it revealed that the mean NOT of aspirated voiceless stops (M = 114 ms) differed significantly from that of the breathy voiced stops (M = 131 ms), plain voiceless stops (M = 17 ms), and plain voiced stops (M = 12 ms) with p < 0.01. The difference between the plain voiceless and breathy voiced means was significant at the p <0.01 level as well, but the difference between plain voiced and plain voiceless means was not significant.

[3]Note that characters in Devanagari are considered to include an inherent schwa, and are pronounced as such unless overtly marked either with another vowel character or with a symbol (called a *halant)* which indicates that the schwa should not be pronounced.

These results are noteworthy for two interconnected reasons. First, they diverge from the predicted trend in that the breathy voiced stop exhibits the longest NOT. To determine whether this is an anomaly somehow connected to the fact that only one speaker produced the voiceless aspirated stop, NOT values for all stop types across all places of articulation were calculated for the four speakers who produced the full set of data. These data—reported in Figure 4—indicate that the pattern observed in the bilabial stops is mirrored by the stops at the other places of articulation.

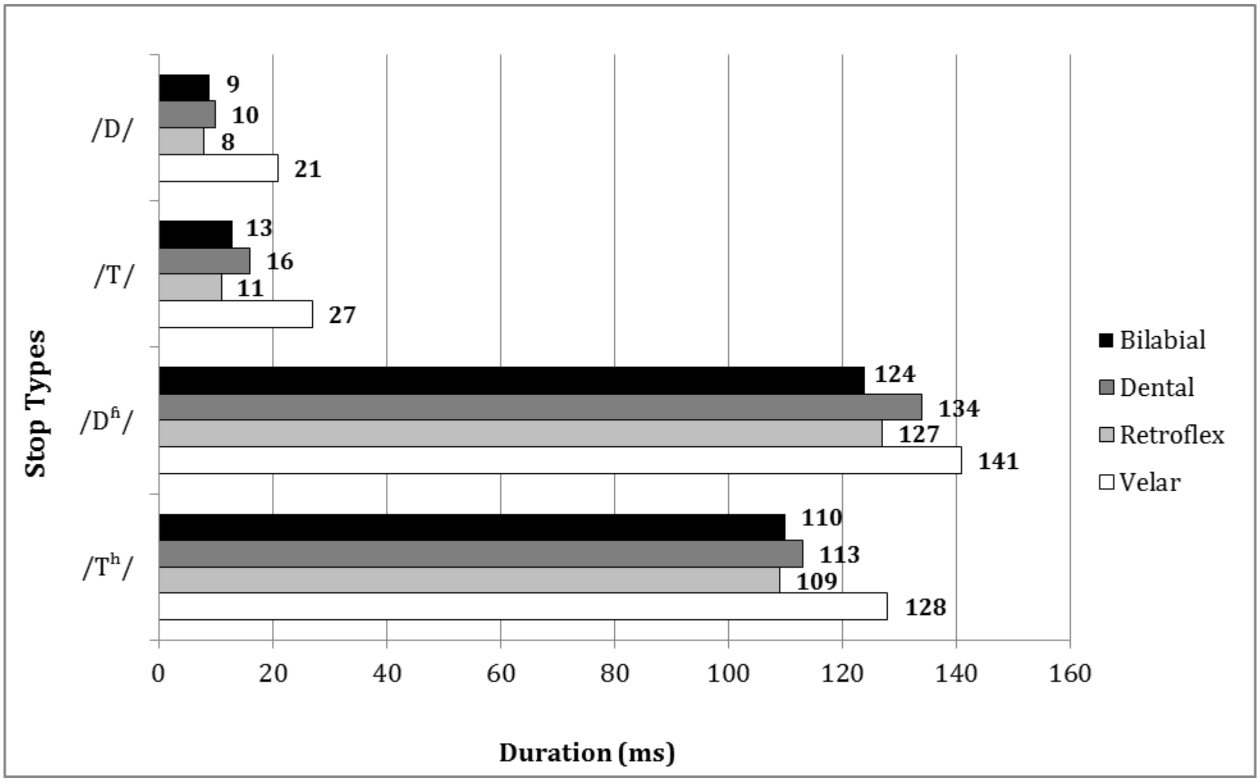

Figure 4: NOT averages by stop type and place of articulation (in ms)

This leads us to the second issue. Recall that Davis's reported NOT average for /gʱ/ was 34 ms. This is wildly different than the breathy voiced NOT values measured in the present study. The divergence in values can be explained as follows. In a spectrogram such as that shown in Figure 5, a pale F2 band appears long before the darkening of F2 associated with the jump in amplitude visible in the waveform. If NOT is assumed to end at the appearance of the pale F2 band, then the breathy voiced values will be more akin to those reported by Davis. Because this pale F2 band is only sometimes present, the point where F2 darkened was considered to be the NOT endpoint in my measurements.

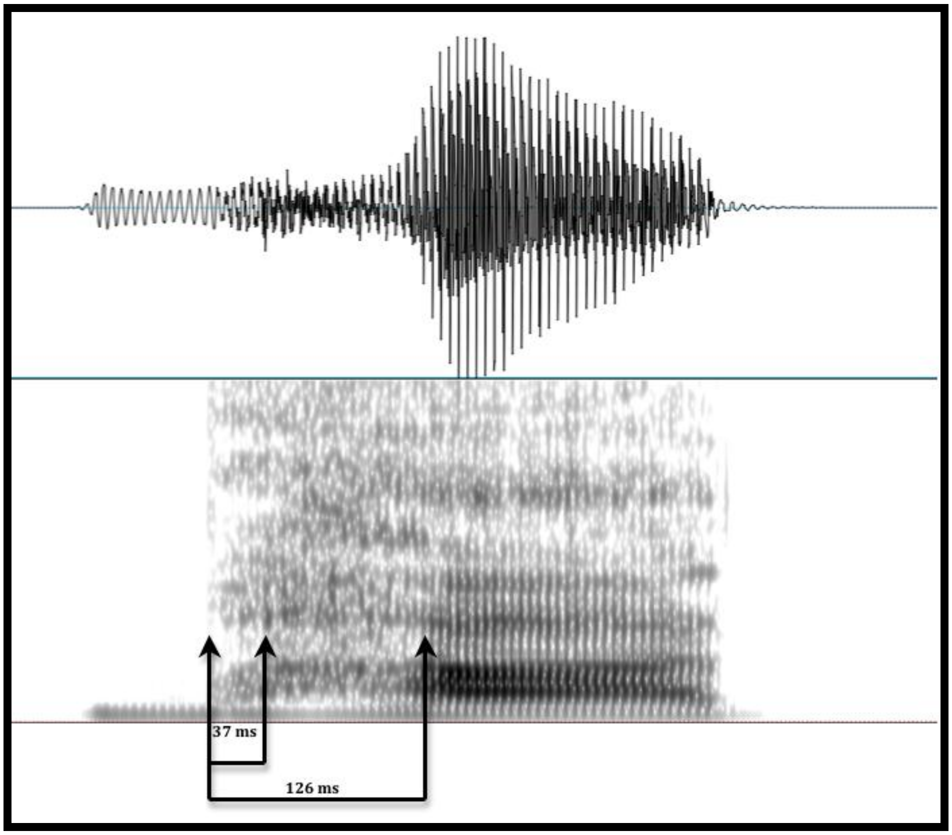

Figure 5: Varying NOT endpoints

Further discussion of this issue appears in section 6, but for now let us note that the divergence illustrated in Figure 5 is not to be lightly dismissed. In fact, it highlights the major drawback of NOT. A great advantage of NOT is that a clear F2 always becomes visible at some point, regardless of speaker and token variation; this cannot be said of the ACT, as will be seen in the coming section. The trouble with NOT is that the point at which F2 becomes clear is truly open to deliberation, particularly when relying solely on a spectrogram. That there is some degree of subjectivity involved in pinpointing the end of the NOT detracts from its value as a replicable measure. We will return to this point shortly.

5.2 ACT

Average ACT measurements for Marathi bilabial stops are found in Figure 6. Here, all sounds patterned as predicted, with the plain stops showing shorter average durations than their aspirated/breathy counterparts, and voiced sounds showing shorter average durations than their voiceless counterparts.

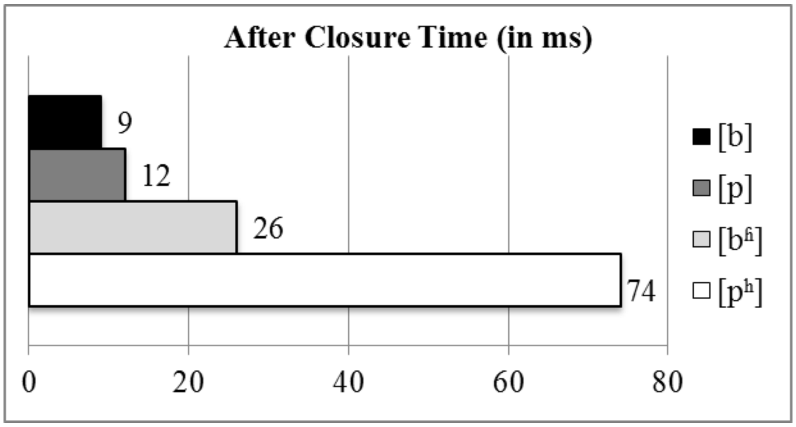

Figure 6: ACT Averages (in ms)

The issue that arose in the NOT data—namely, that the current study found much longer NOT duration averages for the breathy voiced stops than those reported by Davis (1994)—was not mirrored in the ACT data. The numbers reported herein are in line with those reported by Mikuteit and Reetz (2007). This is perhaps best explained by reference to the technical details involved in the measuring techniques. Recall that after closure time measures the interval stretching from the release burst to the onset of periodicity associated with the following vowel. As such, one of the ways in which ACT measurements differ from NOT procedurally is that they rely primarily on the waveform rather than on the spectrogram, thereby avoiding the subjectivity associated with pinpointing the appearance of a clear F2.

For interest's sake, we may consider the average ACT durations of each of the 4 stop types across all places of articulation—again, recall that the data in Figure 7 are from four speakers only, not the six whose data are reported in Figure 6.

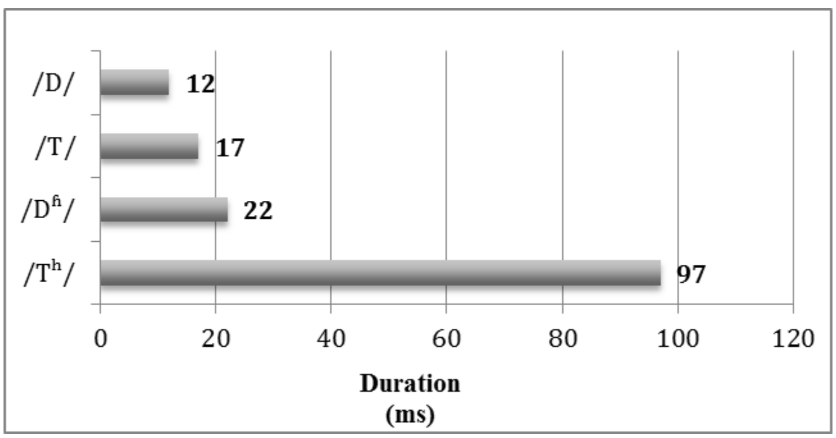

Figure 7: ACT averages by stop type (in ms)

As predicted, aspirated voiceless stops exhibit the longest average ACT duration, while plain voiced stops exhibit the shortest. A one-way analysis of variance was conducted to determine whether mean ACT values differed significantly across stop types. The test was significant overall (F (3, 127) = 883, p < 0.001), and revealed that the mean ACT of aspirated voiceless stops (M = 97 ms) differed significantly from that of the breathy voiced stops (M = 22 ms), plain voiceless stops (M = 17 ms), and plain voiced stops (M = 12 ms) with $p < 0.01$. The difference between the plain voiceless and breathy voiced means was significant at the $p < 0.05$ level. The difference in mean ACT values between plain voiced and plain voiceless sounds was not significant, however.

We can also separate these data according to place of articulation, as shown in Figure 8. While the present study has not focused on differences across places of articulation, there are several points of interest. Velar stops unfailingly exhibit ACTs of the longest average duration, but bilabial stops do not exhibit the shortest average ACTs as we might expect. Rather, retroflex stops have the shortest ACTs, albeit by only one or two milliseconds in most cases. This is in line with the findings of Mikuteit and Reetz, whose Bengali data showed that while the shorter duration of all things related to bilabial stops (closure duration, VOT, etc.) is often thought to be a crosslinguistic universal, durations associated with retroflex stops seem to be shorter. This point bears deeper investigation in the future.

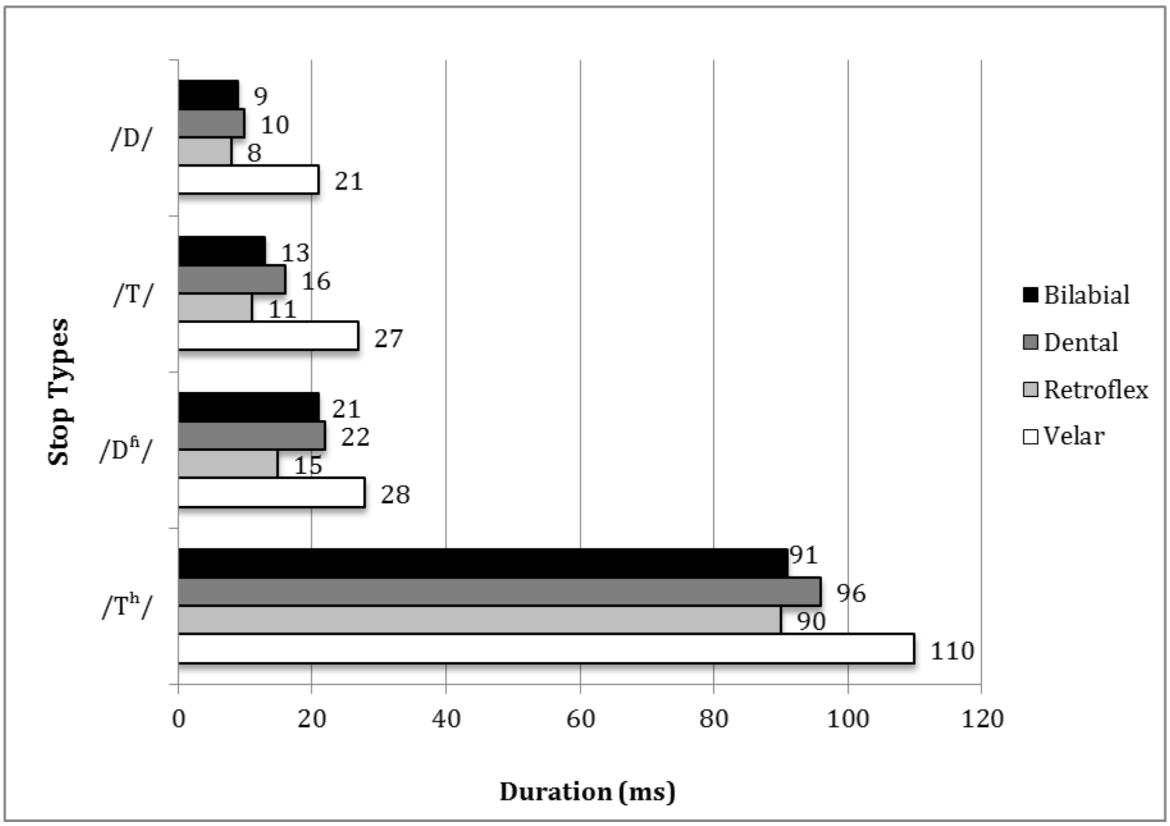

Figure 8:ACT averages by stop type and place of articulation (in ms)

Recall that Mikuteit and Reetz took superimposed aspiration (SA) measurements in addition to measuring ACT. The trend they observed, which held across all places of articulation, was that SA was about twice as long for breathy voiced than for voiceless aspirated stops, with means of 47 ms and 24 ms, respectively (Mikuteit and Reetz 2007: 264). As illustrated in Figure 9, the general trend held for the Marathi data analyzed herein—that is to say, breathy voiced sounds had significantly more SA than voiceless aspirated sounds.

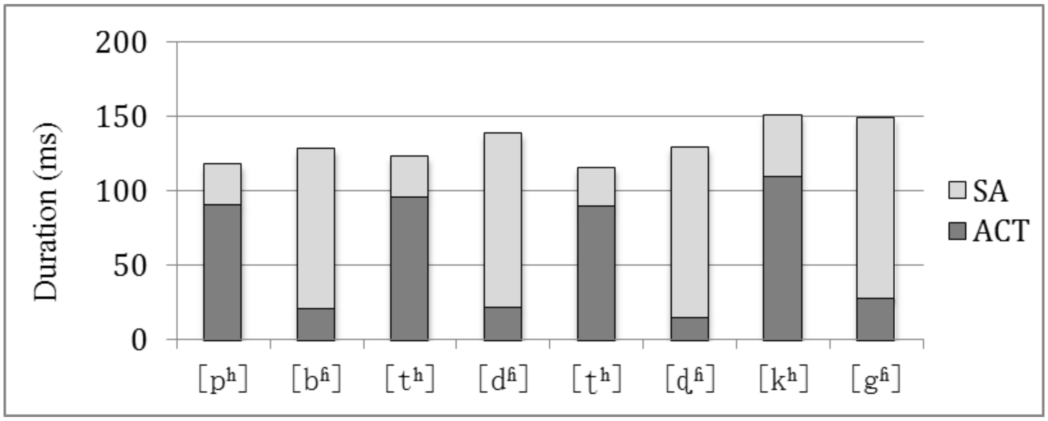

Figure 9: Average ACT and SA Durations (in ms)

Note that the difference between mean SA values for voiced and voiceless sounds is greater than that reported in Mikuteit and Reetz—it is on the order of 4:1, while the differences reported in Mikuteit and Reetz (2007) were more on the order of 2:1. One potential explanation for this divergence could be that aspiration is realized differently in Marathi than in East Bengali, perseverating farther into the vowel. For the time being, however, this remains an open question.

6 A third proposal: pre-vocalic interval (PVI)

The results presented in Section 5 reported on the use of both ACT and NOT in distinguishing between stop types in Marathi, a language which employs both voicing and aspiration/breathiness distinctions in its phonemic inventory. Based on the preliminary findings of this pilot study, both measurements have advantages; similarly, both have drawbacks. We will now investigate these in more detail, and propose a third measurement technique which draws on the strengths of both after closure time and noise offset time while avoiding the associated pitfalls.

As noted, the advantage of noise offset time is that it makes use of a landmark which is unfailingly present in the spectrogram: namely, F2. The drawback, however, is that pinpointing the exact onset of a clear F2 may be open to interpretation. The advantage of measuring after closure time is that, procedurally, it relies on landmarks found in the waveform.

I propose that with some slight modifications to measurement techniques, the so-called landmarks in the waveform are both reliable and replicable. Modifications are necessary because of something regularly encountered in the present analysis of Marathi data. Namely, the period referred to as ACT and characterized by random noise and a lack of periodicity in the waveform was very often absent. This is illustrated in Figure 10.

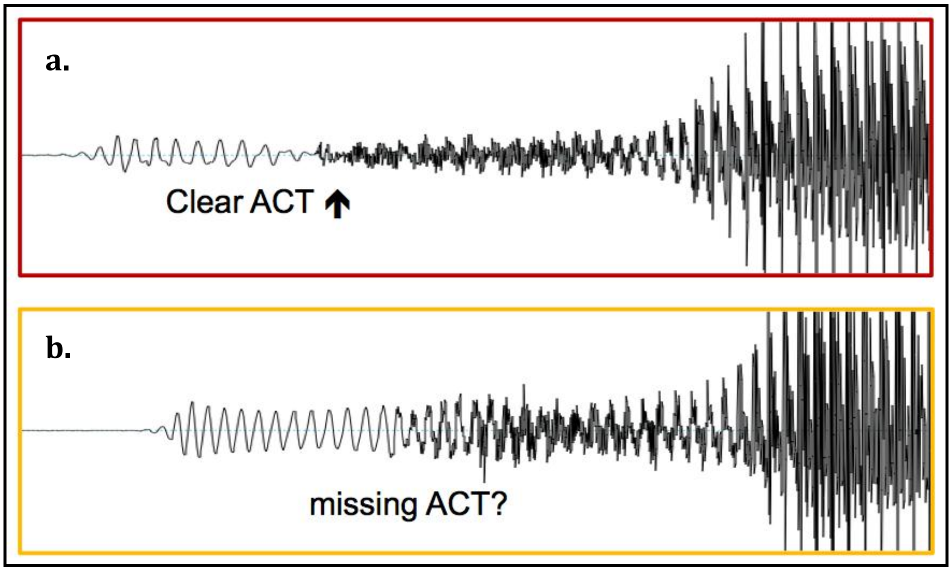

Figure 10: Two waveforms of the [bʱ] in [bʱat], *'rice'*

In Figure 10a., we see a clear—if brief—cessation in periodicity, which we can confidently segment as the ACT. In 10b., however, there is no such interval; rather, the smooth periodicity of the waveform representing the prevoicing of the stop gradually becomes slightly jagged; this is followed by a section of jagged periodicity which transitions into the regular periodicity of the vowel after a lag of 96 ms.

This jagged periodicity is what Mikuteit and Reetz referred to as superimposed aspiration, and there is no difficulty in describing what we see: the jagged periodicity results from the breathiness of the release extending into the following vowel. Dutta (2007) reports that, in Hindi stops, the breathiness extends through approximately the first 30% of the subsequent vowel. The difficulty comes in attempting to divide this interval into the two portions defined by Mikuteit and Reetz, for in the tokens which show no clear ACT there is no abatement in voicing—and therefore, no cessation in periodicity—once the word has begun. The initial stop is voiced, the subsequent vowel is voiced, and there is voicing throughout the release closure. Thus while the observation that the post-release interval of aspirated/breathy voiced stops is composed of two portions is clearly accurate, division of the interval into the portions defined as ACT and SA is not clear-cut.

Before determining how to address this difficulty, let us consider sample waveforms from each of the four stop types in order to get a clear understanding of the interval we are interested in measuring. Images appear in Figure 11; commentary on each follows.

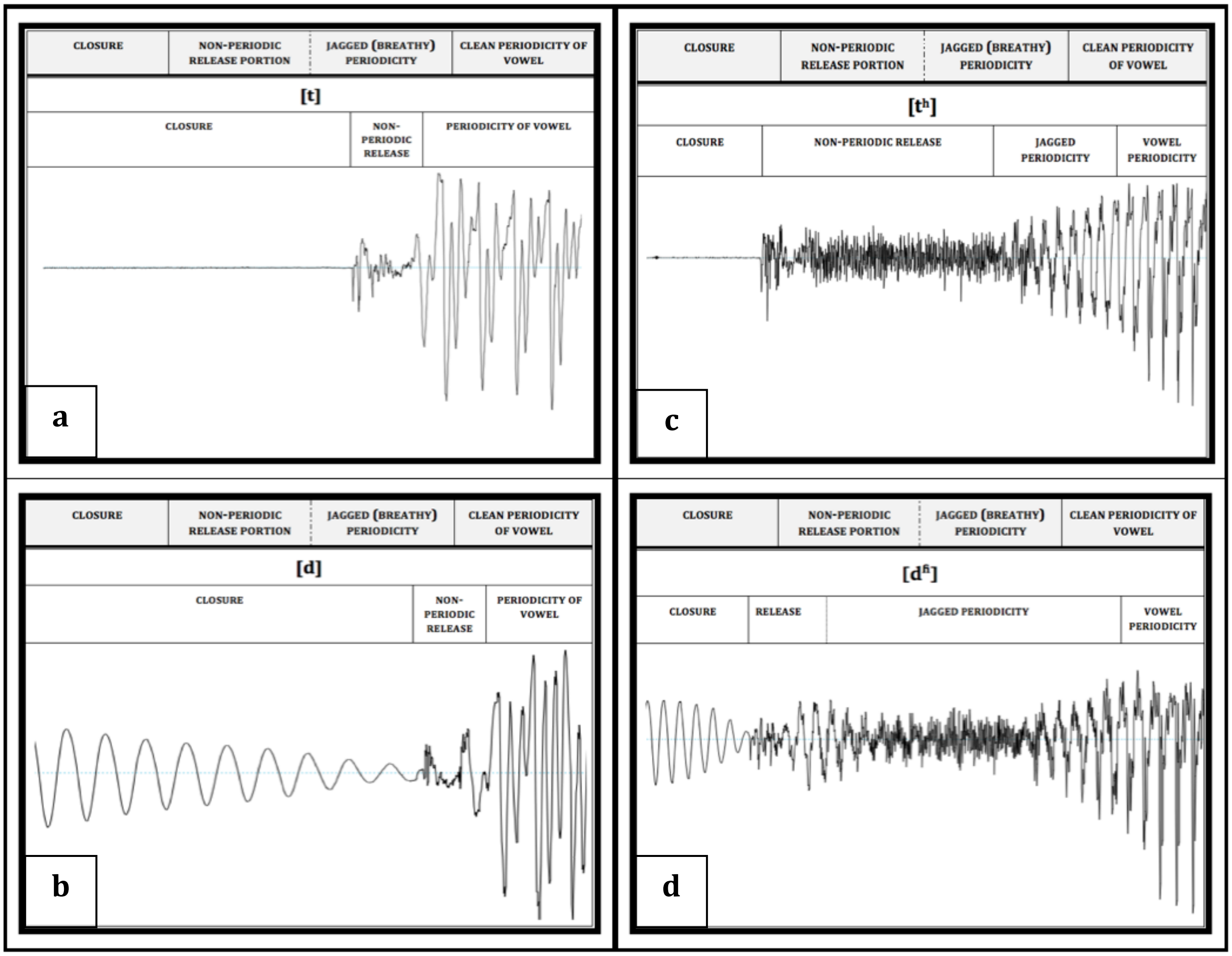

Figure 11: The interval of interest: sample waveforms of [t] (a), [d] (b), [tʰ] (c), [dʱ] (d).

Oral stops may contain four possible intervals, as is indicated at the top of each illustration in Figure 11: the stop closure, a non-periodic release, jagged periodicity associated with breathiness, and the clear periodicity of the subsequent vowel. The stop closure contains the clean periodicity of stop voicing for voiced stops. Beginning at the release burst is the interval which is of primary interest to us; this may be composed of both a nonperiodic portion (Mikuteit & Reetz's ACT) and a jaggedly periodic portion (Mikuteit & Reetz's SA). The division between these two portions may or may not be clear; thus they are divided by a dotted rather than a solid line. Finally, the cessation of breathiness in the vowel yields clean periodicity.

Both the plain voiceless stop illustrated in 11a. and the plain voiced stop in 11b. demonstrate clear non-periodic release portions, and an obvious transition into the periodicity of the vowel. Neither breathiness nor aspiration are present to induce the jaggedness associated with buzzy or aspirated noise, and the transition into the vowel periodicity is marked by a jump in amplitude. Although the spectrogram is not included in this image, a clear darkening of the vowel formants in the spectrogram consistently coincides with the point in the waveform that shows a jump in amplitude.

The voiceless aspirated stop illustrated in 11c. is characterized by a clear release burst, an interval of random, aperiodic noise, and a short interval of jagged, low-amplitude periodicity.

Jaggedness does not persist for the entire low-amplitude interval, and the jump in amplitude again coincides with the appearance of clear vowel formants in the spectrogram.

The breathy voiced stop is illustrated in 11d. This image is characterized by some of the difficulties previously mentioned. While there is clear stop closure voicing and an obvious release burst, no clear divisions are obvious in the interval of periodicity between the release and the rise in amplitude associated with the vowel. This interval is composed of distinct parts; it's simply not clear that they can be divided into ACT and SA as proposed by Mikuteit and Reetz. This token of [dʱ] shows a less jagged periodic interval—here identified as the release portion—followed by a low-amplitude interval identified as jagged periodicity. This interval contains clearly jagged periodicity which first transitions into something slightly less periodic but with no greater amplitude and then becomes more periodic before exhibiting the jump in amplitude which is familiar from the other three waveforms. Again, this jump coincides with the darkening of vowel formants in the spectrogram.

At issue is the fact that the acoustic realization of the breathy voiced stops varies from token to token; we have seen waveforms of three breathy-voiced tokens, two in Figure 10 and one in Figure 11, all of which differ from one another. This trend of variation holds from speaker to speaker, and within a single speaker's tokens. It renders application of ACT as a durational measurement difficult if not impossible, for many tokens simply fail to exhibit a clear ACT. The challenge, then, seems to be to determine which portions of the acoustic signal remain consistent and measurable across all stop types and tokens.

As a final illustration of the complications inherent in this task, consider the intervals demarcated in the spectrogram and waveform found in Figure 12.

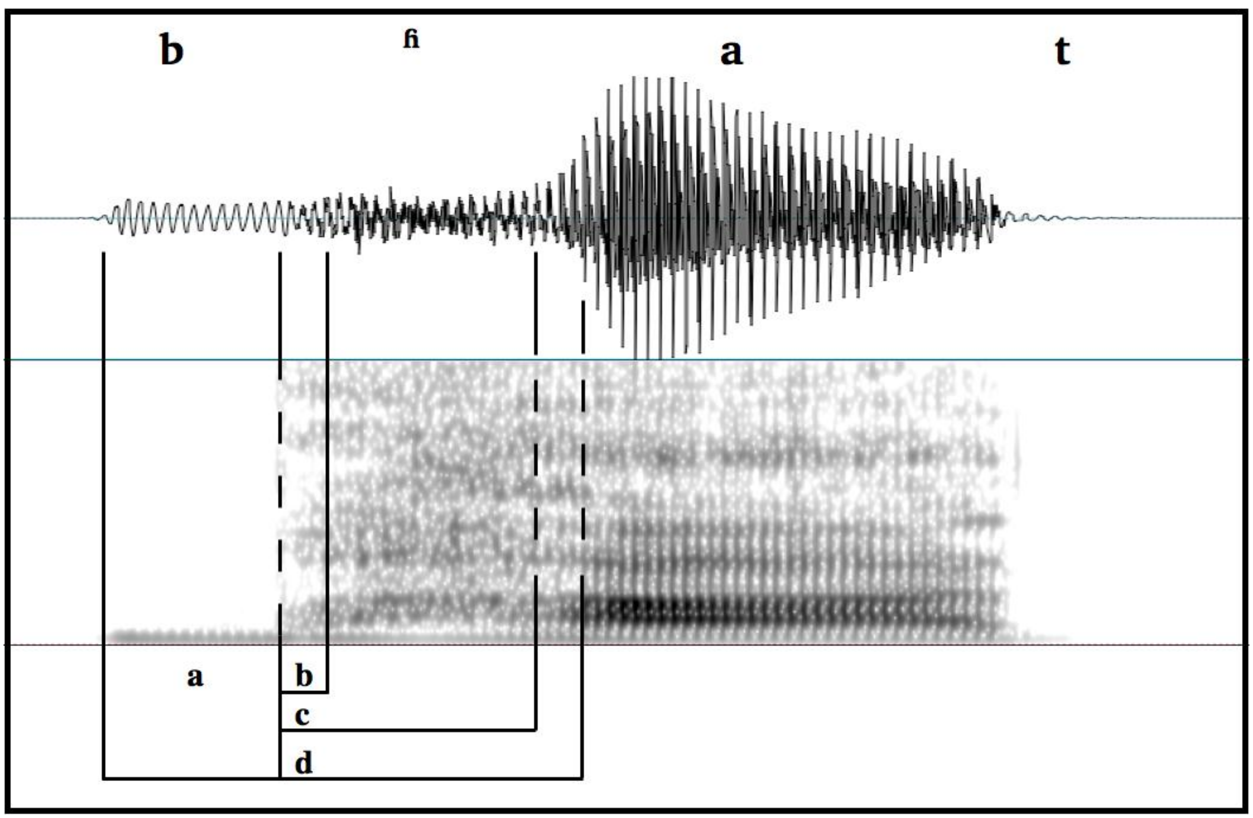

Figure 12: Waveform and spectrogram of [bʱat], *'rice'*. (See text for details.)

Demarcated in 12a is the stop closure. As shown in 12b, which begins with the release burst, there is no measureable ACT in this token because there is no aperiodicity; rather, the wave moves from the smooth periodicity of the stop closure directly into the jagged periodicity of the breathy portion of the sound. 12c, which measures the NOT, ends at the onset of a visible F2 in the spectrogram; as previously noted, however, the start of a visible F2 is mildly subjective at best and wildly subjective at worst. Finally, the interval labeled 12d begins with the release burst and ends with the beginning of a consistent rise in amplitude in the waveform.

The images presented thus far are representative of the data observed in this study, and they indicate that reliably identifiable landmarks exist in the signal regardless of inter- and intra-speaker variability. The post-release interval clearly begins at the release burst, which is generally visible in the waveform; in rare instances when it is not, however, it can be identified in the spectrogram. Further, evident in all stop types and in all tokens reviewed thus far is a jump in the amplitude of vowel periodicity in the waveform. This jump in amplitude often coincides with the darkening of vowel formants in the spectrogram, as seen in Figure 12.

I propose utilizing these consistent hallmarks to measure what I refer to as the Pre-Vocalic Interval (PVI). PVI measurements for aspirated and breathy voiced stops include not only the closure release duration, but also that portion of the vowel which is heavily flavored by the breathy release—this is because the breathiness perseverates into some portion of the vowel, causing the aforementioned jump in amplitude to occur sometime after vowel onset. (If Marathi exhibits a pattern similar to the Hindi pattern reported in Dutta (2007), breathiness may persist for approximately 30% of the subsequent vowel.)

PVI measurements were taken for a subset of the tokens produced for this study—namely, tokens that included bilabial stops in word-initial position before [a] produced by the four female speakers recorded. Mean PVI values for each of the four bilabials appear in Figure 13.

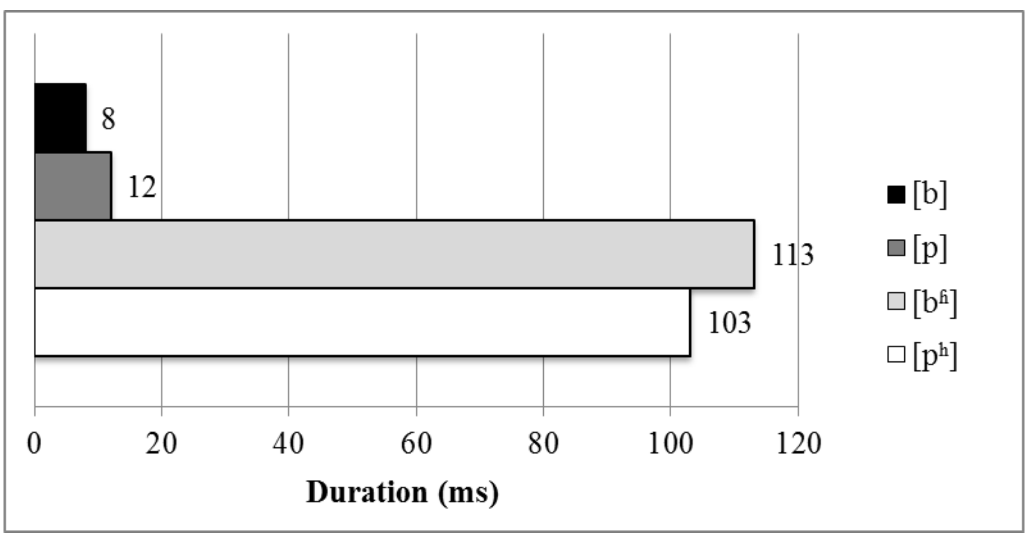

Figure 13: Mean PVI values of bilabial stops in Marathi

Again, recall that only one subject produced the voiceless aspirated stop, and so the results must truly be taken as preliminary particularly with regards to [p^h].

A one-way analysis of variance was used to evaluate whether the mean PVI values differed significantly across stop types; assumptions of normalcy were met despite the small sample size.

For the one speaker who produced all four sounds, the overall test was significant (F (3, 15) = 287, $p < 0.0001$). The mean pre-vocalic interval of the aspirated voiceless stop (M = 103 ms) and of the breathy voiced stop (M = 125 ms) differed significantly from one another and from the means of both the plain voiced stop (M = 6 ms) and the plain voiceless stop (10 ms) with $p <$ 0.01. The difference between the plain voiced and plain voiceless stop means was not significant, however.

An ANOVA comparing the [p, b, b^{ɦ}] data—the data produced by all four speakers—was significant overall, with F (2, 41) = 470 and $p < 0.0001$. As expected, both plain stops differed significantly from the breathy voiced stop; as in all previous analyses reported in this paper, however, they were not significantly different from one another.

7 Discussion

Based on these preliminary results, the PVI is a promising alternative to VOT, ACT, and NOT measurements. While it is true that PVI alone cannot distinguish the plain voiced from the plain voiceless stops, this was the case for both the ACT and the NOT as well, and future work with a wider scope of inquiry can and will utilize alternative acoustic measurements such as closure duration to capture this particular distinction. Thus with regards to distinguishing all four stop types, PVI has no clear advantage over ACT or NOT. As a temporal measure used to capture the post-release interval, however, it has clear advantages over the other measures, not least of all because it is replicable. This was true for neither the ACT nor the NOT.

The ACT relies on landmarks in the waveform which are not consistently present in the data analyzed herein, rendering it difficult or impossible to measure on many tokens. Further, ACT measurements are relatively long for voiceless aspirated sounds, and much shorter for the other three sounds. Average NOT durations are fairly long for aspirated voiceless and breathy-voiced stops, while plain stops of both types exhibit shorter NOT durations. This grouping more closely mirrors the perceptual experience listeners have when hearing these types of stops; the breathy-voiced part of breathy-voiced stops is perceptually salient, and clearly distinguishable from the following vowel. While perceptual data cannot help us untangle the acoustic puzzle, we can look to it for clues as to whether we are moving in the correct direction. The NOT endpoint coincides with the appearance of a clear F2 in the spectrogram, however, and pinpointing this endpoint turns out to be highly subjective. Thus measuring both the ACT and the NOT is, in practice, problematic.

In contrast, the PVI is taken using landmarks which are consistently present in the acoustic signal, despite the amount of within- and between-speaker variability found in the tokens analyzed herein. This consistency renders the PVI highly replicable. Future studies may reveal that the variability observed in the Marathi data is not present in all languages; given Mikuteit and Reetz (2007), for instance, we can presume that the East Bengali data yielded reliable ACTs and less variability. Even in a language with less variability, however, the landmarks utilized in measuring PVI will be clearly visible, rendering it applicable across the board: it can be reliably measured in languages with a great deal of variability, as well as in languages with more consistency. This leads back to the point mentioned previously: by relying on characteristics of the waveform which are consistently present, the PVI avoids the subjectivity inherent in looking for landmarks in the spectrogram and ensures replicability.

In fact, there is at least some empirical evidence to support the intuition that the waveform yields more reliable measurements. Francis *et al.* (2003) analyzed a number of acoustic measurements which are used fairly commonly to assess voicing onset after stop consonants in syllable onset position; these included measures made from the waveform as well as from the spectrogram. Their findings indicate that those measures which rely on the waveform boasted the greatest accuracy, and were also the least variable, as compared with the true onset of voicing determined via an Lx waveform generated from a laryngographic recording. The finding that a waveform-based measure most closely captured the onset of voicing bolsters the claim made here when proposing the utility of the PVI: namely, that the waveform provides more reliable landmarks. Thus PVI is valuable both because it is made relying primarily on the waveform—and too, because it utilizes landmarks which are reliably present.

8 Conclusion

This paper assessed two proposals for post-release temporal measures for oral stops in languages like Marathi, which employs a four-way contrast between plain voiced, plain voiceless, breathy-voiced, and aspirated voiceless stops: namely, Noise Offset Time (NOT) and After Closure Time (ACT). While both measures show differences across stop types, they are each subject to some significant drawbacks. Due to these drawbacks, the alternative measure of Pre-vocalic Interval (PVI) was proposed. PVI is measured from the burst associated with the stop closure release to the jump in amplitude in the waveform that coincides with a darkening of formants in the spectrogram, landmarks which are present regardless of the variability of the data. PVI is proposed as a measure that is both more reliable and more replicable than the others assessed.

This is a preliminary study, and much remains to be investigated. Future studies should include more tokens, with stops in additional positions. Word-final and word-medial stops can and should be included—there are no syllable-position restrictions on Marathi oral stops. This will make it possible to measure stop closure duration. Finally, this study looked at a single vowel context; others should be considered in the future.

It is also important to note that a temporal measure for use with oral stops is but one of the tools we will need moving forward. Ultimately, it is hoped that analysis of Marathi consonants will contribute to a more complete understanding of laryngeal contrasts in general and of breathy voiced sounds in particular.

Breathy voiced consonants are fairly rare typologically—only 13 out of 451 (about 2.9%) of the languages in UPSID (the UCLA Phonological Segment Inventory Database) contain breathy voiced stops. Breathy voiced sonorants are even more rare, appearing in fewer than 1% of the languages in UPSID. Marathi has both, however, making it an ideal language in which to investigate the consistency among and divergence between breathy voiced sonorants and obstruents. The findings of such investigation have the potential to shed light on the crosslinguistic rarity of breathy voiced sounds in general and of breathy voiced sonorants in particular. As such, the present study serves as the starting point for a much larger investigation.

Appendix A: Stimulus List

1.	[pats]	*five*	9.	[ʈats]	*heel*
2.	[pʰaʈa]	*fork (in road)*	10.	[ʈʰam]	*firm (mentally)*
3.	[bara]	*twelve*	11.	[ɖag]	*stain*

4.	[bʱat]	*rice*	12.	[ɖʱal]	*shield*
5.	[tas]	*hour*	13.	[kal]	*yesterday*
6.	[tʰaɭi]	*plate*	14.	[kʰas]	*special*
7.	[dar]	*door*	15.	[gal]	*cheek*
8.	[dʱar]	*sharpness*	16.	[gʱar]	*hawk*

References

Deshpande, Dinker. (1990). Palatal sibilant: Aspect of aspiration. *Osmania Papers in Linguistics*, 16: 1-4.

Census of India. (2001). *Statement 1: Abstract of speakers' strength of languages and mother tongues.* Census of India Website: Office of the Registrar General & Census Commissioner, India. Online resource available at <http://censusindia.gov.in/ Census_Data_2001/Census_Data_Online/Language/Statement1.htm>. Accessed September 20, 2011.

Davis, K. (1994). Stop Voicing in Hindi. *Journal of Phonetics*, 22. 2: 177-193.

Dhongde, R.V. & K. Wali. (2009). *Marathi.* Amsterdam: John Benjamins Publishing Co.

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

Francis, A., V. Ciocca & J. M. Ching Yu. (2003). Accuracy and variability of acoustic measures of voicing onset. *JASA*, 113.2: 1025 – 1032.

Ladefoged, P. and I. Maddieson. (1996). *Sounds of the World's Languages.* Cambridge, MA: Blackwell Publishers, Inc.

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

Masica, C. (1991). *The Indo-Aryan languages.* Cambridge: Cambridge University Press.

Mikuteit, S. & H. Reetz. (2007.) Caught in the ACT: the timing of aspiration and voicing in Bengali. *Language and Speech,* 50.2: 247-277.

Pandharipande, R. V. (1997). *Marathi.* London: Routledge.

Schiefer, L. (1987). 'Voiced aspirated' or 'breathy voiced' and the case for articulatory phonology. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München,* 27: 257-278.

UCLA Phonological Segment Inventory Database. http://web.phonetik.uni-frankfurt.de/cgi-bin/upsid_sounds.cgi. Accessed October 14, 2011.

Author information:

Kelly Harper Berkson: kelly764@ku.edu