# An Acoustic Analysis of Sylheti Phonemes

Amalesh Gope and Shakuntala Mahanta

Phonetics and Phonology Lab, Department of HSS, IIT Guwahati, India
amaleshgope5sept@gmail.com. smahanta@iitg.ernet.in

## ABSTRACT

This paper examines the acoustic properties of Sylheti phonemes. Sylheti is generally regarded as one of the varieties of Bangla. The historical development of this language witnessed significant reduction and reconstruction of its phoneme inventory. The phoneme inventory is considerably reduced due to the phonological process of deaspiration [+spread glottis], spirantization and deaffrication (Gope & Mahanta, 2014). We conducted an acoustic experiment and measured the voiced onset time (VOT) of all the voiced stops. The result of one-way ANOVA did not show any significant interaction among the obstruents in terms of aspiration ($p > 0.05$, [$F$ (1, 359) = 0.095, $p$ = 0.76)]. In a separate experiment, we examined the acoustic qualities of Sylheti vowels. Results confirm the presence of 5 vowels in Sylheti. A one way ANOVA confirmed significance effect on vowel quality in terms of duration [$F$ (4, 600) = 57.77, $p$ = 0.00] and (first three) formants values.

**Keywords**: phoneme, voice onset time, duration, formants.

## 1. INTRODUCTION

Sylheti is generally regarded as one of the varieties of Bangla. There are approximately 10,300,000 people using Sylheti as their primary language (7,000,000 in Bangladesh) (Lewis et al. 2013). The most significant properties that distinguish Sylheti from Standard Colloquial Bangla (henceforth SCB) or other regional varieties is the extensive application of obstruent weakening involving de-aspiration, spirantization and deaffrication. Consequently, the consonant inventory (especially the obstruents) of Sylheti exhibit a major reduction and restructuring compared to that of SCB. The loss of (underlying) breathiness contrast [+spread glottis] of the entire stop series due to the phonological process of obstruent weakening eventually led to deaspiration of voiced ([d̪$^{h}$an > **d̪**an] 'paddy', [d̪ud̪$^{h}$ > d̪u**d̪**] 'milk') and voiceless stops ([t̪$^{h}$ala > **t̪**ala] 'plate', [mat̪$^{h}$a > ma**t̪**a] 'head'). The process of spirantization targeted the underlying voiceless bilabial stop (both aspirated and unaspirated) [p] and [p$^{h}$] ([por] > [**ɸ**ɔr] 'read', [p$^{h}$ul] > [**ɸ**ul] 'flower') and voiceless velar stop (both aspirated and unaspirated) [k] and [k$^{h}$] ([kali] > [**x**ali] 'ink', [k$^{h}$al] > [**x**al] 'drain/channel'). There was a parallel process of deaffrication of voiceless alveo-palatal affricate [tʃ] and [tʃ$^{h}$] ([tʃa] > [**s**a] 'tea', [tʃ$^{h}$uti] > [**s**uti] 'holiday') and voiced alveo-palatal affricate [dʒ] and [dʒ$^{h}$] ([dʒal] > [**z**al] 'net', [dʒ$^{h}$al] > [**z**al] 'spicy') and together these processes reduced and restructured the phoneme inventory of the language (Gope & Mahanta 2014).

Our findings also suggest that even the vowels in Sylheti have been affected and restructured. There are seven oral vowels in SCB (viz., [i], [e], [æ], [a], [u], [o] and [ɔ]) (Chatterjee 1926, Bhattacharya 1910). However, we have found that the half-open front vowel [æ] and half-closed back vowel [o] have been merged with [e] and [u] respectively and thus reducing the number of vowels to five ([i], [e], [a], [ɔ] and [u] in Sylheti.

This paper demonstrates the phoneme inventory of the language under study. Results of two acoustics experiments have been discussed in this study. The first experiment is meant to explore the status of the underlying property of aspiration [+spread glottis] of Sylheti obstruents in terms of durational measurements (VOT). The second production experiment is conducted to understand and determine total number of Sylheti vowels. The process of spirantization and deaffrication is discussed with the help of spectrographic evidence along with adequate examples.

## 2. SYLHETI PHONEMES: CONSONANTS

To understand how the remnant feature [+spread glottis], may have affected consonants, we conducted an acoustic experiment and measured the voiced onset time (VOT) of all the voiced stops present in the language following the theory proposed by Lisker and Abramson (1964).

### 2.1. Experimental Methodology: VOT

For the acoustic experiment, 18 different words representing each of the target voiced stops occurring word initially were carefully chosen (Table 1). As could be seen from the dataset given in the Table 1, the bilabial voiced stop [b] is represented word-initially in 4 different words (viz., [**b**at̪] 'arthritis', [**b**an] 'tie', [**b**ala] 'bracelet' and [**b**ari] 'home'). The other set of words in the contrastive pairs (viz., [**b**at̪] 'rice', [**b**an] 'pretend', [**b**ala] 'good' and [**b**ari] 'heavy') represent the (underlying) aspirated voiced stop [**b$^h$**] which at some point of history had underlying aspiration, and appeared as distinct phonemes. The vowel following the target stops was an [**a**] and form (near) minimal pairs. Thus, the dental voiced stop [**d̪**] (along with the underlying aspirated counterpart) is represented in two words, while the retroflex voiced stop [ɖ] and the velar voiced stop [g] (along with their aspirated counterparts are represented in 4 different words each. Our major goal was to measure the VOT of each of the target stops and compare the intrinsic phonetic variations involving place of articulations and aspiration.

**Table 1**: Dataset considered for VOT experiment

| Sylheti words | Gloss | Sylheti words with a history of underlying aspiration | Gloss |
|---|---|---|---|
| [bat̪] | 'arthritis' | [bat̪] | 'rice' |
| [ban] | 'tie' | [ban] | 'pretend' |
| [bala] | 'bracelet' | [bala] | 'good' |
| [bari] | 'home' | [bari] | 'heavy' |
| [d̪an] | 'donate' | [d̪an] | 'paddy' |
| [ɖax] | 'roaring of cloud' | [ɖax] | 'drum' |
| [ɖala] | 'tray' | [ɖala] | 'pour' |
| [ga] | 'body' | [ga] | 'wound' |
| [gai] | 'cow' | [gai] | 'stroke' |

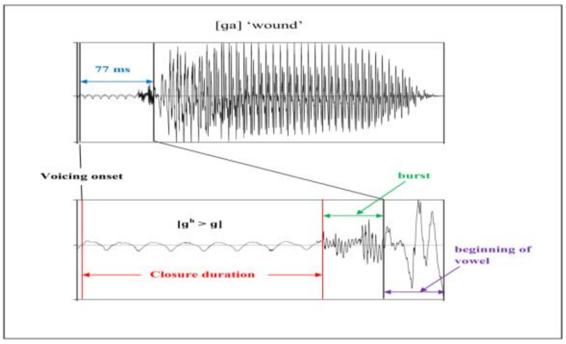
*2.1.1. Participants and recording Procedure*

Eight native speakers of Sylheti (six male, two female) were recorded in a quiet environment in Dharmanagar district of North Tripura, India. Apart from Sylheti, the speakers were also fluent in Hindi, and English. Speech data was recorded with a *Shure* unidirectional head-worn microphone connected to a *Tascam* linear PCM recorder (confirming a constant mike-to-mouth distance) via xlr jack. The material with target stops was displayed on a computer screen. The meaning of each word was written along with the sentence frame. Subjects were asked to pronounce each word with natural intonation. To avoid the effect of neighbouring sounds, each word was recorded individually with a considerable amount of gap between two words. Apart from the target stimuli mentioned in Table 1, an additional 20 words were also placed as fillers in the dataset. All the words (along with the target stimuli) were randomized and presented on three different lists, thus ensuring each stimuli was recorded three times.

*2.1.2. Acoustic Measurement*

Dutta (2007) claimed that voice lead time (VLT) (along with the $f_0$ measurements of the following vowel) can successfully lead to a distinction between voiced aspirated and unaspirated counterparts; VLT of voiced aspirate stops will be shorter and hence will reduce the $f_0$ of the following vowel more than the unaspirated voiced stop. Besides, we did not observe any noise like signal following the release in the interval that was proposed by Lisker and Abramson (1964) as the vital factor to distinguish the two voiced categories of aspirated and unaspirated (Figure 1). Hence we decided to consider onset of the release till the offset of the voicing burst for the durational measurements of the voiced stop categories.

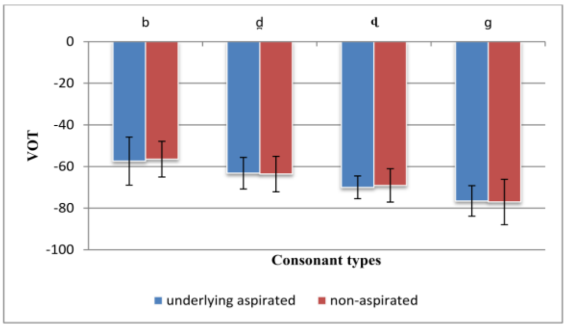
**Figure 1**: Waveform of Sylheti words [ga] 'wound' [g$^{h>}$g]; The enlarged portion of the target phoneme considered for VOT measurement is shown along with.




**2.2. Results and Discussion**

A one way ANOVA was performed on the durational values calculated for underlying aspirated and unaspirated voiced stops using SPSS software (version 20). Altogether 360 tokens were considered for acoustic and statistical analysis. Few tokens were left out due to either distortion from external noise, or in the presence of noises emanating from fits and starts in speaking into the microphone. For the ANOVA test, Voicing types were kept as categorical factor (Independent factor, with two levels viz.; underlying aspiration and unaspiration) and duration values as variables. As expected, the underlying voiced aspirate stops did not show any difference with their unaspirated counterparts ($p > 0.05$, [$F$ (1, 359) = 0.095, $p$ = 0.76)]. The duration values of individual voiced stop are shown in Figure 2.

**Figure 2**: Average duration of the voiced stop consonants with standard deviation as error bars




To observe the interaction among the voiced stop categories a subsequent post-hoc Tukey test was also conducted on the data from all the speakers (Table 3). The findings of post-hoc Tukey test revealed that Sylheti stop categories significantly differ from each other in terms of their place of articulation (POA) except for the pair [d̪] and [ɖ] which could be due to less number of tokens considered for the statistical analysis. As assumed no significant interaction was observed between individual pair of unaspirated stops and their aspirated counterparts. No significant interaction was also observed between [b] and [d̪], however, the underlying aspirated [d̪] and [b] differ significantly from each other where the former was found to be significantly shorter (in duration) than the later. Similarly, underlying aspirated [b] and unaspirated dental [d̪] and underlying aspirate [d̪$^{h}$] is observed to be significantly different.

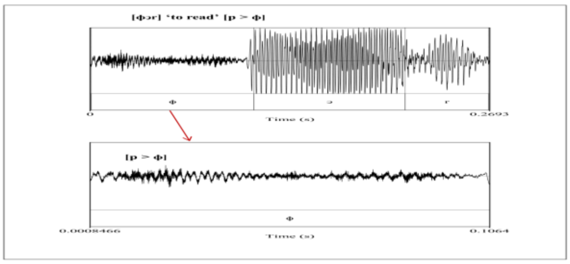
**Table 2**: Significant matrix for duration of Sylheti stop consonants

| Consonant Types | N | b | b$^{h}$ | d̪ | d̪$^{h}$ | ɖ | ɖ$^{h}$ | g |
|---|---|---|---|---|---|---|---|---|
| b$^{h}$ | 80 | ! | | | | | | |
| d̪ | 22 | ! | * | | | | | |
| d̪$^{h}$ | 22 | * | * | ! | | | | |
| ɖ | 36 | * | * | ! | ! | | | |
| ɖ$^{h}$ | 34 | * | * | ! | ! | ! | | |
| g | 43 | * | * | * | * | * | * | |
| g$^{h}$ | 43 | * | * | * | * | * | * | ! |

## 3. SPIRANTIZATION AND DEAFFRICATION

The underlying forms of voiceless bilabial stop (both aspirated and unaspirated) [p] and [p$^{h}$] ([por] > [**ɸ**ɔr] 'read', [ʃap] > [ʃa**ɸ**] 'snake', [p$^{h}$ul] > [**ɸ**ul] 'flower', [lap$^{h}$] > [la**ɸ**] 'leap/jump') and voiceless velar stop (both aspirated and unaspirated) [k] and [k$^{h}$] ([kali] > [**x**ali] 'ink', [kak] > [**x**a**x**] 'crow', [k$^{h}$ali > **x**ali] 'empty', [muk$^{h}$] > [mu**x**] 'face') have been spirantized due to the phonological process of consonant weakening. The transformation of voiceless bilabial stop [p] to voiceless bilabial fricative [**ɸ**] is shown in Figure 3. The enlarged portion of the target phoneme is also shown along with.

**Figure 3**: Waveform display of the phoneme [**ɸ**] as captured in the word [**ɸ**ɔr] 'read'



Further, the underlying form of voiceless alveolar affricate (both unaspirated and aspirated [tʃ] and [tʃ$^{h}$]), [tʃira] > [sira] 'flatten rice', [katʃ] > [xas] 'glass', [tʃ$^{h}$ad̪] > [**s**ad̪] 'roof', [matʃ$^{h}$] > [mas] 'fish', and voiced alveolar affricate (both unaspirated and aspirated [dʒ] and [dʒ $^{h}$]), [dʒal > [**z**al] 'net', [dʒ $^{h}$al] > [**z**al] 'hot', exhibit the process of deaffrication and change to alveolar fricatives [s] and [z] respectively. Table 3 shows the consonant inventory of Sylheti

**Figure 4**: Waveform display of the phoneme [**s**] as captured in the word [**sira**] 'flatten rice'

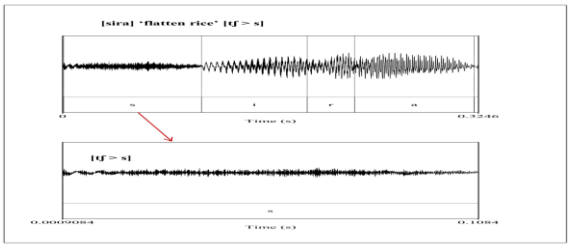


**Table 3**: Consonant inventory of Sylheti

| Manner of Articulation | Bilabial | Dental | Alveolar | Retroflex | Palatal | Velar | Glottal |
|---|---|---|---|---|---|---|---|
| Stop/Plosive | b | t̪ d̪ | | ʈ ɖ | | g | |
| Nasal | m | n | | | | ŋ | |
| Fricative | ɸ | | s z | | ʃ | x | h |
| Affricate | | | | | | | |
| Approximent | | | r | | | | |
| Lateral | | | l | | | | |

## 4. ACOUSTIC ANALYSIS OF VOWELS

To validate the number of vowels and to provide a detailed acoustic analysis of the same, we conducted another production experiment and recorded seven native Sylheti speakers (five male, two female), using the data given in (Table 4). Four different consonantal contexts such as [bVl], [sVɸ], [xVl] and [zVl] (V being the target vowel) were prepared as stimuli that represented all the possible vowels in the language. All the target vowels (embedded in different consonantal contexts) were recorded in a fixed carrier sentence such as [ami cVc xɔiar] 'I cVc said', where cVc is one of the four different consonantal environment and V is the target vowel (see Table 4). Two of the target words (comprising the target vowel) were disyllabic and trisyllabic respectively (viz., [bala] ‘ornament, and [bulbuli] ‘name of a bird’), however, we considered only the first vowel of those two words for the purpose of analysis. All the words were randomized and repeated thrice. Additional thirty words were also used as fillers.

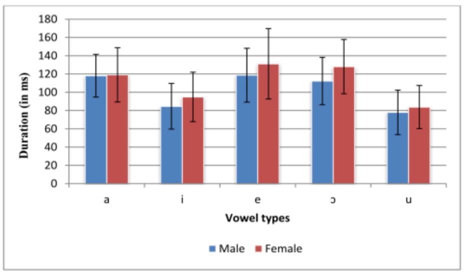
**Table 4**: Dataset prepared for acoustic analysis of the vowels

| Sylheti words | Gloss | Sylheti words | Gloss |
|---|---|---|---|
| [bil] | ‘pond’ | [kil] | ‘fist’ |
| [bel] | ‘wood-apple’ | [xe] | ‘who’ |
| [bala] | ‘bracelet’ | [xal] | ‘skin’ |
| [bɔl] | ‘ball’ | [xɔl] | ‘tube-well |
| [bulbuli] | ‘name of a bird’ | [kul] | ‘lap’ |
| [zil] | ‘lake’ | [siϕ] | ‘press’ |
| [zel] | ‘prison’ | [seϕ] | ‘spit’ |
| [zala] | ‘body’ | [saϕ] | ‘pressure’ |
| [zɔl] | ‘water’ | [sɔϕ] | ‘a type of snack’ |
| [zul] | ‘gravy’ | [suϕ] | ‘quiet’ |

### 4.1. Results and Discussion: Duration

Duration and formant values of each of the target vowels were measured using a Praat script (version 5.3.04_win32) after manually labelling each phoneme. A one way ANOVA conducted on the data from all the speakers confirmed significance effect of vowel quality on duration [$F$ (4, 600) = 57.77, $p$ = 0.00]. The high vowels [u] (Mean duration = 81.04 milliseconds, N = 120) and [i] (Mean Duration = 89.77 milliseconds, N = 120) appear to be the shortest. A successive post-hoc Tukey test was also conducted to observe the interaction between individual vowel pairs. The results revealed that the high vowels [i] and [u] do seem to differ (significantly shorter) from the remaining three vowels in terms of duration. In all the other cases, vowel durations were not found to be significantly different from each other. This pattern was observed for both male and female speakers’ data; i.e., out of all the 5 vowels, only the high vowels [i and u] were found to be systematically shorter than the remaining three vowels.

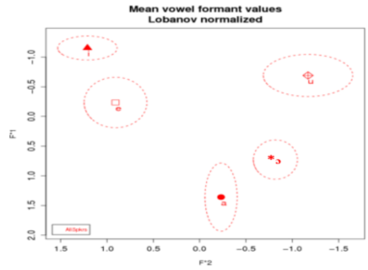
**Figure 4**: Average duration of Sylheti vowels with standard deviation as error bars




### 4.2. Results and Discussion: Formant frequencies

The first three formant values of each vowel were calculated at the mid-point and the values were measured in Mel. Since formant frequencies usually vary across male and female genders (Peterson and Barney 1952), values for the first two formants (F1 and F2) were normalized using Lobanov’s (1971) normalization procedure and were plotted on an F1 – F2 plane using NORM (Thomas and Kendell 2007) (Figure 5).

**Figure 5**: Vowel diagram showing average Lobanov normalized formant frequencies of the first two formants with one standard deviation ellipses. N= 120 per vowel.




The results of the ANOVA test (conducted on non-normalized formant values separately for male and female) confirmed a strong interaction between vowel types based on the first three formant values (for both genders) for each vowel. For all the three formants analysed in this study, we noticed a significant interaction among the vowels (for male speakers: F1: $p$ < 0.05 [($F$ (4, 300) = 83.89, $p$ = 0.000], F2: $p$ < 0.05 [($F$ (4, 300) = 17.38, $p$ = 0.000]), F3 ($p$ < 0.05 [($F$ (4, 300) = 8.56, $p$ = 0.000]; and for female speakers: F1: $p$ < 0.05 [($F$ (4, 300) = 233.37, $p$ = 0.000], F2 ($p$ < 0.05 [($F$ (4, 300) = 87.62, $p$ = 0.000]), F3 ($p$ < 0.05 [($F$ (4, 300) = 20.2, $p$ = 0.000]).

## 5. CONCLUSION

This paper presents the first comprehensive acoustic study of Sylheti phonemes. With the durational measurements, we have shown that the underlying aspirated series do not show any significant difference both in terms of acoustic signal and VOT measurements. We did not observe any noise following the burst of underlying voiced stops (as shown in Figure 1) which prompted us to consider the onset of the voicing start till the burst point. Through acoustic waveforms, we have also discussed the process of consonant weakening involving spirantization and deaffrication. An acoustic analysis of the vowels has also been conducted and vowel inventories of Sylheti have been presented. To conclude, the study at large, lays down the foundation of a previously undocumented language.

## 7. REFERENCES


[1] Abramson, A. S. (1962). The vowels and tones of Standard Thai: Acoustical measurements and experiments. *International Journal of American Linguistics*, 28; 2, part II.

[2] Anisuzzaman et.al. (1988). *Byabaharik Bangla Uchcharan Abhidhan* [A Dictionary of Standard Bengali Pronunciation]. National Institute of Mass Communication, Dhaka.

[3] Bhattacharya, Krishna. (1910). *Bengali Phonetic Reader*. CIIL, Mysore.

[4] Boersma, Paul & Weenink, David (2008). Praat: doing phonetics by computer (Version 5.3.04_win32) [Computer program]. Retrieved on December 22, 2011, from http://www.praat.org/Bradley 1982,

[5] Chatterji, Suniti Kumar. (1971). *The Origin and Development of the Bengali Language*. London: George Allen and Unwin.

[6] Das, Shyamal. (1996). *Consonant Weakening in English and Bangla: A Lexical Phonological Investigation*. M. Phil Dissertation, EFLU (Formerly CIEFL), Hyderabad.

[7] Disner, S. (1980). Evaluation of vowels normalization procedures. *Journal of the Acoustical Society of America* Vol. 67, pp. 253-261.

[8] Dutta Indranil. (2009). *Acoustics of stop consonants in Hindi: Voicing, fundamental frequency and spectral intensity*. Verlag Dr. Müller: Saarbrücken.

[9] Dixit, R. Prakash. (1989). Glottal Gestures in Hindi Plosives. *Journal of Phonetics* 17, pp. 213-237.

[10] Gope, Amalesh. and Irene, Nisha. V. (2010). Linguistic Analysis of West Tripura Variety of Bangla: A Descriptive Study. In C. Sivashanmugam et al. (Eds.). *NewPerspectives in Linguistics* (pp. 198-205). Coimbatore: Department of Linguistic, Bharathiar University.

[11] Gope, Amalesh and Mahanta, Shakuntala. (2014). Lexical Tone in Sylheti in TAL-2014 (pp 10-14).

[12] Gordon, Mathew. (1996). The Phonetic Structures of Hupa. *UCLA Working Papersin Phonetics* Vol. 93, pp. 164-187.

[13] Iseli, markus, Yen-Lian Shue and Abeer Alwan. (2007). Age, sex and vowel dependencies of Acoustic measures related to the voice source. *Journal of the Acoustical Society of America* Vol. 121, pp. 2283-2295.

[14] Kingston, John. (1985). *The phonetics and phonology of the timing of oral and glottal event.* PhD dissertation, UCLA.

[15] Ladefoged, Peter and Ian Maddieson. (1996). *The sounds of the world's languages*. Malden, MA: Blackwell.

[16] Ladefoged, Peter. (1996). *Elements of Acoustic Phonetics* (2nd Edition). The University of Chicago Press, Chicago & London.

[17] Ladefoged, Peter. (1971). *Preliminaries to linguistic phonetics*. Chicago, IL: University of Chicago Press.

[18] Matisoff, J.A., "Sino-Tibetan linguistics: present state and future prospects", Annual Review of Anthropology, vol. 20, 469-504, 1991.

[19] Ohala, Manjari. (1983). *Aspects of Hindi Phonology*. Delhi: Motilal Banarsidass.

[20] Qayyum, Abdul Md., and Sultana Razia. (2007). *Prachin o Madhyajuger Bangla Bhashar Abhidhan* [Dictionary of Old and Medieval Bengali Language]. Bangla Academy, Dhaka.

[21] Rose, P. J. (1987). Considerations on the normalization of the fundamental frequency of linguistic tone. *Speech Communication* Vol. 10, pp. 229-247.

[22] Sapir, Edward. (1925). Sound patterns in language. Language (1): 37–51.

[23] Shahidullah, Md. (1965). *Bangladesher Anchalik Bhashar Abhidhan* [A Lexicon of Bangladeshi Dialects]. Bangla Academy, Dhaka.

[24] Stevens, kenneth and Helen Hanson. (1995). Classification of glottal vibration from acoustic measurements. In Fujimura and Hirano (eds.), pp. 147-170.

[25] Thomas, E. and Kendall, T. 2007. *NORM*: The vowel normalization and plotting suite. Online version: http://ncslaap.lib.ncsu.edu/tools/norm/.