THE PHONETICS AND PHONOLOGY OF SYLHETI TONOGENESIS

BY

AMALESH GOPE

A

DISSERTATION

Submitted in Partial Fulfilment of the Requirements

For the Degree of Doctor of Philosophy in Linguistics

Department of Humanities and Social Sciences

Indian Institute of Technology Guwahati

Guwahati, Assam-781039

2016

DECLARATION

This is to certify that the dissertation entitled "The Phonetics and Phonology of Sylheti Tonogenesis", submitted by me to the Indian Institute of Technology Guwahati, for the award of the degree of Doctor of Philosophy in Linguistics, is an authentic work carried out by me under the supervision of Dr. Shakuntala Mahanta. The content of this dissertation, in full or in parts, have not been submitted to any other University or Institute for the award of any degree or diploma.

Signed:

Amalesh Gope.
Department of Humanities and Social Sciences
Indian Institute of Technology Guwahati
Guwahati -781039, Assam. India.

Date: 05.09.16

CERTIFICATE

This is to certify that the dissertation entitled "The Phonetics and Phonology of Sylheti Tonogenesis" submitted by Mr. Amalesh Gope (Registration Number: 11614106), a research scholar in the Department of Humanities and Social Sciences, Indian Institute of Technology Guwahati, for the award of the degree of Doctor of Philosophy in Linguistics, is a record of an original research work carried out by him under my supervision and guidance. The dissertation has fulfilled all requirements as per the regulations of the institute and in my opinion has reached the standard needed for submission. The results embodied in this thesis have not been submitted to any other University or Institute for the award of any degree or diploma.

Signed:

Supervisor: Dr. Shakuntala Mahanta.
Department of Humanities and Social Sciences
Indian Institute of Technology Guwahati
Guwahati -781039, Assam. India.

Date: 05.09.16

To my Parents and my Love...

Acknowledgements

This has been a long journey that had started on 21st July, 2011. Finally the dissertation is ready, the journey is about to complete. Among so many hardships and challenges that I encountered during my PhD program, it was actually a golden opportunity to experience the kindness of so many people around me, who helped me in numerous ways, and made things easier for me. They definitely deserve to be acknowledged!

First and foremost, my sincere thanks and gratitude are due to Dr. Shakuntala Mahanta, my advisor and mentor, and above all the exceptional human being who I admire the most. I was fortunate enough to be selected by her to begin my journey as a PhD scholar. During this long journey I was offered complete intellectual freedom and left uninterrupted to discover my own path. At the same time, her attentive eyes made it sure that I was never deviated from the actual journey. Working with her definitely extended my perception of phonology as a whole. I am yet to overcome the higher ranked constraint *Intellectual. This work would not have been possible without her extensive knowledge, constant encouragement and support. Thank you so much Ma'am for being so patience and kind.

Words can also not express entirely, my gratitude for my doctoral committee members. I owe a debt of gratitude to Dr. Priyankoo Sarmah for all his insights, suggestions and critical feedback into this study. Dr. Sarmah has been a major source of encouragement and knowledge especially when I started my venture into the world of *Tone*. The long debates and caring advice definitely raise the bar of this work and helped me to look into phonetics with a whole new perspective altogether.

I would also like to thank Prof. Rohini Mukashi Punekar- the chairman of my Doctoral committee for all the support and inspirations. My sincere gratitude is also due to Dr. Bidisha Som for her well wishes, comments and valuable advice.

I am truly indebted to Prof. Bert Remijsen for his kind advice, informative suggestions and all the positive feedback. Thank you so much for the detailed comment on the third chapter of this dissertation. I would also like to thank Prof. Peri Bhaskara rao for sharing one of his work and for all the encouragement and interest in my work.

Thanks are also due to Prof. Arupjyoti Saikia, HoD, Department of HSS, IITG and all the other faculty members at the Department. Sincere thanks are also due to all the office bearers of our department- Bandana ba, Mr. Para gKalita, and Durga da.

Over the years, I was fortunate enough to spend quality times with all my friends and colleagues- I candidly thank my fellow researchers and all the project staffs of *Phonetics and Phonology Lab*, IITG. Special thanks to Asim da for all the support, encouragements and inspirations during my stay at IITG. Thank you Kalyan da, Kanak da for all your well wishes. Thank you so much Priti for all the little insights in Sylheti. Thanks are also due to my other colleagues in the department- Luke, Anirudha, Pranti, Mahsina, Munmiba, Aleendra da, Arfan, Bandana, Madhu, Nirmalaba, Pankaj da, Rupan da, Rajashree, Prarthana, Leena, to name a few. A big hug to all of you guys.

I also express my gratitude to all my friends at IITG- Arpan, Tanmoy, Polash, Subhranshu, Mukesh, Madhu, Debu, Shibsankar, Rohan, Soshi, Suju, Ebria, among so many. Special thanks to Anirban da for helping me out with the statistics related problems. I thank my dear friend Mahesweta Chakraborty for being so kind and caring. This journey wouldn't have been possible without your help and support. Thanks for arranging my stay at your home and all the delicious food during my several field visits. Thanks are also due to all your students and relatives who actively participated in the recordings during my multiple round of field trips. Thank you so much for arranging all the subjects for my experiments. I am extremely thankful to Uncle and Aunty for providing all the details of Sylheti. I would also like to thank Avik mama (uncle) for his constant patience to make the recordings all perfect.

On the personal front, I have to thank my fellow linguist friends from various parts of India- Tariq, Abhijit, Shilendra, Atul to name a few. Thank you so much Nisha for everything. Your silent support and encouragement truly helped me to complete this project. Thank you so much for all the care and inspiration.

My humble gratitude is also due to all my school buddies- Pranab, Uttam, Biplab, Abhi, Nabin, and Baban. Special thanks are due to all the members of WhatsApp group *Friends Union Zindabaad* for giving me company whenever I needed a break late at night. You guys are just amazing.

Thank you so much Anusuya. I was really confused how to express my gratitude for you. Thank you so much for the unconditional support and help, encouragements and motivations. Your extra ordinary confidence on me really inspired me to cross several hurdles so easily.

Personally I would also like to thank Mr. Narayan Pal and Mrs. Aparna Pal for all the encouragement, care and love.

I would also like to thank, Dr. Asish Das Gupta and his family- your encouragement and care definitely inspired me to work harder.

I am truly grateful to all the stuff members of Lakshminath Bezbaroa Central Library, IITG. My sincere gratitude is due to respected Librarian sir Dr. Tamal Kumar Guha, Senior Library Information Officer Mrs. Gitasri Bordoloi, and Mr. Chandan Kumar Goswami.

I am thankful to all my uncles and aunts- Dr. Nirod Ghosh and Mr. R. R. Ghosh and my late uncles Mr. Sudhir Gope and Mr. Sunil Gope and Mr.Parimal Ghosh. I am sure you must be so happy for me and I seek your blessings as always. To my aunty Mrs. Shanti Gope, Mrs. Parul Gope and Mrs. Anjali Ghosh. Thanks are also due to my cousins Mr. Pradip Gope and Mr. Sushanta Gope.

A big shower of gratitude to all my sisters- *Fuldi*(Anindita), *Mezdi* (Smriti), *Sezdi* (Priti), and *Rangadi* (Dipti) and my brother-in-laws- Mr. Haridas, Mr. Babul, Mr. Bijoy and Mr. Radheshyam. Thanks are also due to all my niece and nephews Hirendra, Biplab, Joydeep, Sayentika, Bishal, Pawani, Om and Rishita. Your uncle would be home soon.

I am indebted to my parents in several ways. Thank you so much for instilling belief in my abilities and providing unconditional support, love beyond imagination, and the right foundation for pursuing life with dedication, honesty and courage. I am truly sorry for hurting you both on several occasions, especially when I refused to be a Doctor or an Engineer! See the irony! Your son ended up earning a PhD (a Doctorate) and that too from one of the premier Engineering institutions of India. Thank you so much for everything and bearing with my absence at home. I promise to be back soon, and this time I would come leaving my entire burden behind.

I would also like to express my heartfelt gratitude to my in-laws- Mr. Sanjay Choudhury and Mrs. Gouri Dutta (Choudhury) for all the motivations, encouragements and support. I also thank my sister-in-law Mrs. Paramita Choudhury, brother-in-law Mr. Rana Sarma and the cutie-pie of the family- Adi.

Last but not the least; I am truly grateful to my wonderful wife Ranita who sacrificed her post-marital happiness just to see me complete this project. Thank you so much for bearing with my long silence and excessive use of `hmmmm' to every possible question. Thank you for shouldering up all the burdens and responsibilities of the family in my absence. And thank you so much for the most precious gift of my life- our daughter Anurita! To my daughter, Pa loves you so much baby! I am dying to hold you in my arms and play with you whole day; I wish to grow up with you once again. Much love and blessings.

IIT Guwahati.
25th January, 2016.

Abstract

The physical constraints of articulatory and/or auditory speech mechanism are vital to the sound change of human languages (Hombert 1977, Ohala 1974, 1993). These constraints often affect the way speech sounds are produced and perceived by listeners. In this dissertation, with the help of acoustic experiment(s) we first explored the phoneme inventory of a previously undocumented language i.e. Sylheti. The phoneme inventory of this language is substantially reduced due to the loss of underlying breathiness contrastt: both voiced and voiceless aspirated stops became unaspirated ([d̪^{h}an > **d̪**an] 'paddy', [t̪^{h}ala > **t̪**ala] 'plate'), and the voiceless labial and dorsal stops went one step further and spirantized to homorganic fricatives([pɔr] > [**ɸ**ɔr] 'read', [p^{h}hul] > [**ɸ**ul] 'flower', [kali] > [**x**ali] 'ink', [k^{h}al] > [**x**al] 'drain/channel'). The post-alveolar affricates (both aspirated and unaspirated) also spirantized to dental fricatives[tʃa] > [**s**a] 'tea', [tʃ^{h}uti] > [**s**uti] 'holiday', [dʒal] > [**z**al] 'net', [dʒ^{h}al] > [**z**al] 'spicy'). Among the voiceless stops, only the dental and retroflex stops remained stops.An apparent consequence of the loss of underlying breathiness contrast is the development of high tonein words that originally contained an aspirated stop, while a low tone developed in other words ([d̪án] 'paddy' [d̪^{h} > d̪], and [d̪àn] 'donate', [bát̪] 'rice' [b^{h} > b], and [bàt̪] 'arthritis'. This study explores the tone system of Sylheti and provides a detailed phonetic and phonological account of Sylheti tonogenesis.

The foundation of canonical theories of tonogenesis is driven by the idea of interaction between consonantal features and the predicable pitch patterns of the following vowel. Generally voiced obstruents are expected to lower the f_0 of the following vowels whereas their voiceless counterparts may even raise it. However, such predictions do not seem to work in case of Sylheti. Further, we have also observed that in Sylheti, each homophonous pair is marked with a contrastive tone, regardless of the voicing property of adjacent (onset) consonant. As such the loss of (underlying) breathiness property of Sylheti obstruents is not limited to a high tone (only)- a low tone can also evolve if the conditioning environment of the homophonous pair witnesses a similar pattern of evolution (such as the loss of underlying breathiness in both the words of homophonous pair):[([xál] 'drain/channel' [k^{h} > x] [xàl] 'skin' [tʃ^{h} > x]). Similarly,

a low can also evolve following a voiceless obstruent ([ɸɔ́r] 'read' [p > **Φ**], and [ɸɔ̀r] 'guard' [p > **Φ**]). The phonological patterns, which seem to be divergent as observed in case of Sylheti, i.e., the unpredictability of consonant-tone interaction, are explained through the phonetic motivations for these changes. A separate acoustic experiment was conducted to understand the phonation qualities of the vowels carrying contrastive tones: results confirm that the vowels bearing low tones have slightly 'laxer' voice qualities than the vowels bearing a high tones. This study further reveals that contrastive tones in Sylheti are perceived in a categorical manner. Finally, this study also confirms that tone assignment in morphologically derived words (as observed in the process of Sylheti suffixation)yields a process of tonal complexities- the suffixes attached to noun roots are toneless and takes the tone opposite to that on the root they are attached to; thus displaying a case of tonal polarity. The verbal suffixes, on the other hand, bear an (underlying) high tone, and displays an instance of tone reversal; i.e., when added to a verb form with underlying high tone, the derived verb form surfaces with a post affixation low tone.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

General Introduction

Any human language system is considered as a system comprising of a set of [minimal] meaningful sound units termed as phonemes. Human languages of the world exploit different phonemes (consonants and vowels forming different minimal pairs) to distinguish the lexical meanings of words. Thus, the Standard Colloquial Bangla word [kal] 'tomorrow' is different from [k^hal] 'canal/drain' or [gal] 'cheek' since the first consonant of each word [k], [k^h] and [g] are different. Similarly, the word [kal] 'tomorrow' is also different from the word [kil] 'a fist', since their vowels [a] and [i] are different from each other. Whereas such minimal pairs of words are universally present in all human languages, the number of phonemes (consonants, vowels, semi-vowels and diphthongs) used to employ lexical differentiation varies from one language to the other and are subjected to change. The physical constraints of articulatory and/or auditory speech mechanism are vital to sound change of human language (Hombert 1977, Ohala 1974, 1993). Researchers (Hombert 1977, Hombert *et al.* 1979) also suggest that these constraints may affect the way speech sounds are produced and perceived by listeners. A speaker's pronunciation may get distorted and may not be perceived as intended, thus creating space for inclusion and/or reduction of phoneme inventory. Several factors causing sound change over generations are found in all human speakers.

One of the striking outcomes of such sound change such as the loss of voicing contrast, or loss of breathiness property among the obstruents is the emergence of lexical tone in many of the world's languages. A language is categorized as a tone language if any variation in pitch of a certain lexical item can indicate change the meaning of that word. Languages such as Mandarin use a particular pitch pattern on a syllable in order to create semantic distinction. Therefore, by tone we understand that, it is used to determine the contrastive pitch patterns that act as minimal word-differentiating elements, comparable to the inventory of vowels or consonants of a language. Yip (2002) pointed out that 60-70 per cent of the world languages are tonal.

In addition to being the source of lexical contrast, pitch variation is also used to convey discoursal meaning, thus various other information such as emphasis, contrast, emotions, etc. are expressed by contrastive pitch height on larger linguistic units of phrases and sentences. Those languages are called intonational languages. Intonation therefore provides specific discoursal meaning to the utterances that is independent of the meanings of the words themselves.Languages such as English, use intonation to add functional meaning to a word or phrase.

The principal thrust of this dissertation is to investigate the former function of pitch variation- that is, to explore the phonetics and phonology of lexical tone in Sylheti, often regarded as a variety of Bangla[1] (Chatterjee 1971), that remains essentially undocumented till date.

1.1 Aims of this dissertation

A general survey of this language indicates substantial reduction and reconstruction of its phoneme inventory (see Chapter 2). The language has lost the entire series of aspirated stops. The phoneme inventory is further reduced due to the phonological process of spirantization and deaffrication. These changes in turn give birth to numerous homophones. Keeping these insights in mind this dissertation aims to concentrate on the following objectives:

1. To explore the phoneme inventory of Sylheti and to investigate the status of

[1]see chapter two of this dissertation for a detailed discussion on this.

underlying breathiness property among the obstruents.

2. To what extent the loss of the underlying property of breathy voice contrast [+spread glottis] among the obstruents and the process of lenition (viz. spirantization and deaffrication) affect the pitch contrast among the homophones?

3. To determine the number of lexical tones in Sylheti and the way they occur in underived monosyllables, derived and underived disyllables, and derived polysyllables.

4. To investigate the phonation qualities of the vowels carrying contrastive tones and to explore the co-relation/interaction between tone and voice qualities of the vowels.

5. To examine how the contrastive tones are perceived in Sylheti? Further to investigate as to what are the perceptual cues employed by the native speakers?

The experimental findings of this dissertation provide a conclusive phonetic and phonological account of Sylheti tonogenesis. It is worth mentioning at this point that till date no studies have been conducted on Sylheti at any (linguistic) level in general and tone in particular. Further, contrary to the 'standard view' of classical theories of tonogenesis where (aspirated) voiced obstruents (stops) are classified as pitch suppressor of the following vowel (Yip 2002, Hombert 1978, Bradshaw 1999, Dutta 2007), the extensive experimental findings of this study confirm that such predictabilities do not seem to work in Sylheti. The loss of underlying breathiness contrast of Sylheti obstruents (both voiced and voiceless) is compensated with a higher f_0 in the neighbouring vowel in most of the cases- though not always. Further, this study also observed that each homophonous pair is marked with contrastive tones regardless of the voicing property of the adjacent (onset) consonant. As such a low tone was also observed after a voiceless consonant ([ɸɔ́r] 'read' [p > ɸ], and [ɸɔ̀r] 'guard' [p > ɸ]) (Gope and Mahanta, 2015, 2014). This dissertation also provides a detailed analysis of the phonation qualities of the vowels carrying contrastive tones. The findings through various spectral measurements indicate that the vowels carrying higher f_0 are in the continuum of modal to creakiness. The following section provides an overview of tone, tone languages of the world in general and discusses the process of tonogenesis and the interaction between consonant and tone in particular.

1.2 Overview of Tone

Tone languages are distinguished from the non-tonal languages in the sense the lexical distinction of a word can also be achieved (apart from the phonemic contrast in a minimal pair) by varying the pitch height or pitch contour of a syllable/vowel in a tone language. Pitch is one of the major characteristic features of phonologically distinctive tones in many languages of the world. At this point it is relevant to understand the concept of tone.

Researchers often do differ in generalizing the concept of tone. Pike (1948) is of the opinion that a tone language is the one that has lexically significant contrastive pitch on each syllable; however the pitch on each syllable needs to be relative. In classical generative phonology tone is regarded as a property of segments (Schachter and Fromkin 1968, Sarmah 2009). Goldsmith (1976) also considered the tone bearing unit to be an element of the segmental tier.

Hyman (2001) provides the most plausible definition of a tone language: "a language with tone is one in which an indication of pitch enters into the lexical realization of at least some morphemes" (Hyman 2001, Yip 2002). Yip (2002) further suggested that unlike tone languages, pitch in stress languages does not stay constant on lexical items which draw a thin line of distinction between these two types of languages. She however holds the opinion that the 'accentual' languages (not all) those have a small number of 'sparsely distributed tones' are often analyzed as a sub-type of tone language. For example, languages like Cantonese, Zulu, Swedish, do have accent and lexical tonal distinction, languages like English, Spanish do not have lexical tonal distinction; even though they all fall under accentual language. Whereas the former group is termed as 'accentual tone language'; the latter is called 'stress accent language' (Woo 1969, Sarmah 2009). Languages like Japanese, Mende, Haya, have very restricted tone, and thus form a third category- these languages are called ‘pitch accent languages’. A tone language, therefore, is the one in which the pitch of a lexical element is capable of changing the lexically distinctive meaning of that lexical item.

Yip (2002) also argues that tone is a phonological category that distinguishes two words or utterances. Tone could be discussed in association to fundamental frequency (f_0) and pitch, which is the perceptual correlate of fundamental frequency. f_0 refers to signal, i.e. the number of pulses per second that a signal contains where each pulse is produced by a single vibration of the vocal cord. The frequency of these pulses is measured in Hertz (Hz) and one Hz is one cycle per second. Pitch is considered as the perceived f_0, i.e., the way a hearer perceives a signal: whether it is heard as high or low in pitch. The pitch of sound thus grasps the foundation of tone.

In Mandarin, for example, four pitch patterns referred to as tones function as phonemes: High (Tone 1), Rising (Tone 2), Low (Tone 3), and Falling (Tone 4) (Yip 2002). They alter the lexical meanings as shown below:

- Tone 1 /ma/ [high] ‘mother’
- Tone 2 /ma/ [rising] ‘hemp’
- Tone 3 /ma/ [low] ‘horse’
- Tone 4 /ma/ [falling] ‘scold

In languages like Dagaare (a Gur language spoken in Ghana) tonal distinction is achieved by the contrastive positioning of restricted tones in different syllables:

- LH [yuori] ‘penis’
- HL [yuori] ‘name’ (Yip 2002)

Yip (2002) also mentions that the exact location of tone may also change depending on the morphological complexity and phonological environment.

1.2.1 Tone languages of the world

Literatures on tonal languages identify three areas of the world viz., Africa, East and South-East Asia and the pacific, and the Americas as the home to so many tone languages (Yip 2002). Whereas Sub-Saharan Africa is the home to the largest concentration of tone languages of the world, the existence of tonal languages in the regions such as Australia, New Zealand are almost next to none (Yip 2002). In the East and South-East Asia most of the languages belonging to Sino-Tibetan and Mon-Khmer language families form the belt of tonal languages in that region. However, in East-Asia, Japanese is often considered an accentual language whereas in case of Korean, pitch differences are realised by the laryngeal properties such as voicing, aspiration and glottalization of onset consonants (rather than lexical tone) (Jun 1998). In the Pacific region, Papua New Guinea and Iranian Jaya are home to large number of tonal languages such as Siane (James 1981), Golin (Hayes 1995), Iau (Edmondson *et al.* 1992). Based on the data collected from languages of those regions, the generalizations of tonal phenomena are drawn. Similarly most the languages belonging to Otomanguean family of Central America are recognized for their tonal property. For example, the most prominent aspects of African tone languages are the mobility of tone to adjacent morphemes (as could be seen in Bantu languages). Other tonal phenomena include the process of spreading, deletion and metathesis. Most of these languages are marked with only one phonological (H) tone, and toneless syllable is often associated with a default L tone. Asian and Pacific tone languages, on the other hand, tend to have lesser complex syllables- in most of these languages morphemes are entirely or mostly monosyllabic (Yip 2002).

1.2.2 Tone languages of the Indian sub-continent

Among the language families of Indian sub-continent, Dravidian languages are generally non-tonal, whereas only a few Indo-Aryan languages such as Punjabi, Dogri, (Bahl 1957, Hombert *et al.* 1979, Vijaykrishnan 2003, Bhaskararao 1998), Kalam Kohistani (Baart 1997, Yip 2002), etc are reported to have lexical tones. Among the Indo-Aryan languages, tonal specification in Punjabi is rather well established. In Punjabi, breathy voiced consonants lost the breathy specification and the following vowel acquired a low tone ([b^{h}a] $\longrightarrow$ [pà]) (Hombert, Ohala and Ewan, 1979, Vijayakrishnan 2009, Rao 1998, Bhatia 1975). Bhatia (1975) claimed that tone in Punjabi is developed in four different stages formulating the modification of an aspirated stop. In the first stage, the voiced aspirated and [h] were voiced in all environment in the dialect area of Punjabi and Pahari, thus resulting in low or rising tone on the following vowel and a high or falling tone on the preceding vowel. In second stage, the language along with Pahari exhibit the process of devoicing, thus the low tone following initial voiced aspirate and [ɦ] became unpredictable (since the phonetically voiceless articulation would predict a high tone for these segments). However, the tones were still traceable from the occurrence of voiced aspirates and [h] in word medial position. In the third stage, the process of deaspiration among (old) voiced aspirated was observed in Punjabi along with Pahari and Kashmiri. The loss of voiced [ɦ] in non-initial position was also observed at this stage, thus resulting in complete unpredictable tonal systems in all positions. In the final stage, a slight phonetic adjustment occurred in Punjabi. The initial voiceless lax aspirated stops which became voiceless lax unaspirated stops (due to the process of deaspiration) [b̤^{h} > b̤] further simplified by the merger of voiceless lax stops with the corresponding voiceless tense stops (Bhatia 1975).

Vijayakrishnan (2009) suggested that, at least three way lexical contrasts are possible in Punjabi, viz., lh, hl, and hlh. For every lexical entry, lh pitch sequence is generated. An H tone is inserted to the left of the pitch melody and finally the final h is promoted to an H tone (hlh). The accent too plays a significant role in determining the tonal specification in Punjabi: the stressed syllable always receives an L tone. Also the leftmost tone is always linked to the stressed syllable. Neither Vijaykrishnan, nor others explained the phonetic reasoning of tonogenesis in Punjabi in an explicit way.

Contrary to the less number of tone languages in the other parts of Indian subcontinent, Northeast India is home to a large number of tone languages. Most of the languages of Tibeto-Burman subgroup of Northeast India are tonal. In has to be noted that this region is also home to some other Indo-Aryan languages such as Assamese, Nepali, and Bangla etc. which are primarily non-tonal. Further, a few languages belonging to the Tibeto-Burman sub family (such as Missing and Garo) are reported to be (phonologically) non-tonal or on the verge of losing (lexical) tonal contrast (Deori for example) (Mahanta *et al.* 2014). The overlapping and constant interaction between these tonal and non-tonal languages of this region presents an interesting perspective on the tonal distribution in the languages based on their geographical location- for example the presence of non-tonal language(s) such as Assamese in the western part of this region reduces the tonal complexity in terms of number of lexical tone and their assignment pattern. Bodo, for example has only two lexical tones (Joseph and Burling 2001, Sarmah 2004, 2009). However, the languages such as Mizo or Angami belonging to the eastern boundaries of North East India (where most of the languages are tonal) are known to have as many as four tones (Lalrindikii 1989, Sarmah 2009, Meyase 2014). Similarly other Tibeto-Burman languages of this area such as Ao (Temsunungsang and Sanyal 2004), Rabha, Dimasa (Sarmah 2009), Kok-Borok (Debnath 2014), are also known to have tonal distinctions. It has to be noted that the language under study has a considerable number of native speakers in the Barak valley region comprising districts such as Cachar, Karimganj and Hailakandi of Assam and the northern parts of Tripura comprising districts such as Dharmanagr, Kailasahar and Kumarghat.

1.2.3 Tonogenesis and the interaction between consonant and tone

Over the last few decades, a considerable number of studies were conducted to understand the mechanisms and motivations of tonogenesis in tonal languages of the world (Matisoff 1970, 1973, Hombert 1978, Haudricourt 1954, Yip 2002, Kingston 2004, Abramson 2004, to name a few). The term 'tonogenesis' was first introduced by James Matisoff (1970, 1973) to refer to the concept of 'birth or rise of the tonal contrast' where there were none before. The concept was further broadened to also include the mechanisms of tone splitting promoting more tones (Abramson 2004).

The best known source of tonal contrast could be traced from the loss of voicing contrast in obstruents (Yip, 2002). While voiced obstruents are known to lower down the pitch of the following vowel (Hombert 1978, Maddieson 1977) voiceless obstruents may even raise it (Hombert *et al.* 1979, Yip 2002). These patterns could be attested to many Tibeto-Burman languages like Karen, Gurung, Austronesian languages like Phang, Rang, Mon-Khmer language like Vietnamese and so on (Thurgood 2002). In a few languages like Yabem and Korean, the tone remains redundant in phonation contrasts in the preceding consonants, whereas the same phonation contrast is realised as tones in languages like (Western) Kammu and (Eastern) Cham (Kingston 2004, 2009). The loss ofvoicing contrast of a preceding stop consonant lead to a low tone in Western Kammu ([bu:c] > [pù:c]) and Eastern Cham. On the other hand, the voiceless fricative [s] [spread glottis] induced a high tone in the Western dialect of Kammu.

Accounts showing a closer relationship between tone and syllable initial consonants could be found in the historically developments of tones in many South East Asian languages. According to Haudricourt (1954, 1961) Vietnamese was toneless at one stage. However, the loss of few consonantal features (distinctions) eventually gave rise to tone in Vietnamese. Historically Vietnamese reported to have three syllable types with voicing distinction in initial stops

- Open syllable (those ending in a vowel or nasal [pa], [ba]),
- Syllables formerly cheeked with a final voiceless spirant that had become [h] ([pas > pah], [bas > bah]), and
- Syllables formerly checked with some kind of stop that had become [ʔ] ([paX > paʔ], [baX > baʔ]). (Haudricourt1954, 1961)

The historical development of Vietnamese witnessed the loss of the final consonant that eventually gave birth to three distinct tones.

Researchers however, slightly vary in their opinions about the overall pitch effect occurring from the postvocalic consonants. Matisoff (1973, 1976), Ohala (1973) and Abramson (2004) speculated that the final glottal stop [ʔ] would result in a raising pitch in the preceding vowel, whereas the final [h] will generate a low pitch in the preceding vowel. Mazuadon (1977), Thurgood (2002), on the other hand claims that a more abrupt and complete glottal stop leads to pitch raising while the more imperfect, less abrupt variant of glottal stop leads to sharp pitch lowering. It is worth mentioning that the phonation contrast in prevocalic consonants may also split the existing tones as could be seen in language families like Sino-Tibetan, Hmong, Tai, Vietic and Kam-Sui families (Kingston 2009).

Loss of voicing contrast among the onset stops (consonants) is also capable of similar tonal contrast. For example, by the end of 12th century, the three tones in Vietnamese (those that occurred due to the loss of coda consonants [h] and [ʔ]) had been developed into six tones due to the loss of voicing contrast among the syllable initial stops (Abramson 2004, Haudricourt (1961)). Haudricourt suggested that final aspirate consonant [h] caused relaxation of the vocal cord and lowered the vibration rate. Further, the higher airflow required for aspirated consonants force the glottis to be wide open. These two factors together led to lower the f_0 in the adjacent vowel. On the other hand, the vibration of the vocal folds is increased for glottal stop [ʔ] due to the tensing of the vocal folds which in turn increase the f_0 of the adjacent vowels. Such correlation is also found in other linguistic groups such as Hottentot (Beach 1938, Hombert *et al.* 1979).

Halle and Steven (1971) proposed a different unification that replaces the feature [voicing] with [stiff] and [slack] obstruents. The height of the larynx in pharynx puts forth horizontal tension of vocal cord, the rate of tension is changed significantly to maintain the voiced/voiceless distinction that affects in f_0 perturbations in the onset of the adjacent vowels. However, experiments show that there is no significant tendency for the voicing distinction in consonants to affect a distinctive f_0 of the preceding vowel. Therefore, it is believed that horizontal or longitudinal tension of vocal folds which is controlled by the cricothyroid muscles controls the f_0 fluctuations in the neighbouring vowels. Stop consonants produced with a [slack] vocal folds [+voice] are likely to lower the f_0 on succeeding vowel, whereas, stops produced with a [stiff] vocal folds [-voice] would result in higher f_0 in the following vowel. While explaining the phonetic realization of consonant tone interaction Kingston and Diehl (1994, 1995), and Kingston *et al.* (2008) argue that the lower f_0 in the vowels next to a voiced stop incorporates perceptually with voicing in the stop closure itself in order to increase the perception of the presence of the low frequency energy in and near the stop. It is the low frequency energy and not the individual voicing or low f_0 that may provide the clue that the stop is [voice]. The larynx is deliberately lowered to uphold the intraoral air pressure from rising rapidly by lowering the f_0 of the vowel next to a voiced obstruent.

However, Hombert and Ladefoged (1977), Homebert *et al.* (1979) pointed out that the degree of voicing of a voiced obstruent does not correlate well with the degree of pitch lowering. Kingston (2009), Thurgood (2002) provide a plausible answer to this. They argue that it is rather various consonant classes (such as distinctive laryngeal gestures) and not the individual consonant that is associated with pitch assignment. For example, voiced obstruent onsets may result in breathy voice, the phonetic by-product of laryngeal gestures that causes a lower pitch which also produces the breathiness. Thus, the voice quality of the laryngeal gesture affects the pitch perturbations. Thurgood (2002) claimed that there are two laryngeal gestures associated with pitch production: one, the downward movement that expand the supraglottal cavity in order to maintain the voicing in obstruents, and two, the tilting of the thyroid cartilage in relation to the cricoid cartilage below it. The cricothyroid muscle is believed to be the main pitch controller since it largely affects the vibrations of the vocal cord. Thus, the laryngeal features are capable of raising or lowering pitch.

We have already discussed that the final glottal stop [ʔ] promotes a raising pitch as in the case in Vietnamese; however, there are also cases where pitch is lowered due to the loss of final glottal stop. For example, the less complete and abrupt glottal stricture accompanying a final glottal stop results in a tense or creakiness in Burmese, and this tenseness in vocal fold in return, produces a sharp low tone (Mazaudon 1977, Thurgood 2002).

Unlike obstruents, the interaction between fricatives and sonorants is way more restricted. It is noticed that the laryngeal articulation in sonorants and fricatives is easily shifted in the beginning of consonantal articulation than in stop consonants as sonorants and fricatives lack a stop burst. These consonants lack any phonetic constraints on laryngeal articulation; therefore, these do not trigger a pitch difference when they occur in syllable final position in Proto-Athabaskan. The stem final glottalic consonants may trigger a high tone (as in the language like Chipewyan) as well as a low tone (as in the language like Gwichín). The contrastive tone could be the result of contrastive laryngeal articulation. Till this point we have noticed that the tones are induced or split on vowels due to loss of phonation contrast among consonants. There are instances of tone splitting at least in two languages (U and Lugbara) due to vowel height and advanced tongue root feature(Kingston 2009). In U, the high level tone in closely related language Hu has split, the high vowels raised the pitch of the syllable with a high vowel; on the other hand, the pitch is lowered when the syllable contains a non-high vowel. In Lugbara, the [+ATR] vowels [i e o u] triggers an extra high tone in a disyllabic verb stem [átsí] whereas [-ATR] vowels [ɪ ɛ a ɔ ʊ] introduce a high tone in a verb stem [áví] 'play'. In case of nouns, high tones occur on [-ATR] vowels only, but extra high tone can occur on both [-ATR] and [+ATR] vowels (Anderson 1986, Kingston 2009).

In this regard it has to be noted that the effects of voiceless aspiration and glottalization on obstruents are not nearly as consistent as those of voicing. Researchers also claim that voiceless aspirated stops often tend to give rise to higher tones than unaspirated stops (Yip, 2002). However, the reverse process is also reported for few languages (Haudricourt 1972, Kingston and Solnit 1989). For example, the voiceless aspirated series are realized in higher pitch than the unaspirated voiced and voiceless stops in the Karen language Palaychi, on the other hand the glottalized series are higher than the voiceless aspirated stops in the language like RenliKam (Yip 2002).

Apart from the effect of consonantal features on tone as mentioned above, the voice quality too can play a decisive role in tonogenesis. Thurgood (2002) argued that three laryngeal configurations produce three basic voice qualities (among the nine suggested by Ladefoged (1971)) which play a significant role in tonogenesis: viz., the breathiness, the clear voice (often regarded as normal or modal) and the tense voice. Breathiness is produced with abducted arytenoid cartilages, thus often lowers the larynx and hence correlates with a low pitch, the tenseness (or creakiness) of the vocal folds often raise the larynx, thus is generally associated with raising pitch. On the other hand, the clear or modal voice is not capable of significant changes of the larynx, hence it is not capable of modifying the pitch to any significant level (Thurgod 2002).

1.3 Layout of the dissertation

The chapters of this dissertation are organized in the following manner-

- **Chapter 1**: *Introduction*- This chapter pinpoints the motivation of the research under consideration. It highlights the brief historical and theoretical background of tone languages of the world and the tone languages of the Indian subcontinent -especially those belonging to Indo-Aryan language family and the languages of the northeast India. The objective and the broad research questions of the dissertation are discussed in this chapter. Finally this chapter talks about the way this dissertation is organised.

- **Chapter 2**: *Sylheti Phonemes*- This chapter demonstrates the phoneme inventory of Sylheti. Results of two acoustics experiments have been discussed in this chapter. The first experiment is meant to explore the status of the underlying breathy voice property [+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.

- **Chapter 3**: *Lexical tones*- description and analysis. This is one of the core chapters of this dissertation. A third production experiment has been conducted on 51 homophones (24 monosyllabic words and 27 disyllabic words) to examine and explore the status of lexical tone in Sylheti. With the help of acoustic parameters (such as f_0 [measured at various points], duration and intensity) along with the support of ample statistical evidence, the phonologization of lexical tone in Sylheti is established. The results clearly establish two way tonal contrasts in Sylheti- in most of the cases a high tone is realized following the loss of underlying breathy voiced contrast among Sylheti obstruents (for both voiced and voiceless). In a few other case where the homophonous pair witnessed similar feature loss such as the loss of underlying breathiness contrast, one of the word in the homophonous pair is realized with a low tone ([xál] 'drain/channel' [k^h > x], [xàl] 'skin' [$tʃ^h$ > x]).

- **Chapter 4**: *Correlation between tone and phonation types and Sylheti tonogenesis*- this chapter investigates the qualities of the vowels carrying contrastive tones. An experiment is conducted on 20 monosyllabic words (comprising the vowel [a]) to explore the phonation qualities of the vowels by exploiting various spectral properties- such as first and second harmonics H1 and H2 respectively, the difference in the amplitude of the first and second harmonic (H1-H2), the difference in the amplitude of the second and fourth harmonic (H2-H4), first formant bandwidth (H1-A1), second formant bandwidth (H1-A2), and the overall spectral tilt (H1-A3). The findings revealed that the vowels carrying contrastive tones are (significantly) different from each other in terms of their voice quality- vowels bearing low tones have slightly but consistently greater spectral tilt than the vowels bearing high tones. We conclude that the vowels bearing low tones have slightly 'laxer' voice qualities than the vowels bearing high tones which are 'tenser' in nature.

- **Chapter 5**: *Perception of Sylheti tones*- this chapter discusses the way lexical tone is perceived by the native speakers. For that purpose, an experiment has been conducted using synthesized stimuli. The stimuli were prepared by varying (shifted up or down) the pitch contours of the contrastive tones at every 20% of the rhyme and were embedded in a natural sentence frame. Subjects were asked to listen and respond to each item and identify the meaning (embedded with the stimuli) from three possible options- the real meaning, the contrastive meaning and third option being NOT SURE. Results indicate that Sylheti native speakers perceive tone in a categorical manner- f_0 fluctuations provide the most important cue and this fluctuation must continue (at least) till the first 60% of the total rhyme/voicing part.

- **Chapter 6**: *Morpho-phonemics in Sylheti*- this chapter examines the complexities of tone assignment in morphologically derived environments. Suffixes such as inanimate plural marker and definite classifiers are attached to (common) noun root (underived monosyllable and derived disyllables) carrying contrastive tones. Similarly various verbal suffixes such as (tense aspect markers) are attached after various verb forms with contrastive (underlying) tones. Results of this production experiment indicate the presence of tonal complexities in Sylheti. The suffixes attached to noun roots are toneless and takes the tone opposite to the roots they are attached to; thus displaying a case of tonal polarity. The verbal suffixes, on the other hand bears an (underlying) high tone, and displays an instance of tone reversal; i.e., when added with a verb form with underlying high tone, the surface tone of the verb form alters to low tone after suffixation.

- **Chapter 7**: *Conclusion*- the final chapter of the dissertation, summarises the findings of all the chapters discussed in this dissertation. Subsequently the limitations and further scope of future work have also been highlighted in this chapter.

Chapter 2

Language Description: Sylheti

General Background: Bangla and its Dialects

Genetically Bangla is derived from Indo-Aryan (IA) or the Indic sub branch of the Indo-Iranian branch of the Indo-European (IE) language family (Chatterjee, 1926). The language is considered as one of the world's densely inhabited languages with more than 170 million first-language speakers across the world (Gordon 2005). Apart from where it is the national language, the Bangla speaking zones in India includes parts of Bihar, Orissa, Assam and the whole of West Bengal and Tripura within India. Substantial Bangla inhabited pockets are also found in the Middle-East, Europe and the U.S, north-western Burma (also known as Myanmar), parts of Nepal (in the regions such as Mechi Zone, Jhapa district, Kosi Zone, Morang and Sunsari districts and so on). In India, Bangla is the second most-spoken language (behind Hindi and Urdu) and is considered as an official language of the state of West Bengal and the co-official language in the state of Tripura and the union territory of Andaman and Nicobar Islands.

Naturally, the wide spread of this language over a vast region witnessed various historical developments and thus formed numerous dialects of Bangla. Many of these dialects appear to be unintelligible due to their wide spread over a large continuum. Grierson (1928) while classifying the speech varieties spoken in the province of Bengal divided Bengali into two major branches: the Eastern Branch and the Western Branch (ignoring the geographical and/or political boundary). The Western branch can further be divided into three different types, viz. Central Bengali (Standard Bengali: includes the varieties of Kolkata, Howrah in India), Northern Bengali (includes the varieties spoken in East Malda, Koch-Bihar of West Bengal in India and Rajshahi, Dinajpur, Bograand in Bangladesh),and Western Bengali (includes varieties such as Kharia, Thar, Mal Paharia, Sarkietc). The Eastern Branch on the other hand, could further be divided into four different groups; viz., Eastern Bengali (includes the varieties spoken in districts such as Dhaka, Mymensingh, Comilla, Sylhet, Hajong of Bangladesh and the Cachar Region of Assam in India), East-Central Bengali (includes the varieties spoken in Jessore, Khulna, Faridpur districts of Bangladesh), South-Eastern Bengali (varieties include Noakhali, Chittagong, Chakma, Tangchangya etc) and Rajbanshi (varieties spoken in Rongpur districts of Bangladesh, and Siripuria, Jalpaiguri in the state of West Bengal and Goalpara of Assam in India).

Chatterjee (1926) however, categorized the varieties of Bangla spoken in the region of greater Bengal into four major clusters: Radha, Varendra,Kamrupa and Vanga.The region of greater Bengal does not consider the political boundary and thus comprises the varieties of Bangla spoken in entire Bangladesh and Indian states of West Bengal and Tripura, and parts of Assam, Bihar and Jharkhand. This dissertation aims at studying the linguistic development of Sylheti in general and tone in particular. Chatterjee (1926) clubbed Sylheti under the subcategory of Eastern and South-Eastern Vanga under the Vanga cluster (Fig 2-1).

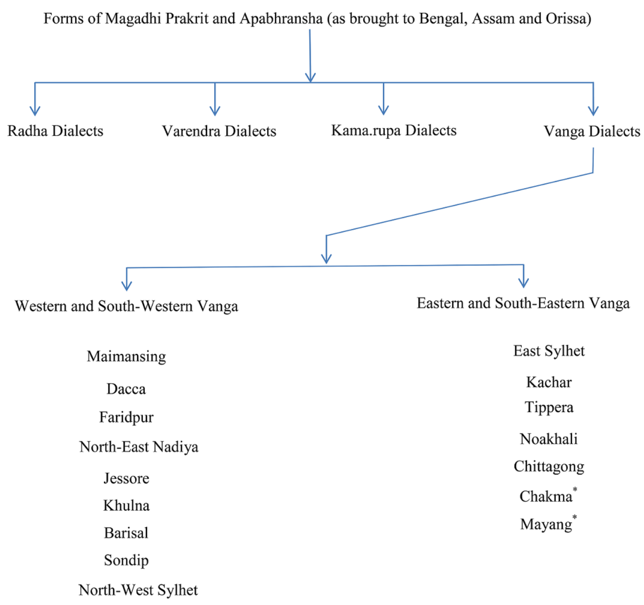

Figure 2.1: Varieties of Bangla classified by Chatterjee (1926)

The languages with an asterisk mark (*) (viz., Chakma and Mayang) reported to be mixed up with the languages of Tibeto-Burman group of languages. While Chakma is considered to be a dialect of hill people of Chittagong (in Bangladesh) and parts of Tripura in India, Mayang or Bishnupriya is spoken by a sub-section in Manipur (Chatterjee 1926). All these dialects differ widely in all aspects of the grammar, especially in terms phoneme inventory, allophony, inflectional morphology and lexicon, thus making many dialects largely unintelligible. For this reason, several dialects such as Chakma, Hajong, KhariaThar, Mal Paharia, Sylheti, Tangchangya, Noakhaliare often considered to be separate languages (Grierson 1928, Chatterjee 1926, Gordon, 2005).

2.1 The language under study: Sylheti

As shown above (Figure 2-1), Sylheti is generally regarded as one of the varieties of eastern Bangla. There are approximately 10,300,000 people using Sylheti as their first language (7,000,000 in Bangladesh) (Lewis *et al.* 2013). The name of this language has been derived from the place where it is being typically spoken; i.e., the Sylhet District (also known as Surma Valley including districts such as Sylhet, Habiganj, Maulvi Bazar, and Sunamganj) located in the north-eastern region of Bangladesh. In India, a considerable amount of Sylheti speakers inhabit the northeast state of Assam (the Barak Valley region comprising districts such as Cachar, Karimganj and Hailakandi) and Tripura (mostly the northern districts such as Dharmanagar, Kailasahar and Kumarghat). At present Sylheti exhibits two divisions; mostly due to the speakers' religious affinity. The Islamic followers (especially those from Bangladesh and Karimgang and Hailakandi districts of Assam in India) use a lot of borrowed words and phrases from Persian and Arabic (including many of the kinship terms). The Sylheti spoken by the Hindus of Cachar district of Assam and North Tripura has somehow kept their original form intact. Data for this dissertation were collected mostly from the Dharmanagar district of north Tripura and partially from the Cachar district of Assam.

Interestingly, Sylheti had a distinct script in Syloti-Nagari (better known as Nagori) interrelated with *Kaithi alphabet*. This script appears to be very different from the Bangla script. The exact origin of this script of Sylheti is unknown and believed to be developed around the beginning of 14th century (Chatterjee 1926). Books written in Nagori (manuscripts or printed) are known as *Puthis*. The main purpose of this writing system was to record religious poetry.

The distinct phonetic and phonological property distinguishes this variety from Standard Colloquial Bangla (henceforth SCB). One of the properties that distinguish Sylheti from SCB or other regional varieties is the significant 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 languages belonging to Indo-Aryan language family are not generally recognized for tone [Punjabi is known for two levels of tone; high and low (Bahl, 1957; Chatterji, 1969)]. The literature on tonogenesis reveals that the loss of one or the other laryngeal features such as voicing contrast, aspiration or breathiness contrast (also called murmur voicing), glottalization, etc. of adjacent consonants are the usual sources of tonal contrast (Yip 2002, Thurgood 2002, Kingston 2009).Interestingly as mentioned above, the language under study exhibited significant reduction and restructuring of its phoneme inventory. 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[1] of voiced ([b^{h}ai > **b**ai] 'brother', [lab^{h} > la**b**] 'profit') and voiceless stops ([t̪^{h}ala > **t̪**ala] 'plate', [lat̪^{h}i > la**t̪**i] 'kick'). Further, the lenition process involving spirantization targeting the underlying voiceless bilabial stop (both aspirated and unaspirated) [p] and [p^{h}] ([por] > [**ɸ**ɔr] 'read', [p^{h}ul] > [ɸul] 'flower') and voiceless velar stop (both aspirated and unaspirated) [k] and [k^{h}] ([kali] > [**x**ali] 'ink', [k^{h}al > **x**al] 'drain/channel') and 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') reduced and restructured the phoneme inventory of the language. These processes led to the birth of numerous homophones and prompted us to investigate the tonal property of this language. It has to be noted that, till date no studies have been conducted on Sylheti at any (linguistic) level in general, and tone in particular.

This chapter discusses the phoneme inventory of Sylheti. To understand the status of the feature [+spread glottis], we conducted an acoustic experiment and measured the durational properties of Sylheti voiced stops using the concept of voice onset time (henceforth VOT), proposed by Lisker and Abramson (1964). We speculated that these voiced stops would differ only in terms of their place of articulation. The lenition process of spirantization and deaffrication is discussed with the help of spectrograms. Further, to provide a comprehensive account of Sylheti vowels, we conducted another acoustic experiment on Sylheti monophthongs (section 2.2.4). Various statistical analyses were performed on different acoustic properties such as duration, first three formants to confirm the types of vowels and gender effects. Our findings confirmed that there

[1] The concept of '*deaspiration*' as mentioned here and elsewhere in this dissertation denotes the idea of '*loss of breathiness property [+spread glottis]*' (phonological) of Sylheti obstruents. The properties of (breathy) aspirated voiced stops available in Standard Colloquial Bangla are entirely different from those currently available in Sylheti (loss of the breathiness property during the historical development of this language) as the two varieties are geographically segregated with many other dialects and languages occurring between the two.

are five (oral) vowels in Sylheti as compared to seven in SCB.

2.2 Phoneme inventory of Sylheti

Chatterjee (1926) recorded twenty nine consonantal sounds and fourteen vowel phonemes of which seven are nasalised in SCB. The language maintains a four way contrast of voicing and aspiration among the stops. Of the twenty nine consonants ten are aspirated stops [p^h], [b^h], [t̪^{h}], [d̪^{h}], [t^h], [d^h], [k^h], [g^h], [dʒ^h] and [tʃ^h]. The stop consonants also have their unaspirated counter parts. There are six sonorants consisting of three liquids [l, r, ɾ] and three nasals [m, n, ŋ]. As discussed above, Sylheti underwent drastic phonological changes resulting in large scale reduction in the overall number of phonemes and stops in particular. Assuming SCB as the base of roots, we will argue that all the aspirated stops in Sylheti lost their underlying breathiness property [+spread glottis] due to the phonological process of obstruent weakening. Das (1996) also showed a similar reduction process of obstruent weakening in Noakhali spoken in Southern Tripura.

Since the obstruents in Sylheti Bangla are the most affected ones, we decided to provide a detailed analysis of the obstruent inventory. We observed only twelve stops in Sylheti.

2.2.1 Loss of underlying breathiness property [+spread glottis] of voiced aspirated Obstruents

The entire set of voiced aspirated stops [b^h, d̪^{h}, d^h, g^h, dʒ^h] lost the underlying breathing breathiness property in all the positions of occurrence (viz., word initially, medially, as well as word finally) in Sylheti. The loss of the phonological feature [+spread glottis] could be seen in the following examples:

a) [bʰ] ⟶ [**b**]

Initial	Gloss	Medial	Gloss	Final	Gloss
[**b**ai]	'brother'	[na.**b**i]	'naval'	[la**b**]	'profit'
[**b**ab]	'affection/pretend'	[ɔ.**b**ab]	'poverty'	[lu**b**]	'greed'

b) [d̪ʰ] ⟶ [**d̪**]

Initial	Gloss	Medial	Gloss	Final	Gloss
[**d̪**an]	'paddy'	[ga.**d̪**a]	'donkey'	[d̪u**d̪**]	'milk'
[**d̪**ɔn]	'wealth'	[a.**d̪**a]	'ginger'	[ʃu**d̪**]	'repayment'

c) [ɖʰ] ⟶ [**ɖ**] [the consonant [ɖʰ] occurs only in the initial position of a word]

Initial	Gloss
[**ɖ**ax[2]]	'drum'
[**ɖ**a.lu]	'sloping'

d) [gʰ] ⟶ [**g**]

Initial	Gloss	Medial	Gloss	Final	Gloss
[**g**um]	'sleep'	[mɔ.**g**a]	'inauspicious star'	[ba**g**]	'tiger'
[**g**ɔr]	'house'	[rɔ.**g**u]	'name'	[ma**g**]	'name of bengali month'

[2]notice the word-final voiceless stop [k] is also spirantized to [x] ([**ɖʰ**ak > **ɖ**ax]).

2.2.2 Loss of underlying breathiness property [+spread glottis] of Voiceless Aspirated Obstruents

Similarly voiceless obstruents also display the loss of the feature [+spread glottis] in Sylheti. The entire set of voiceless aspirated obstruents [pʰ, t̪ʰ, tʰ, tʃʰ, kʰ] lose the underlying breathiness property and thus fall together with the already existing set of corresponding sounds with the feature specifications of [-son, +cons, -voice]. Interestingly,the underlying voiceless bilabial stop (both aspirated and unaspirated) [p] and [pʰ] and the underlying voiceless velar stop (both aspirated and unaspirated) [k] and [kʰ]are the most affected sounds. Not only these sounds lose their underlying breathiness property, these are further spirantized to [ɸ] and [x] respectively. The details of these patterns have been discussed in section 2.2.5. Let us consider the following examples:

a) [t̪ʰ] $\longrightarrow$ [t̪]

Initial	Gloss	Medial	Gloss	Final	Gloss
[t̪a.la]	'a plate'	[ma.t̪a]	'head'	[rɔt̪]	'chariot'
[t̪a.ka]	'to stay'	[la.t̪i]	'kick'	[nɔt̪]	'nose-ring'

b) [tʰ] $\longrightarrow$ [**t**]

Initial	Gloss	Medial	Gloss	Final	Gloss
[**t**ik[3]]	'exact'	[la. **t**i]	'stick'	[ma **t**]	'field'
[**t**a.x[4]ur]	'deity'	[x[5]a. **t**al]	'jackfruit'	[x[6]a. **t**]	'wooden piece'

One major goal of this dissertation is to examine how the loss of the feature [+spread glottis] among the Sylheti obstruents have affected the phonological and/or phonetic property of this language. The historical development of this language resulted in

[3] [k] is an allophone of [x] which occurs only when preceded by high vowels [i] and [u] (x $\longrightarrow$ k[+high, -consonantal] _).

[4] Spiratization [k] $\longrightarrow$ [x].

[5] [k] $\longrightarrow$ [x].

[6] [k] $\longrightarrow$ [x].

a reduced and restructured phoneme inventory therefore requires to be investigated carefully. To examine the properties of stop consonants and to understand the nature of the (underlying) feature [+spread glottis] (such as the outcome of the loss of the underlying feature [+spread glottis] on the following vowels), we first conducted a production experiment and measured the voiced onset time (VOT) of all the voiced stops[7] following the theory proposed by Lisker and Abramson 1976). Since the voiceless stop series, by and large, is the most affected ones (involving two-way loss viz., spirantization and also the loss of underlying breathiness property [+spread glottis], [p, p^{h} $\longrightarrow$ ɸ] and [k, k^{h} $\longrightarrow$ x]), and also partially because we did not observe any difference in terms of their VOT (voicing lag as proposed by Lisker and Abramson 1976) between the pairs of underlying aspirated and non-aspirated series of the two voiceless stops [**t̪**] and [**t**], we decided to concentrate on the voiced series instead. This was also done because of the fact that the onset of most of the homophonous-like pairs considered for f_0 analysis (mostly) involved voiced stops. The phonation qualities of the vowels adjacent to the homophonous pairs of stop consonants (with and without underlying breathiness property) are discussed in chapter four of this dissertation.

2.2.3 Acoustic analysis of stop consonants

To understand how the remnant feature [+spread glottis] may have affected consonants of Sylheti, we measured the voice onset time of each of the voiced stops present in the language. Abramson (1977) argued that voice onset time (VOT) could be considered as the most reliable acoustic cue to be able to draw a distinction between the voicing categories (+ voice versus - voice). In general, the presence or absence of glottal buzz during the oral closure is assumed to be the relevant indicator of voicing. For example,

[7] In chapter one, we have specifically mentioned that the principle thrust of this dissertation is to investigate the nature of tonal properties (if any) possibly borne out due to the loss of the feature [+spread glottis]. The finding of this experiment (determining the status of the feature [+spread glottis]), therefore, is intended to provide supporting evidences towards that goal. The scope of investigating the intrinsic phonetic variations involving place of articulations and voicing contrast in much detail is thus kept open for future investigations. The outcome of loss of the underlying feature [+spread glottis] is thoroughly discussed in chapter 3 and chapter 4.

in a language like German, voiced stops in initial prevocalic and prestressed position exhibit a relatively short VOT (negative) and show no or little presence of aspiration, while voiceless stops have a relatively long VOT (positive) and are indicative of the presence of strong aspiration (Lisker 1986, Docherty 1992, Mikuteit 2009). With the help of data from 11 different languages, Lisker and Abramson (1964) claimed that the single dimension of VOT or duration measurement of the time lag between the start of voicing and the start of the consonant featuring stop consonants (those which occur word-initially) is sufficient to be able to distinguish the voicing contrasts in most of those languages.

This temporal characteristic of stop consonants reflects the complex timing of supralaryngeal laryngeal coordination. VOT therefore, is defined in relation to the time difference between the release of closure for a stop consonant and the onset of vocal fold vibration. Voicing detected before the release or during stop occlusion is called the voicing lead, while voicing starting after the release is called voicing lag. VOT thus stands for the temporal relation between the onset of glottal pulsing and the release of the initial stop consonant, thus making the stop release as the measuring reference point (Lisker and Abramson, 1964, 1967). A negative value is attributed to the VOT [+voice]if the onset of voicing is perceived before the release and a positive value is assigned if the voicing onset takes place after the release [-voice]. Thus it is generally assumed that, the voiceless unaspirated stops is attributed the highest and positive VOT, their aspirated counter parts will have a relatively less and positive VOT, and the voiced unaspirated stops are associated with negative and relatively shorter VOT, and the voiced aspirated do have the lowest and negative VOT.

While the postulations derived from the VOT theory is capable of successfully distinguishing the voicing contrast and the contrast between voiceless aspirated stops with their unaspirated counterparts (thus, appears to be three dimentional), the theory is insufficient in accounting for the voiced aspirates (Lisker and Abramson 1964, Lombardi 1994, Dutta 2007, Mikuteit 2009). This constraint surfaces due to the fact that the clue ‘voicing lead’ (-VOT) employed for both voiced aspirated stops and voiced stops appears to be inadequate for making a distinction between these two stop types.

To overcome this problem, Lisker and Abramson (1964) proposed to consider the low amplitude buzz combined with noise in the interval following the release as the vital factor to distinguish these two voiced categories. In our data of Sylheti voiced phonemes (with and without underlying breathiness property), we did not observe any low amplitude buzz and/or noise like signal.

Davis (1994), however, claimed that the lag time differences of 'noise offset' measurements (measurement taken from the release burst to the onset of the second formant), of all four velar stops in Hindi ([k, k^h, g, g^h] is capable of successfully distinguishing the voiced unaspirated stops from the voiced stops. In her studies she observed a significantly longer lag time for voiced stops than their voiced aspirated counterparts. On the other hand,many researchers such as Schiefer (1986), Lombardi (1994), Dixit (1989), Yadav (1984), Ladefoged (1971), Dutta (2007) debated on the inadequacy of the VOT theory especially for the voiced aspirated stops and suggested additional acoustic measurement techniques for successfully distinguishing the voiced aspirated stops from their unaspirated counterparts. While on one hand, Schiefer (1992) proposes the notion of 'Voicing Lead Time' (to be measured depending on the period of voicing during closure), Ladefoged (1971) on the other hand, proposed the concept of independent mode of relationship between the voiced aspirated stops and phonation types (such as breathy, modal and murmured) of the adjacent vowels.

Dutta (2007) argued that the four way stop distinction (voicing [+/-], and aspiration [+/-]) in Hindi stops could well be analysed as a cumulative effect of several acoustic cues such as the effect of f_0 perturbation (of neighbouring vowel) following the stops. While analysing the stop categories of Hindi,he observed that both voiced aspirated stops and their unaspirated counterparts are stronger f_0 depressor[8]; voiced aspirated

[8]Contrary to the findings of Dutta (2007) we will argue that the loss of the underlying breathiness contrast [+spread glottis] among the Sylheti obstruents indeed raised the f_0 of the neighbouring vowels, creating a two way tonal contrast in Sylheti (discussed in Chapter 3). Thus Sylheti presents a special case of tonogenesis where a high tone is generated (largely) due to loss of underlying breathing property of stop consonants (as observed in both voiced and voiceless stops). Following the findings of Dutta (2007), and the findings of VOT analysis discussed in this chapter, we will argue that voiced stops in Sylheti do contrast only in terms of their place of articulation.

stops seem to be lowering the f_0 even further. Keeping in mind all the theoretical issues discussed above, we conducted a production experiment on all the possible voiced stops in Sylheti as discussed in the section 2.2.1.

2.2.4 Experimental Procedure

Stimuli

For the acoustic experiment, 18 different words representing each of the target voiced stops occurring word initially were carefully chosen(Table 2-1). As could be seen from the dataset given in the Table 2-1, 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 contrastive pairs (viz., [**b**at̪] 'rice', [**b**an] 'pretend', [**b**ala] 'good' and [**b**ari] 'heavy') representing the (underlying) aspirated voiced stop [$\mathbf{b^h}$] had underlying breathiness property at some point of history, and appeared to be 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 [d] 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.

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

Table 2.1: Dataset considered for VOT experiment

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. Apart from Sylheti, the speakers were also fluent in Hindi and English. The age of the participants ranged between 18 and 42 years. None of speakers had any history of speech disorders.

Speech data was recorded with a *Shure* unidirectional head-worn microphone connected to a *Tascam* linear PCM recorder (ensuring 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 each individual word. Subjects were asked to pronounce each individual word first at normal utterance speech with preferably natural intonation. To avoid the effect of any neighbouring sounds, each word was recorded individually with a considerable amount of gap between two words. The experimenter constantly monitored the whole recording procedure and ensured that each subject understood the task correctly. Apart from the target stimuli mentioned in Table 2-1, 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 each.

Acoustic Measurement

Dutta (2007) claimed that voice lead time (VLT) along with the f_0 measurements of the following vowel can successfully draw a distinction between voiced aspirated and unaspirated counterparts (in Hindi); where, in case of VLT, voiced aspirated stops will be shorter and hence will reduce the f_0 of the following vowel the most than the unaspirated voiced stop. He also observed that the breathy portion following the voiced aspirated stop incline to infiltrate into a considerable portion of the vowel, hence he suggested to consider f_0 of the initial portion of the vowel (instead of the whole f_0) where a lowering in f_0 of the following vowel could be observed. Contrary to that, in chapter 3 of this dissertation we will show that percentage wise pitch measurement[9] of

[9]measurement was taken at every 10% of the total duration of the vowel.

the vowel following the stop consonants those are specified with (underlying) breathiness property (and subsequently lost that property in course of historical development of this language) resulted in a higher f_0 instead which continued to be (almost the) same till the end.

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 2-2, and Figure 2-3). 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.

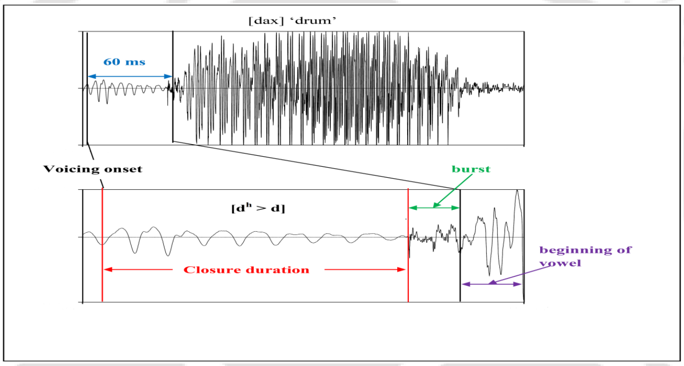

Figure 2.2: Waveform of Sylheti words [dax] 'drum' [d^h> d]; the enlarged portion of the target phoneme considered for VOT measurement is shown along with.

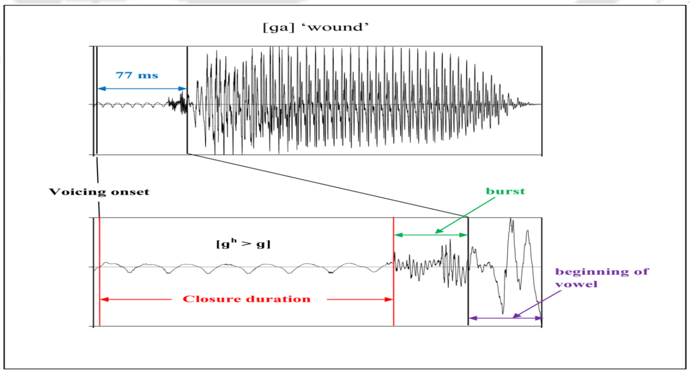

Figure 2.3: Waveform of Sylheti words [ga] 'wound' [g^h > g]; the enlarged portion of the target phoneme considered for VOT measurement is shown along with.

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 problem external noise, or if the token was disturbed due to individual speaker's starting problem. 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 aspirated stops did not show any difference from their unaspirated counterparts (Table 2-2). The duration values of individual voiced stops are shown in Figure 2-4.

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	14.314	1	14.314	.095	.757
Within Groups	53665.895	358	149.905		
Total	53680.209	359			

Table 2.2: Statistical analysis of variance among Sylheti voiced categories.

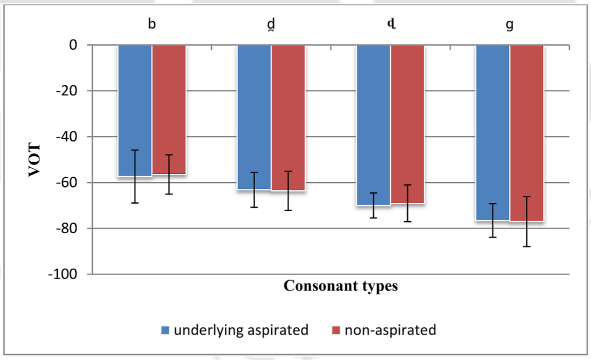

Figure 2.4: Average duration of the voiced stop consonants with standard deviation as error bars.

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 2-3). 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

[d] which could be due to less number of tokens considered for the statistical analysis (the consonant [d̪] has only 22 tokens). As assumed no significant interaction was observed between individual pair the unaspirated and (underlying) aspirated stops. No significant interaction was also observed between [b] and [d̪], however, the underlying aspirated [d̪] and [b] significantly differ from each other where the former was found to be significantly shorter than the latter. Similarly, underlying aspirated [b] and unaspirated dental [d̪] and underlying aspirate [d̪^{h}] is found to be significantly different from each other.

Consonant Types	N	b	b^{h}	d̪	d̪^{h}	ɖ	ɖ^{h}	ɡ
b^{h}	80	!						
d̪	22	!	*					
d̪^{h}	22	*	*	!				
ɖ	36	*	*	!	!			
ɖ^{h}	34	*	*	!	!	!		
ɡ	43	*	*	*	*	*	*	
ɡ^{h}	43	*	*	*	*	*	*	!

Table 2.3: Significant matrix for duration of Sylheti stop consonants are presented in above table. An asterisk [*] symbols represents the significant while [!] represents the non-significant pairs.

2.2.5 The process of Lenition

The lenition process of spirantization and deaffrication (along with the loss of underlying breathiness property) further reduced and restructured the phoneme inventory of Sylheti. With the help of waveforms, these changes are discussed below.

Spirantization

The underlying forms of voiceless bilabial stop (both aspirated and unaspirated) [p] and [p^{h}] and voiceless velar stop (both aspirated and unaspirated) [k] and [k^{h}] have been sprantized due to the phonological process of consonant weakening. Consider the following examples:

a) [p] ⟶ [ɸ]

Initial	Gloss	Medial	Gloss	Final	Gloss
[ɸan]	'betel'	[ru.ɸa]	'silver'	[ʃaɸ]	'snake'
[ɸɔr]	'read'	[ʃaɸ.ta]	'week'	[baɸ]	'father'

b) [pʰ] ⟶ [ɸ]

Initial	Gloss	Medial	Gloss	Final	Gloss
[ɸul]	'flower'	[t̪u.ɸan]	'storm'	[maɸ]	'forgive'
[ɸɔl]	'fruit'	[bɔr.ɸi]	'plum'	[laɸ]	'leap/jump'

c) [k] ⟶ [**x**]

Initial	Gloss	Medial	Gloss	Final	Gloss
[**x**ali]	'ink'	[bɛ.**x**a]	'curve'	[xa**x**]	'crow'
[**x**ɔ.la]	'banana'	[dɛ.**x**a]	'calf'	[bu**x**]	'chest'

d) [kʰ] ⟶ [**x**]

Initial	Gloss	Medial	Gloss	Final	Gloss
[**x**a.li]	'empty'	[ma.**x**on]	'butter'	[mu**x**]	'mouth'
[**x**up.ri]	'smaill hut'	[rɛ.**x**a]	'line'	[d̪u**x**]	'pain'

The transformation of these stops to fricatives has been shown in Figure 2-5 and 2-6. The enlarged portion of the target phoneme is also shown.

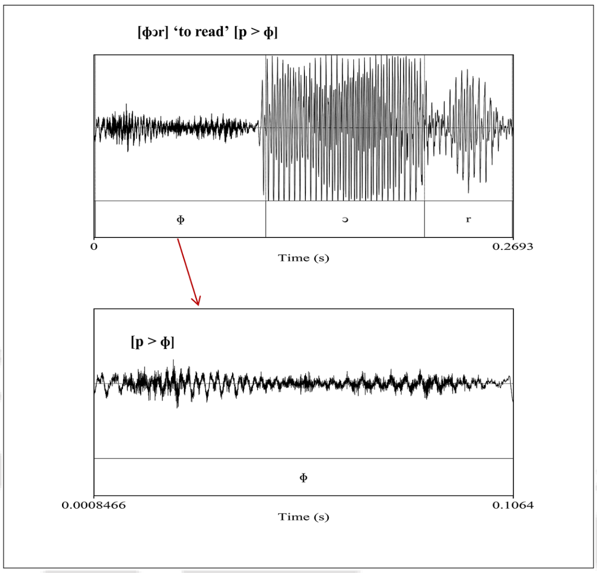

Figure 2.5: Waveform display of the phoneme [ɸ] as captured in the word [ɸɔr] 'read'.

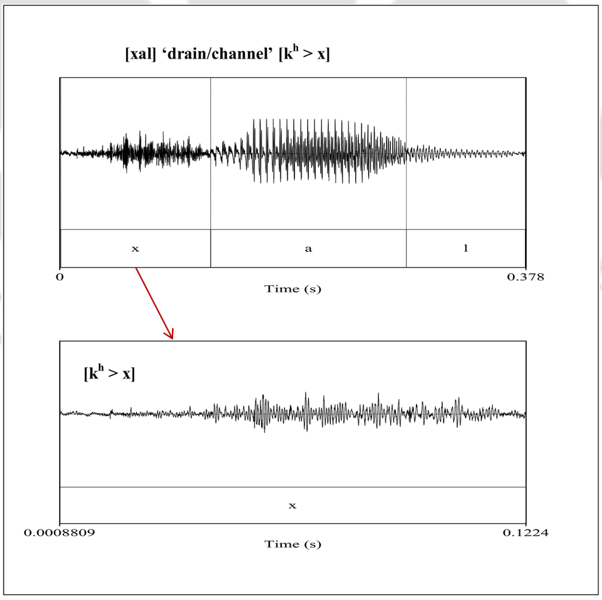

Figure 2.6: Waveform display of the phoneme [**x**] as captured in the word [**x**al] 'drain/channel'. The enlarged portion of the target phoneme is also shown.

Deaffrication

The post-alveolar affricates (both aspirated and unaspirated, viz. [tʃ], [tʃʰ], [dʒ] and [dʒʰ]) also exhibit the process of deaffrication and change to alveolar fricatives [s] and [z] respectively. Let us consider the following examples:

a) [tʃ] ⟶ [s]

Initial	Gloss	Medial	Gloss	Final	Gloss
[sa]	'tea'	[xa.sa]	'raw'	[xas]	'glass'
[sal]	'rice'	[ɸe.sa]	'owl'	[ɸas]	'five'

b) [tʃʰ] ⟶ [s]

Initial	Gloss	Medial	Gloss	Final	Gloss
[sa.gi]	'she goat'	[ma.si]	'fly'	[mas]	'fish'
[sad̪]	'roof'	[bis.na]	'bed'	[gas]	'tree'

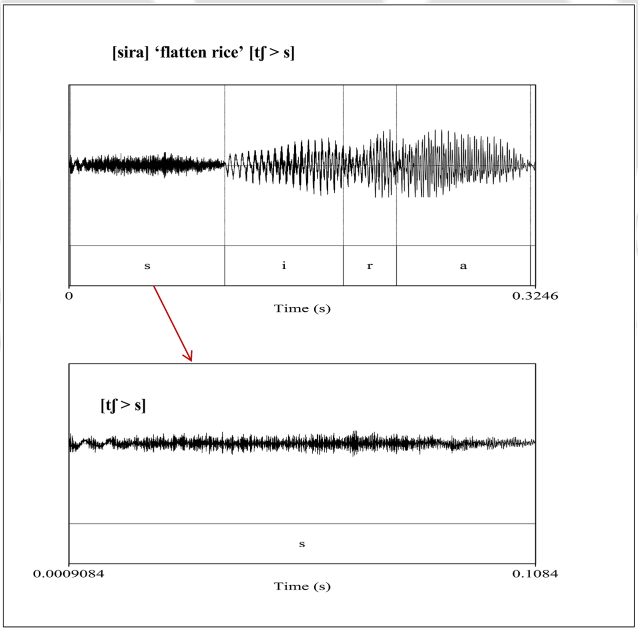

Figure 2.7: Waveform display of the phoneme [s] as captured in the word [**sira**] 'flatten rice'. The enlarged portion of the target phoneme is also shown.

c) [dʒ] ⟶ [**z**]

Initial	Gloss	Medial	Gloss	Final	Gloss
[**z**al]	'net'	[ba**z**.na]	'instrument'	[ba**z**]	'fish'
[**z**at̪]	'caste'	[ba**z**.ar]	'market'	[ʃa**z**]	'make-up'

d) [dʒ^{h}] ⟶ [**z**]

Initial	Gloss	Medial	Gloss	Final	Gloss
[**z**al]	'hot'	[ma.**z**i]	'boatman'	N/A	
[**z**ap.ʃa]	'blur'	[ma.**z**a.ri]	'medium'	N/A	

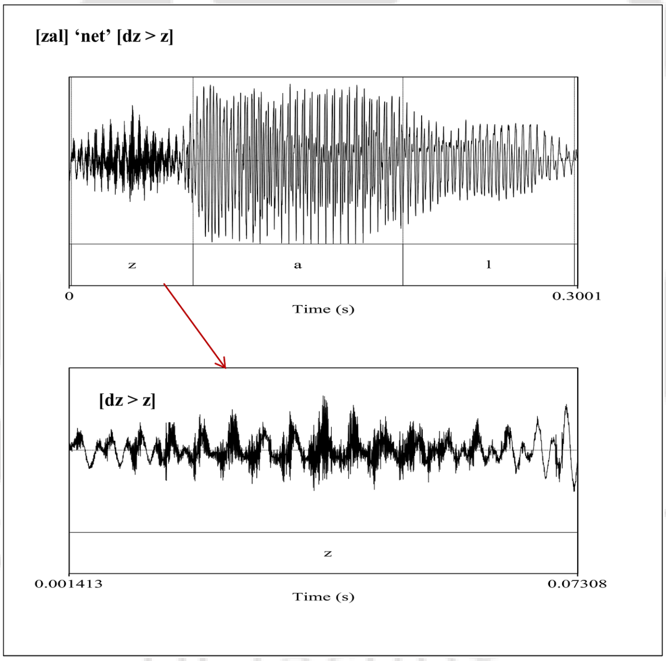

Figure 2.8: Waveform display of the phoneme [**z**] as captured in the word [**zal**] 'net'. The enlarged portion of the target phoneme is also shown.

We have seen the way consonants especially the stops and affricates in Sylheti exhibit a drastic change and thus resulted in reducing the total number of consonant phonemes. Table 2-4 represents the consonant inventories 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				

Table 2.4: Consonant inventory of Sylheti

2.3 Vowels in Sylheti

From our initial fieldwork on Sylheti, we got the impression that not only the consonant inventory but also the number of vowels in Sylheti have got reduced and restructured when compared to SCB. There are seven oral vowels in SCB (viz., [i], [e], [æ], [a], [u], [o] and [ɔ]) (Chatterjee 1926, Bhattacharya 1910). Recordings from our pilot studies gave us an impression 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], [ɛ], [a], [ɔ] and [u] . To validate the number of vowels and to provide a detailed acoustic analysis of the vowels, we conducted another production experiment and attempted to understand the vowel system of Sylheti. The methodology of the acoustic experiment and the findings are discussed in the subsequent sections.

2.3.1 Methodology

Stimuli, speakers and recording procedure

After consulting two of our primary informants, a dataset containing all the possible vowels were prepared. Four different consonantal contexts such as [bVl], [sVɸ], [xVl] and [zVl] (V being the target vowel) were prepared as stimuli. All the target vowels (embedded in different consonantal contexts) were recorded using 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 2-5). Two of the target words (comprising the target vowel) were disyllabic (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.

Sylheti words	Gloss	Sylheti words	Gloss
[bil]	'pond'	[kil]	'fist'
[bɛl]	'wood-apple'	[xɛ]	'who'
[bala]	'bracelet'	[xal]	'skin'
[bɔl]	'ball'	[xɔl]	'tube-well
[bulbuli]	'name of a bird'	[kul]	'lap'
[zil]	'lake'	[siɸ]	'press'
[zɛl]	'prison'	[sɛɸ]	'spit'
[zala]	'body'	[saɸ]	'pressure'
[zɔl]	'water'	[sɔɸ]	'a type of snack'
[zul]	'gravy'	[suɸ]	'quiet'

Table 2.5: Dataset prepared for acoustic analysis of the vowels.

All those words (as shown in Table 2-5) were randomized and additional thirty words were also used as fillers. Subjects were asked to produce those words (embedded in the fixed sentence frame) as naturally as possible. The list was prepared in such a way that each of the target stimuli (along with the fillers) embedded in the fixed sentence frame were repeated thrice.

Seven (5 male, 2 female) out of the eight speakers who participated in the first production experiment (on VOT) participated in the current experiment.Apart from these seven native speakers three more female speaker belonging to the same age group from the Cachar district of Assam, India were also included in this experiment, thus making the total number of speakers ten (five male, five female). Recordings were done in the same way as mentioned in section 2.2.4.2 of this chapter. A total of 600 vowel tokens were considered for acoustic and statistical analysis.

Acoustic and statistical analysis of Sylheti vowels

The start and the end time of each token (of the target vowels) were manually labelled using Praat 5.2 (Boersma and Weenick, 2013). Labels were placed at zero crossings at the onset consonant till the end of the vowel by ensuring the surrounding speech sounds were not audible in the remaining signal. The frequencies of the first three formants (F1, F2 and F3) of each monophthong were calculated at vowel midpoint to avoid possible onset consonantal effect.Further, the Hertz values were transformed to Mel using Praat's inbuilt function HertzToMel. The extracted formant values calculated in Mel (in dB) were further normalized for speaker variations using the Lobanov normalization method in NORM (Thomas and Kendell, 2007). As speculated, the results confirm the presence of only 5 vowels (monophthongs) in Sylheti (see Figure 2-10). To confirm the vowel types differences and gender effects, we conducted one-way Analysis of Variance (ANOVA) tests keeping vowel types/gender as factor and duration and formant values (F1, F2, F3 calculated in Mel) as dependent variables. We used SPSS (version 20) to test the statistical differences of the analyzed data.

Results and discussion

Duration

Figure 2-9 shows the average duration of each vowel drawn separately for male and female speakers. Result suggests that on average vowel durations do differ from each other. A one way ANNOVA conducted on the data from all the speakers confirmed significant effect on vowel quality in terms of 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.

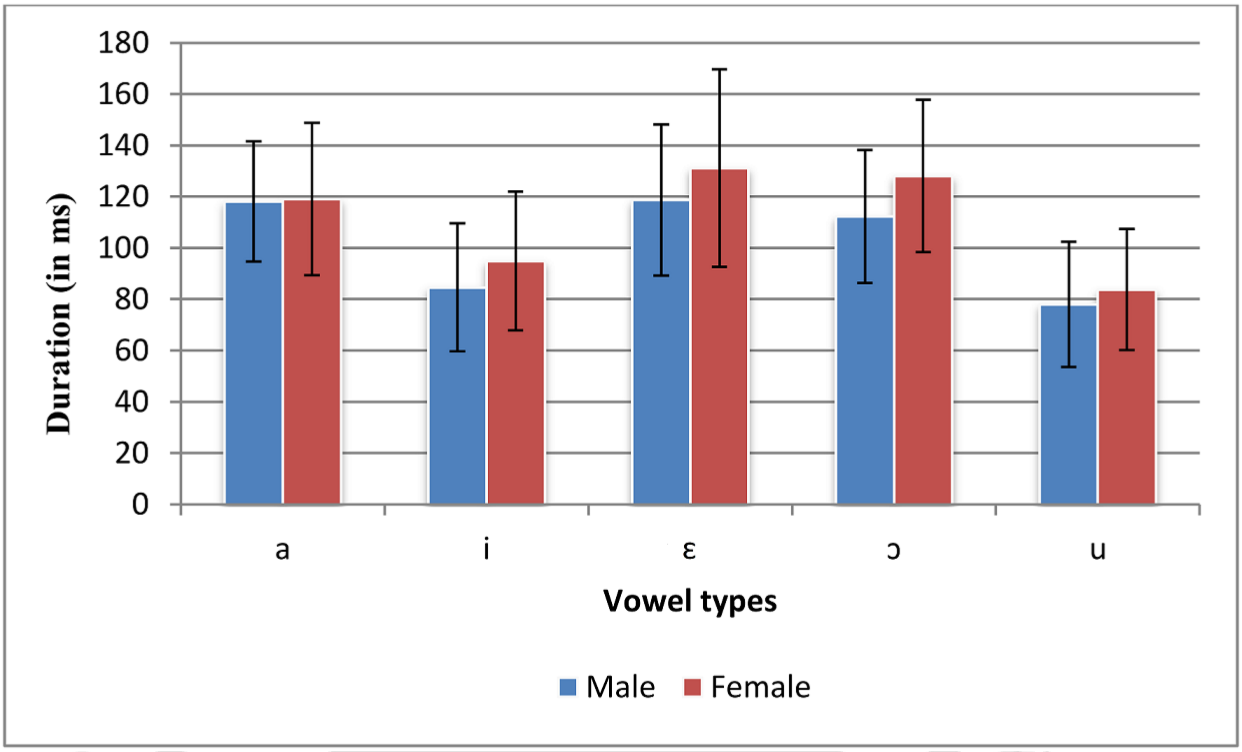

Figure 2.9: Average duration of Sylheti vowels with standard deviation as error bars.

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 significant factor (which is fine since the remaining 3 vowels are mid to low vowels). 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. Significant pairs are marked with (*), and (!) is used to represent the non-significant pairs.

Vowel Types	N	a	i	ɛ	ɔ
i	120	*			
ɛ	120	!	*		
ɔ	120	!	*	!	
u	120	*	!	*	*

Table 2.6: Significance matrix for duration of all the vowels in Sylheti.

Formant frequencies

As we mentioned above, the first three formant values of each vowels 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 2-10).

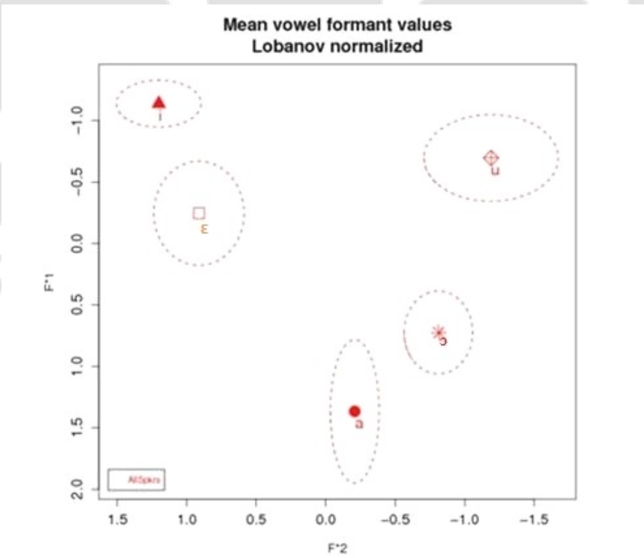

Figure 2.10: Vowel diagram showing average Lobanov normalized formant frequencies of the first two formants with one standard deviation ellipses. N= 120 per vowel.

Non-normalized F1, F2, and F3 formant frequencies (measured in Hz and Mel) of Sylheti vowels are shown for male (Table 2-7 and Table 2-8) and female (Table 2-9 and 2-10) speakers respectively.

Male (measured in Hz)	F1 (SD)	F2 (SD)	F3 (SD)
a	876.79 (93.37)	1735 (222.65)	2813.45 (204.83)
i	512.03 (48)	2038.82 (293.96)	2830.63 (363.19)
ɛ	614.31 (63.05)	1948.94 (257.11)	2933.41 (185.25)
ɔ	667.51 (86.32)	1685.17 (276.35)	2727.02 (252.47)
u	523.76 (52.14)	1434.07 (217.52)	2666.06 (220.59)

Table 2.7: Non-normalized formant values (measured in Hz) of Sylheti male speakers with standard deviations (SD), N=60 for each vowels.

Male (measured in Mel)	F1 (SD)	F2 (SD)	F3 (SD)
a	525.34 (29.72)	783 (50.84)	996.01 (30.89)
i	353.44 (78.84)	835.12 (43.26)	994.15 (56.52)
ɛ	409.66 (28.39)	826.09 (54.95)	1013.53 (30.49)
ɔ	435.12 (38.10)	765.01 (50.36)	980.82 (42.83)
u	365.14 (54.36)	701.72 (47.14)	969.18 (38.80)

Table 2.8: Non-normalized formant values (measured in Mel) of Sylheti male speakers with standard deviations (SD), N=60 for each vowels.

Female (measured in HZ)	F1 (SD)	F2 (SD)	F3 (SD)
a	953.90 (82.67)	1590.61 (100.88)	2719.97 (110.56)
i	427.92 (75.36)	2330.61 (220.19)	3046.14 (262.16)
ɛ	644.16 (52.92)	2174.29 (212.48)	3006.08 (182.96)
ɔ	698.19 (52.20)	1343.26 (284.47)	2924.49 (116.70)
u	486.04 (77.39)	1267.19 (242.21)	2870.75 (189.99)

Table 2.9: Non-normalized formant values (measured in Hz) of Sylheti female speakers with standard deviations (SD), N=60 for each vowels.

Female (measured in Mel)	F1 (SD)	F2 (SD)	F3 (SD)
a	557.67 (24.03)	742.48 (12.49)	982.01 (17.70)
i	301.41 (21.96)	958.26 (14.21	1052.83 (21.58)
ɛ	425.10 (44.03)	898.83 (17.45)	1036.38 (15.52)
ɔ	474.81 (34.01)	643.23 (38.72)	1011.42 (17.83)
u	342.14 (28.37)	620.80 (39.31)	1019.69 (18.57)

Table 2.10: Non-normalized formant values (measured in Mel) of Sylheti female speakers with standard deviations (SD), N=60 for each vowels.

Figure 2-10, and the formant values shown in Table 2-7 and 2-8 (for male and female speakers respectively) demonstrate the distinctiveness of Sylheti monophthongs. To see how these observed vowels interact in terms of their formant values, we conducted one way ANOVA tests on F1, F2, and F3 values of each vowel (using vowel types as factor and formant frequencies (F1, F2, F3) as dependant variables). For the statistical tests, we decided to use the raw (non-normalized) data. It was done soas it is assumed that Lobanov's normalization procedure may introduce artefacts when used for measurements based on vowel system that might differ in the overall size and shape (Disner 1980, Clopper *et al.*, 2005, Adank *et al.*, 2007). Therefore, Lobanov's normalization procedure was used to draw the vowel diagram only, and all the statistical tests were done separately for male and female speakers using the non-normalized Mel data. The results of the ANOVA test confirmed a strong interaction between vowel types based on their formant values (for both the gender).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]).

A subsequent post-hoc Tukey test was conducted to observe systematic interaction among the vowel pairs based on formant frequencies. The results revealed that in case of F1, all the vowel pairs differ significantly from each other except for the pairs [i] and [u] and [ɛ] and [ɔ]. A similar pattern was observed for both the gender which suggest that in terms of vowel height, Sylheti vowels significantly differ from each other (except for the pairs [i] and [u] (which are expected to be of similar height, and hence did not differ), and [ɛ] and [ɔ].

In case of F2, we noticed consistent patterns of interaction in female data. Except the pair of [i] and [ɛ] all the remaining vowels do seem to differ from each other significantly. Since, F2 denotes the backness property of a vowel; we concluded that in terms of their backness property, all the vowels are significantly different from each other. However, the male data did not show a consistent pattern. Analysis of male data revealed that the vowel [a] did not differ significantly from the other vowels except [u]. As it was also observed in case of female data the pair of [i] and [e] also did not show any significant interaction (which seems to be obvious) in terms of the backness property of these vowels. The details of significant interaction between the vowel types and their formant values are provided in table 2-9 and 2-10 ([*] indicates a significant interaction, whereas [!] denotes anon-significant pair.

In order to examine the perceptual distance between the vowel pairs in Sylheti, we calculated the Euclidean distances that uses the formula in (1).

$$D = \sqrt{F1_i - F1_j)^2 + (F2_i - F2_j)^2} \tag{2.1}$$

(D is distance being calculated and ‘i’ and ‘j’ stand for two different vowels), F1 = first formant, F2 = second formant. Euclidean distances between vowel pairs in Sylheti vowel inventory were calculated separately for male and female speakers (Table 2-11 and 2-12). As expected, we observed biggest differences among the peripheral vowels [i], [u], and [a]. The pattern was found to be same in both male and female speakers’ data. The pair of [i] and [ɛ] showed the smallest difference for both male and female speakers. The average Euclidean distance and overall vowel space is found to be less for male speakers (108 Mels) as compared to female speakers (168 Mels).

Male	a	i	ɛ	ɔ	u
a	0				
i	163.09	0			
ɛ	126.03	37.60	0		
ɔ	109.45	75.17	53.97	0	
u	192.30	111.26	120.72	88.72	0

Table 2.11: Euclidean distance between vowel pairs of Sylheti male speakers.

Female	a	i	ɛ	ɔ	u
a	0				
i	212.97	0			
ɛ	140.92	84.77	0		
ɔ	125.96	204.99	178.66	0	
u	215.75	201.57	215.67	97.182	0

Table 2.12: Euclidean distance between vowel pairs of Sylheti female speakers.

2.4 Conclusion

The phoneme inventory of Sylheti is discussed in this chapter. With the durational measurements, we have shown that the voiced stops in Sylheti do differ only in terms of place of articulation. The underlying breathiness property of the stop consonants did 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 2-2 and 2-3) which prompted us to consider the onset of the voicing start till the burst point. As expected, no statistical difference was observed between the underlying voice aspirated and their non-aspirated counterparts. 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. Below we represent an inventory of (possible syllable types for underived monosyllables and disyllables in Sylheti (Table 2-13).

Monosyllables	Disyllables
VV- [ɔi] 'that'	VCV- [i. gu] 'this one'
CV- [bɔ] 'sit'	CVCV- [ki. gu] 'who'
VC- [ar] 'and'	CVCVC- [ɸa.zil] 'naughty'
VVC- [aiz] 'today'	CVCCVC- [mɔr.kɔt] 'idiot'
CVV- [bai] 'brother',	VCVC- [u. d̪am] 'fully open'
CVC- [kɔr] 'do'	CVCCV- [kɛm.nɛ] 'how'
CVVC- [kait̪] 'bend',	

Table 2.13: Syllable structures of Sylheti.

We conclude that, the systematic deletion of underlying breathiness property [+spread glottis] and the process of lenition are certain to affect the pitch of the adjacent vowels. The third chapter of this dissertation discuss the possible outcome of these changes in detail.

Chapter 3

Lexical Tones in Sylheti

Introduction

In the previous chapter on Sylheti phonemes, with the help of acoustic components, we have shown that the consonant inventory of Sylheti has undergone significant reduction and restructuring, and lost the underlying breathiness contrast [+spread glottis] of the entire stop series largely due to the phonological process of obstruent weakening ([b^{h}at̪] > [**b**at̪]'rice', [lat̪^{h}i] > [la**t̪**i]'kick'. With spectrographic evidence, we have also shown that the consonant inventory is further reduced due to two types of lenition process; viz. spirantization and deaffrication. While the underlying form 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**]) and velar voiceless stop (both aspirated and unaspirated) [k] and [k^{h}]([kali] > [xali] 'ink', [k^{h}al] > [xal] 'drain/channel') were spirantized, the underlying form of voiceless alveo-palatal affricate [tʃ] and [tʃ^{h}] ([tʃa > sa] 'tea', [tʃ^{h}uti] > [suti] 'holiday') and voiced alveo-palatal affricate [dʒ] and [dʒ^{h}] ([dʒal] > [zal] 'net', [dʒ^{h}al] > [zal] 'spicy') exhibited the process of deaffrication, thus giving birth to numerous homophonous like words in Sylheti. In this chapter[1], we have analysed the acoustic property that distinguish these homophonous pairs.

[1]We sincerely than Prof. Bert Remijsen (University of Edinburgh) for his suggestions and detailed feedback on the draft version of this chapter

3.1 Rationale for the experiment

The literature on tonogenesis reveals that the loss of one or the other laryngeal features such as voicing contrast, aspiration or breathiness contrast (also called murmur voicing), glottalization, etc. of adjacent consonants are the usual sources of tonal contrast. Generally voiced obstruents are known to lower the pitch of adjacent vowels (Hombert 1978, Maddieson 1977), and voiceless obstruents may even raise the pitch of neighbouring vowels (Hombert et al. 1979, Yip 2002). Voiced aspirated stops are believed to be the greatest suppressor of pitch of the adjacent vowels (Dutta 2007). Researchers also speculate that tone often arises from a toneless state (Abramson 2004) - a language therefore, may acquire or lose (lexical) tone due to areal effect (Gussenhoven 2004).

Among the Indo-Aryan languages, tonal specification in Punjabi is rather well established. In Punjabi, breathy voiced consonants lost the breathy specification and the vowel following it acquired a low tone ([b^{h}a]→[pá]) (Hombert, Ohala & Ewan 1979, Vijayakrishnan 2009, Bhaskararao 1998, Bhatia 1975).

Presence of a high tone as a result of deaspiration in the syllable sequence of #BHV (initial breathy voiced consonant followed by a vowel), #HV, and VBH# (syllable containing a vowel and a breathy voiced consonant in the word boundary) was reported to be present in a few other Indo-Aryan languages such as Northern Haryanvi, Mandeli and so on (Bhaskararao 1998). However, the plausible phonetic reason related to the presence of a high tone in those languages seemed elusive. In this dissertation we have attempted to address this issue.

We hypothesized that the overall reductions of Sylheti phoneme inventory(especially the loss of the phonological property [+spread glottis] and the lenition process) leading to a large number of homophonous words might have been compensated and categorized with an additional property: tone. This impressionistic perception is explored and quantified by examining the acoustic correlate of perceived pitch or tone of the homophonous words. To understand the distinct acoustic and phonological properties (if any) of the homophonous pairs distinguishing the lexical contrasts, we conducted our third production experiment and examined the pitch patterns of the homophonous words. The following section describes the methodology adopted for data collection, digitizing them, and the acoustic analysis of speech data.

3.2 Methodology

3.2.1 Data Collection

In the introductory chapter of this dissertation we have mentioned that the total findings of this dissertation are based on multiple rounds of production experiments. Our primary goal of the current experimental study is to explore the tonal inventory (if any) of the language under research. We seek to distinguish and classify different tones on the basis of differences in phonetic pitch (fundamental frequency). We assume that the loss of underlying breathy voice contrast among the obstruents and the lenition process might cause a perturbation in the f_0 of the adjacent vowels (the voiced part of the syllable or the rhyme). Hence, we will regard the pitch of those adjacent vowels (the voiced part of the vowel or the rhyme) as the indicator of tonal contrast in the language under study.

3.2.2 Participants, materials and recording procedure

For the current experiment , the same eight native speakers of Sylheti (six male, two female and aged between 18 and 42 years) from Dharmanagar district of north Tripura who earlier participated in the first production experiment of VOT measurement were recorded following the same recording procedure adopted in the previous experiments (as discussed in chapter 2). However, we could not include one of the male speakers data for the analysis of disyllabic words due to disturbances in the recording. The target words for the production experiment comprised a list of segmentally homophonous pairs of words gathered from two native speakers of Sylheti during our initial pilot study. The dataset was further refined after consulting two older native speakers. The contrastive pairs were chosen in such a manner that historically one of the words had distinctive aspirated stop in either onset or coda position (please see the dataset- Table

3-1, Table 3-4 and Table 3-11 for more details). A few words have undergone both the processes of spirantization and also lost their underlying breathiness property and thus, appeared to be homophonous. Those words were compared with the words which already existed with their unaspirated counterparts. There were two words which exhibited only the process of spirantization - [ɸɔr] 'read' and [ɸɔr] 'guard'. There were also two words which were not the result of any historical changes: [mela] 'fair' and [mela] 'many'. The dataset contained a total number of 24 monosyllabic lexical items and 27 disyllabic words including 9 three-way minimal sets (these words were analyzed separately- section 3.6). The monosyllabic words were of CV, CVV, and CVC syllable types and the disyllabic words were of V.CV, CV.CV, and CV.CVC syllable structure.

The target words were embedded in a fixed sentence frame of "I am saying **X**" [ami **X** xɔiar], X being the target word. For the subjects to be able to maintain the tonal contrast between the words with distinct meanings, a method of priming was used.The priming method involved the recording of an example sentence where the meaning of the contrastive word would be best illustrated. For example, in order to elicit the word [bat̪] 'rice', an example sentence 'I am eating **rice**' [ami **bat̪** xaiar] was recorded first, followed by the target word [bat̪] in the carrier frame. Each sentence containing the target word was displayed on a computer screen and the subjects were asked to produce those sentences in their mother tongue in a natural way. The meaning of each word was written along with the sentence frame. The target word was situated in the sentence medial position in order to ensure that the target word was not influenced by different segmental properties of the adjacent words. This procedure also ensured that the intonational interference on the target words was uniform. Subjects were asked to repeat each sentence 4 times.However, only the best three iterations were considered for the analysis.All the lexical items in the carrier frame sentence were manually randomized in such a way that the minimal set did not occur in shorter succession. Since the entire recording procedure was done during the field-visit in Dharmanagar district of north Tripura, the recording conditions were different for each of the subjects (most of them were recorded at their own home using an empty room ensuring minimal disturbance). Even though utmost care was taken to minimize the effects of noise, however, in rare cases, background noise was audible. A Few tokens were also distorted due to an echo emanating from an empty room. After carefully listening to the entire data of each speaker, those distorted tokens were excluded from further analysis. The entire recording session varied from one hour to one and half an hour, an intermediate break was given after every 20 minutes of the recording time.

3.2.3 Methods: Acoustic analysis of f_0

All the sentences were digitized at a sampling frequency of 44.1 KHz and 32 bit resolution. After each session, participants data was transferred from the recorder to a portable PC using a USB cable. Each iteration of the target words was separated from the sentence using the wide band spectrograms and waveform displays and saved as individual sound file using Praat (version 5.3.04_win32) (Boersma and Weenink, 2012). Individual sound files of target words were manually segmented and a three tier Praat Textgrid file was created for each of them - the first tier representing the vowel (rhyme or the voiced part of the syllable); individual phonemes of each word were marked in the second tier, and the third tier was for the whole word.

The pitch bearing vowels (the rhyme) of all the target words were visually examined and labelled by observing the pitch of the onset (the point of initiation) and offset (the point of termination) in the region of the syllable nucleus or the entire rhyme, as all voiced codas in our data were only sonorants. Thus, each rhyme consists of the vowel plus any voiced coda. All the f_0 related measurement (f_0 measured at various points such mean f_0, maximum f_0, minimum f_0, f_0 at vowel mid-point and so on), were made over the voiced part of the rhyme of the target word, however duration and intensity were measured for the vowel only.

A Praat script was written (by the author) to measure the pitch contour at every 10% of the total duration of each pitch bearing vowel. Pitch was thus calculated at 11 consecutive points- starting from the onset ('start-pitch' [0%]) till the offset ('end-pitch' [100%]) across the duration of each vowel (the rhyme); each point representing 10% of the total length of the pitch track. This was done with a pitch floor of 40 Hz and pitch ceiling of 600 Hz with a default time step of 10 milliseconds.

The percentage-wise pitch values were averaged across all the three iterations of each word produced by each speaker separately and was plotted as a line graph using a spreadsheet in order to observe the distinct pitch contours. As mentioned earlier, we speculated that the loss of underlying breathy voice contrast [+spread glottis] among the obstruents and/or the lenition process might cause a perturbation in the f_0 of the adjacent vowels. Pitch contours of those adjacent vowels (or the rhyme if the coda is a sonorant) would be considered as the indicator of lexical tone in Sylheti.

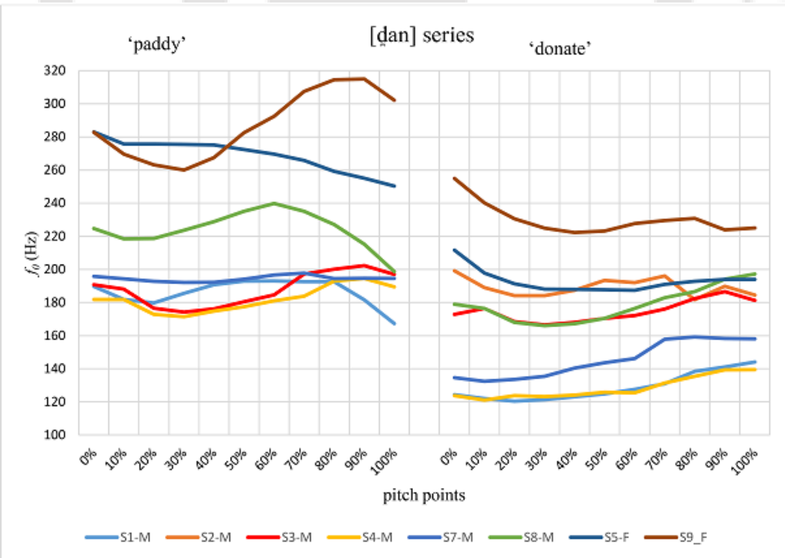

Figure 3.1: Non-normalized pitch tracks for [**d̪an**] as produced by all the speakers are drawn using the percentage wise pitch values (n=3).

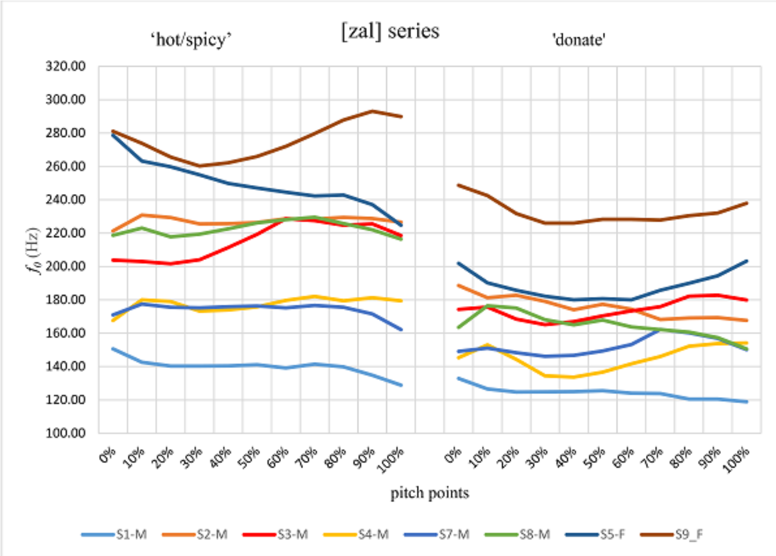

Figure 3.2: Non-normalized pitch tracks for [**x̪al**] as produced by all the speakers are drawn using the percentage wise pitch values (n=3).

Figure 3-1 and 3-2 show non-normalized averaged f_0 tracks of [d̪an] series (meaning 'paddy' and 'donate'), produced by a female and a male speaker respectively. The f_0 tracks of the homophonous set are realized with two distinct pitch contours- the loss of the underlying breathiness property in the word [d̪an] 'paddy' [d̪^{h} > d̪], is realized with higher f_0 (around 40 Hz for female speakers, and 32 HZ for the male speakers when averaged across all the iterations from all the female and male speakers separately) than the homophonous counterpart [d̪an] 'donate'.

Similarly, visual inspection of spectrograms of the remaining homophonous words also shows distinctive f_0 tracks for each individual word, thus confirming the presence of tonal contrasts in Sylheti. Presence of a high tone is observed following the loss of underlying breathy voice property in Sylheti obstruents. Using the same script (that was used to measure percentage-wise pitch values), values for duration, intensity, mean f_0, maximum f_0, minimum f_0, f_0 at vowel mid-point, were also measured. A series of statistical tests were conducted to examine if the differences in pitch observed visually were significant or not.

3.3 Normalization of speech data

The importance of normalizing the raw f_0 values for distinguishing tonal contrast is well attested in the literature (Disner 1980, Nolan 1983, Fant 1973, Rose 1987, 1991). Since the physical anatomy (the size and the length) of the vocal tract of an individual differ from each other, it is very likely that the acoustic properties of speech signal produced by different speaker is also bound to differ- even for the same phonetic sound (Rose 1991). Since f_0 is the basic acoustic correlate of perceived pitch, functioning as a dimension for suprasegmental linguistic systems of intonation, stress, and tone (Lehiste 1970), therefore, it becomes inevitable to exclude the individual acoustic/accentual property in order to define the phonetic and phonological property of a specific suprasegmental form like tone. So for example, since the male speakers tend to have longer and wider vocal tract than their female counterparts, they are very likely to produce lower f_0 values. Consequently, it is possible that a male speakers phonological high tone is produced with a lower f_0 than a female speakers low tone (Rose 1991).

While analyzing the pitch contours of Sylheti homophonous words, we observed that the difference between the contrastive tones (mean f_0) is around 30 Hz for male speakers and around 42 Hz for the female speakers. The non-normalized data showed a significant difference of f_0 among the male and female speakers ($p < 0.05$, $F [1, 414] = 357.15$, $p = 0.01$). It was observed that on average, Sylheti female speakers average fundamental frequency was almost 65 Hz higher than the male speakers. In order to avoid inter/intra-speaker and token variations, we adopted the z-score normalization procedure following Disner (1980) and Rose (1987, 1991) that uses the formula in (1).

$$z = \frac{(f_{0i} - x)}{SD} \tag{3.1}$$

where f_{0i} is the sampling point (such as 'start pitch' calculated at the onset of the TBUs (0%), 10th, 20th, ... 90th, 100th point) (in Hertz), x is the average f_0 (in Hertz) of all the sampling points (i.e., the average f_0 values of all the values for a speaker, across all items, both low and high-toned), and , and SD represents the standard deviation of all the sampling points. Following this method, the percentage-wise pitch values (measured in Hz) of each token produced by each speaker were normalized into z score values. The procedure adopted for calculating z score values give us the freedom of converting the raw f_0 values as multiples of so many standard deviations above or below their mean. The z score values were averaged across all the tokens and all the speakers (for all the 11 points) and plotted on a graph to reconstruct the pitch track. We assume that the pitch track occurring below the standard deviation (below the neutral point 0.00, and thus will possess negative values) will be indicative of low tone whereas those occurring above the standard deviation point (occurring with positive values) will be indicative of high tone.

3.3.1 Results - Tones in monosyllables

The normalized pitch tracks clearly demonstrate the presence of a high tone following the loss of historically breathy voice contrast [+spread glottis]. The monosyllabic words included in the current experiments were of CV, CVV, and CVC syllable types (Table 1).

Sylheti words	**Gloss**	**Sylheti words with a history of underlying aspiration**	**Gloss**
[ɸɔr]	'read'	[ɸɔr]	'guard'
[zao]	'gruel'	[zao]	'tamarisk'
[ga]	'body'	[ga]	'wound'
[gai]	'cow'	[gai]	'stroke'
[bat̪]	'arthritis'	[bat̪]	'rice'
[d̪ax]	'roaring of cloud	[d̪ax]	'drum'
[ban]	'tie'	[ban]	'pretend'
[d̪an]	'donate'	[d̪an]	'paddy'
[xal]	'skin'	[xal]	'channel/drain'
[xua]	'well'	[xua]	'jackfruit cell'
[zal]	'net'	[zal]	'chilly hot'
[gail]	'scold'	[gail]	'husking device'

Table 3.1: The dataset displaying the list of monosyllabic words considered for the current experiment.

In case of CVC syllable types, the voiced coda was either a nasal consonant [n] or a lateral approximant [l], therefore the pitch of the entire rhyme was considered for those words. The loss of underlying breathy voice property [+spread glottis] of the onset voiced stops [d̪] and [b] is realized with an induced f_0 in the adjacent vowel and extended till the offset of the rhyme. Consider the following examples (Figure 3-3 and Figure 3-4):

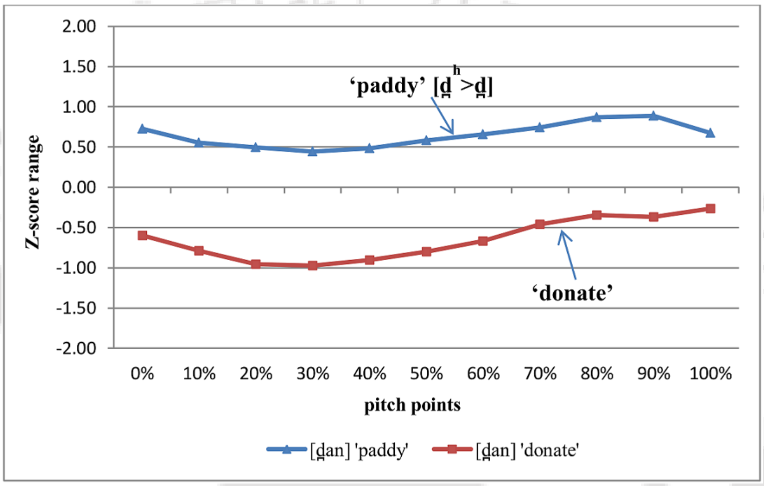

Figure 3.3: Normalized pitch tracks for [**d̪an**](n=24, 8 speakers * 3 iterations each).

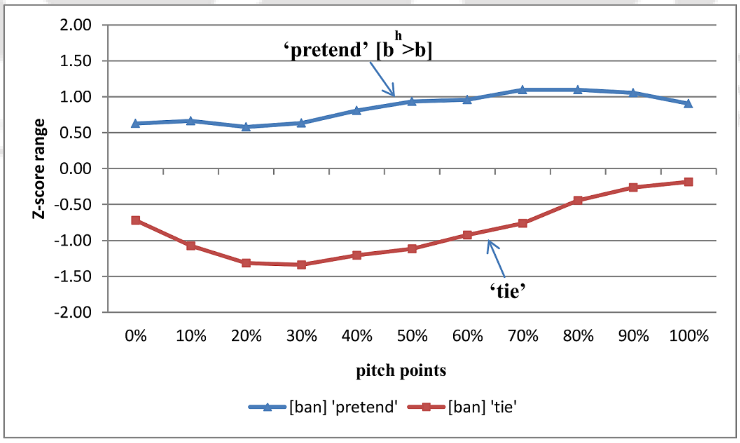

Figure 3.4: Normalized pitch tracks for [**ban**](n=24, 8 speakers * 3 iterations each).

Similarly, distinct f_0 track was also observed in the words with CVC syllable types where the final syllable is either a voiceless fricative or a voiceless stop. This could be seen in the following examples (Figure 3-5 and Figure 3-6):

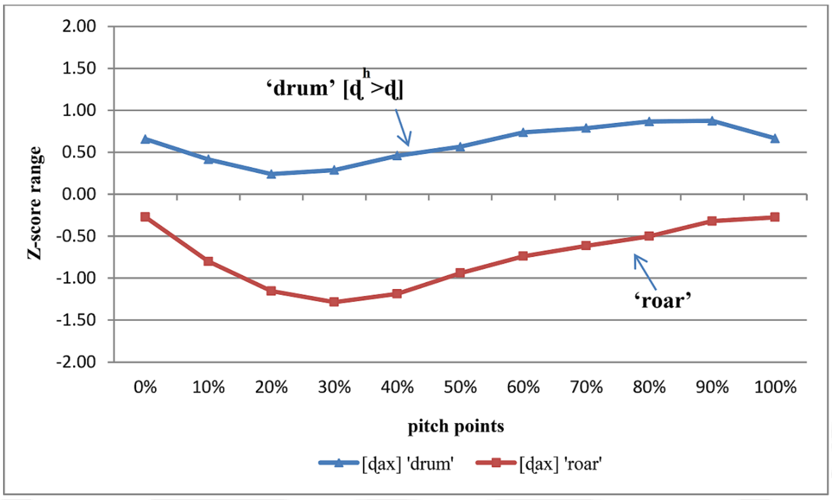

Figure 3.5: Normalized pitch tracks for [**dax**](n=24, 8 speakers * 3 iterations each).

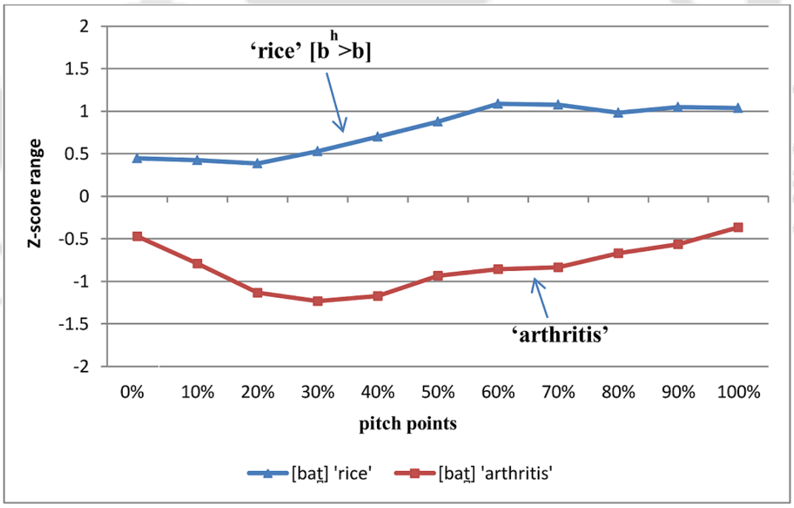

Figure 3.6: Normalized pitch tracks for [**bat̪**](n=24, 8 speakers * 3 iterations each).

The presence of a high f_0 following the loss of underlying breathy voice property is also observed in CV (Figure 3-7) and CVV (Figure 3-8) syllable types words. Interestingly, we have noticed that the effect of f_0 carries on throughout the syllable.

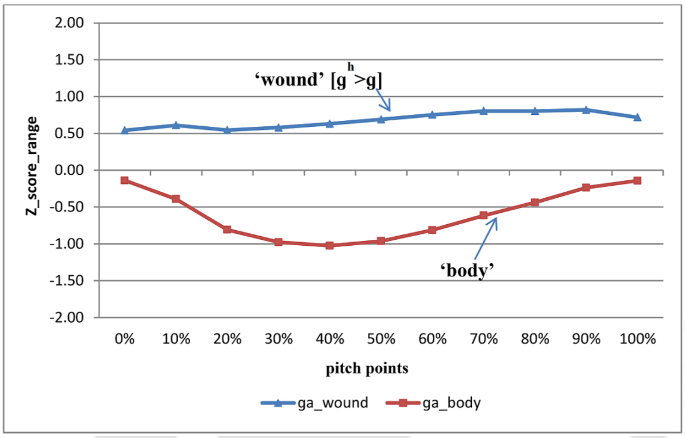

Figure 3.7: Normalized pitch tracks for [**ga**](n=21, 7 speakers * 3 iterations each).

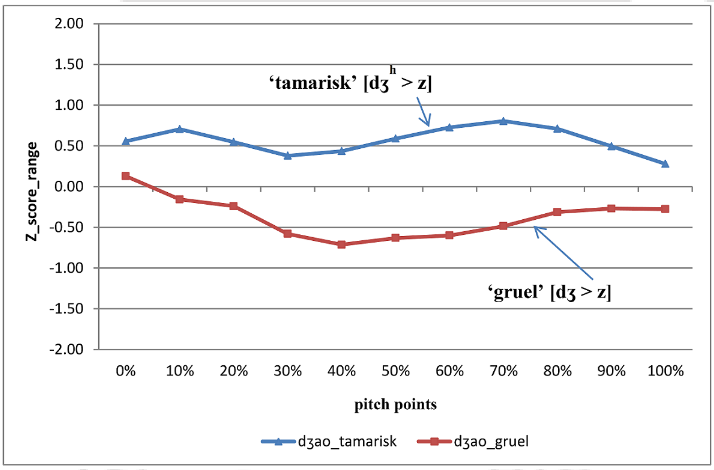

Figure 3.8: Normalized pitch tracks for [**zao**](n=21, 7 speakers * 3 iterations each).

Note that the same tonal property is also maintained in the words that exhibited two way loss; viz. the loss of underlying breathiness [+ spread glottis] and [spirantization] as could be seen in the word [zao] 'tamarisk' [dʒ^h > z].It must be noted that the difference between the traces is greater in the middle of the vowel/ voiced part of the rhyme (as could be seen from all the figures shown above). This confirms that it is not a phonetic side-effect of the onset consonant.

3.3.2 Statistical Analysis

Following visual examination of each of the pitch contours of 24 monosyllabic words, we categorized those words into two distinct tonal categories, viz.; high and low. In order to confirm whether the differences between the tonal categories are significantly different or not, repeated measures-style (RM) ANOVAs test was conducted. Since the mere significance of mean f_0 among the homophonous words might not be sufficient enough to establish the presence of tonal contrast in Sylheti, we also considered a few others existing components such as maximum f_0, minimum f_0, f_0 at vowel-mid point, duration and intensity . To reconstruct the normalized pitch contour, we measured the f_0 values of each rhyme at every 10% of the total duration (11 consecutive points, starting from the pitch onset (0%) till the pitch offset (100%) of the rhyme) and the Hertz values were subsequently converted into z score to reduce the difference between male and female speakers.

To examine the presence of any other possible acoustic cues determining the lexical contrast of the homophonous words, we also included the duration and intensity values of the vowels bearing contrastive f_0 tracks in our statistical tests. The same procedure was applied to analyze both the monosyllabic and disyllabic words. Repeated measures-style ANOVA's were conducted using a fixed factor- tone (high and low), and a random factor speaker. Our dependant variables where various acoustic components such as mean f_0, maximum f_0, minimum f_0, f_0 at vowel-mid point, duration and intensity. The alpha value was set at 0.05 as the criterion to determine significance among the contrastive tones. The results of the ANOVA show that the factor tone has a significant effect on all the acoustic measurements related to f_0. For example, when mean f_0 of the homophonous pairs was compared- Mauchly's test indicated that the assumption of sphericity had been violated χ^2 (2)= 8.98, p = .011. Therefore, degrees of freedom were corrected using Huynh-Feldt (ε = .93). The results show that there was a significant effect of mean f_0 on tone types- F (1.87, 169.9) = .17, p = 0.00. Similarly, all the f_0 measurements were also found to have a significant effect on tone types: maximum f_0 F (1.73, 143.9) = .14, p = 0.01, minimum f_0 F (1.47, 174.9) = .19, p = 0.02, and f_0 at vowel mid-point F (1.45, 113.9) = .11, p = 0.00. However, acoustic components such as duration and intensity weren't found to be statistically significant. In case of duration, Mauchlys test indicated that the assumption of sphericity had been met $\chi^2(2)$= .96, p = .19. The results show that there was no significant effect of duration on tone types- F (2, 182) = 0.27, p = 0.76. Similarly, the intensity values did not show any significant interaction with the tone types- F (2, 117) = 0.93, p = 2.5.

Table 3-2 represents the descriptive statistics below:

		N	Mean	Std. Deviation
Duration	Low Tone	213	189.04	43.13
	High Tone	203	193.69	43.23
	Total	416	191.31	43.19
Intensity	Low Tone	213	86.19	5.16
	High Tone	203	86.96	5.49
	Total	416	86.57	5.33
f0_mean	Low Tone	213	179.50	36.31
	High Tone	203	207.49	45.27
	Total	416	193.16	43.21
f0_mean	Low Tone	213	154.67	27.26
	High Tone	203	175.08	32.59
	Total	416	164.63	31.64
f0_vowel_mid	Low Tone	213	175.22	35.43
	High Tone	203	207.47	44.45
	Total	416	190.96	43.17
f0_vowelmid	Low Tone	213	151.46	26.74
	High Tone	203	175.10	32.01
	Total	416	162.99	31.69
Max_Pitch	Low Tone	213	198.65	41.22
	High Tone	203	221.86	50.02
	Total	416	209.97	47.13
Min_Pitch	Low Tone	213	167.38	37.39
	High Tone	203	192.912	43.85
	Total	416	179.84	42.59

Table 3.2: Descriptive statistics showing the average value (across all the speakers and all the variables) for each acoustic measurement considered for statistical analysis for Sylheti tones in monosyllabic words.

The non-significant interaction between the factor tone and the dependent variable such as duration and intensity (duration: F (1, 414) = 1.21, $p > 0.05$, p =0.27), intensity F (1, 414) = 2.19, $p > 0.05$, p =0.14) confirms that f_0 is the only acoustic component of realization of tonal differences in Sylheti.

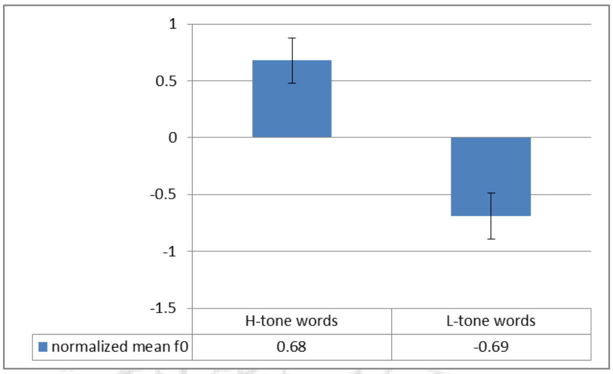

Figure 3.9: Means of normalized pitch tracks of monosyllabic words with standard error bars.

The normalized pitch contours of 24 monosyllabic words along with the statistical tests confirm the presence of tonal contrast in Sylheti. We conclude that the loss of the breathiness property [+spread glottis] along with the phonological process of lenition is substituted with a tonal contrast to ensure the distinctive lexical meaning among the homophonous words. The non-significance of duration and intensity values of the vowels carrying contrastive tones further supports our claim that the native Sylheti speakers indeed rely on pitch differences as acoustic cues to determine the lexical meaning.

3.4 Tones in disyllabic words

Once we were convinced about the presence of tonal contrast in Sylheti, we examined the tonal properties of 27 disyllabic words. 9 (homophonous) words with three-way lexical distinctions were examined separately. Thus the list of disyllabic words with two way tonal contrasts comprised 18 words as presented in Table 3-3.

Sylheti words	Gloss	Sylheti words with a history of underlying aspiration	Gloss
[aɖa]	'ginger'	[aɖa]	'half'
[ali]	'lizard'	[ali]	'unit of four'
[bala]	'bracelet'	[bala]	'good'
[ɖala]	'tray'	[ɖala]	'pour'
[bari]	'home'	[bari]	'heavy'
[ɖaxa]	'call'	[ɖaxa]	'cover'
[baʃɔn]	'utensil'	[baʃɔn]	'lecture'
[gɔsa]	'thorn'	[gɔsa]	'persuaded to take'
[mela]	'fair'	[mela]	'many'

Table 3.3: The dataset displaying the list of disyllabic words (with two way lexical contrast) considered for the current experiment.

Like it was observed in Sylheti monosyllabic words, we presumed to perceive a similar two way tonal contrast in case of disyllabic words too. As it was done for the monosyllabic words, the raw f_0 values were measured over 11 consecutive points-(starting from the onset of the vowel till the offset of the rhyme/voiced part of the syllables)and Hz values were subsequently converted into normalized z score values using the same normalization procedure adopted for analysing the monosyllabic words. As we had mentioned earlier, due to noise intervention on either of the disyllabic pairs, we could not include the data of one of the male speakers for the acoustic measurements. Therefore we concentrated on the data collected from seven native speakers (5 male, 2 female). Using the normalized z score values (of three iterations of each word) we reconstructed the pitch contours of each of the words produced by each speaker separately. Using the data from all the speakers averaged normalized pitch contours were reconstructed for each word. Like the monosyllabic words, the statistical tests were conducted using the z score normalized data taken on various positions on the total f_0 tracks.

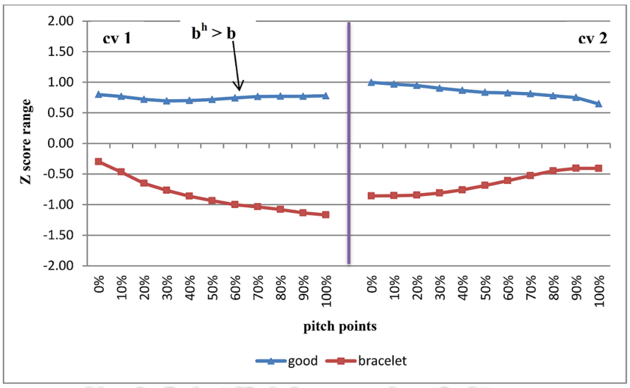

Figure 3.10: Normalized pitch tracks for [**bala**](n=21, 7 speakers * 3 iterations each).

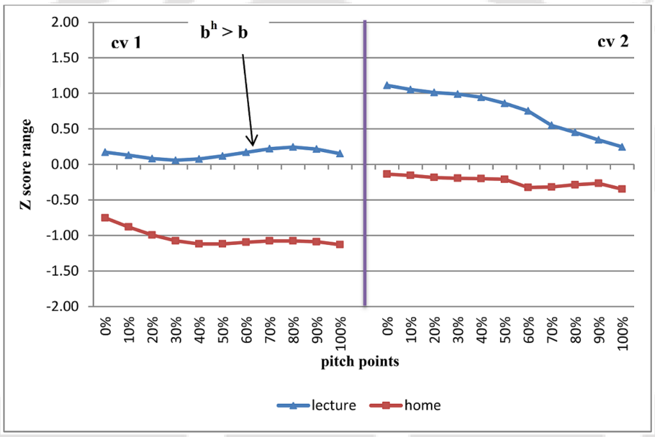

Figure 3.11: Normalized pitch tracks for [**baʃɔn**](n=21, 7 speakers * 3 iterations each).

Like the monosyllabic words, the normalized pitch contours averaged across all the iterations and all the speakers demonstrate two ways tonal contrasts in Sylheti, viz., a high tone emerged either due to loss of underlying breathiness contrast and/or due to the process of spirantization plus loss of underlying breathy voice property. In case of disyllabic words, we observed that, the tonal contrast is realized in both the syllable with contrastive tones; that is, the high tone is realized as high in both the syllable, and the low tone is realized with negative z score values in both the syllables.

When we compared the average duration of both the syllables of the contrastive disyllabic words, we observed that CV1 of both the contrastive words tend to have slightly longer duration value for both the words, however that difference was not significant. Below we present the descriptive statistics for duration, intensity and mean

f_0 (normalized z score) in Table 3-4.

Low tone words	CV 1	CV 2	High tone words	CV 1	CV 2
Vowel Duration (in ms)	89.42	88.59	Vowel Duration (in ms)	92.75	87.67
Intensity	85.84	84.85	Intensity	86.59	84.85
f0_mean (z normalized)	-.84	-.80	f0_mean (z normalized)	0.38	0.22

Table 3.4: Descriptive statistics showing the average value (across all the speakers and all the variables) for duration, intensity and mean f_0 of Sylheti disyllabic words.

Following a keen observation of the f_0 tracks of all the disyllabic words, we shall argue that each syllable of underived disyllabic words in Sylheti is specified for identical tones[2]. Consider Figure 3-12 and Figure 3-13:

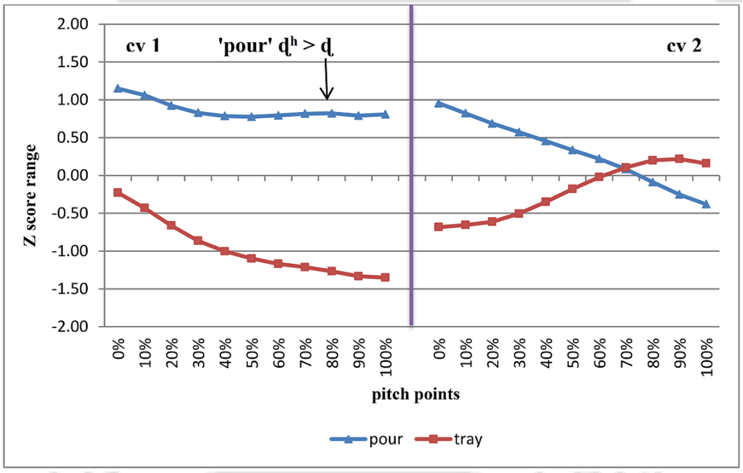

Figure 3.12: Normalized pitch tracks for [**dala**](n=21, 7 speakers * 3 iterations each).

The normalized average pitch tracks shown in Figure 3-13 and 3-14 evidently support our claim that in case of Sylheti underived disyllabic words; both the syllables are specified for identical tone. In that case if the pitch of the first syllable of a high tone word is realized with a relatively higher f_0, then the rhyme of the following syllable

[2]In chapter six of this dissertation we will show that Sylheti exhibits a tonal complexities in morphologically derived nouns, the suffixes bears tone opposite to that of the root, thus presents a case of tonal polarity in Sylheti.

Figure 3.13: Normalized pitch tracks for [**ali**](n=21, 7 speakers * 3 iterations each).

(here second syllable) is likely to exhibit a lowering (Figure 3-12). However, if the first syllable of a high tone word is realized with relatively lower f_0, then the successive syllable is realized with a slightly higher f_0 (see Figure 3-13, Figure 3-14 and Figure 3-15). Like the monosyllabic words the difference between the traces is greater in the middle of the vowel / voiced part of the rhyme.

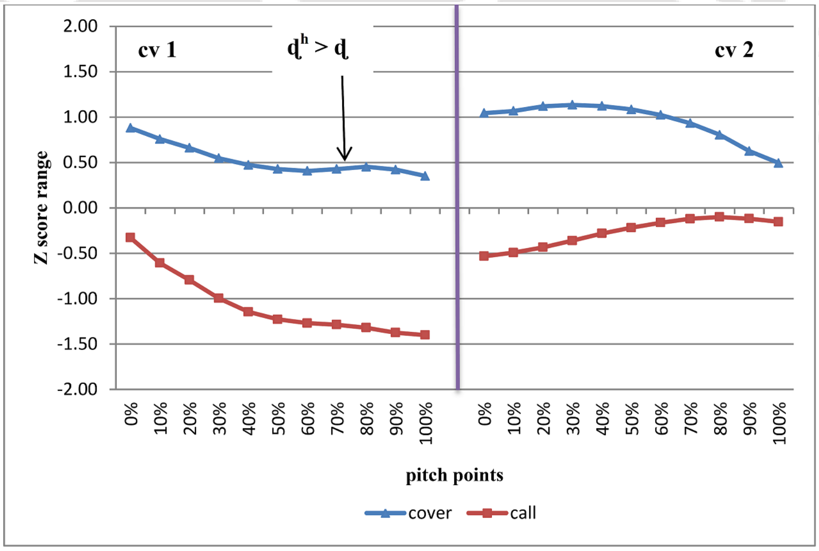

Figure 3.14: Normalized pitch tracks for [**daxa**](n=21, 7 speakers * 3 iterations each).

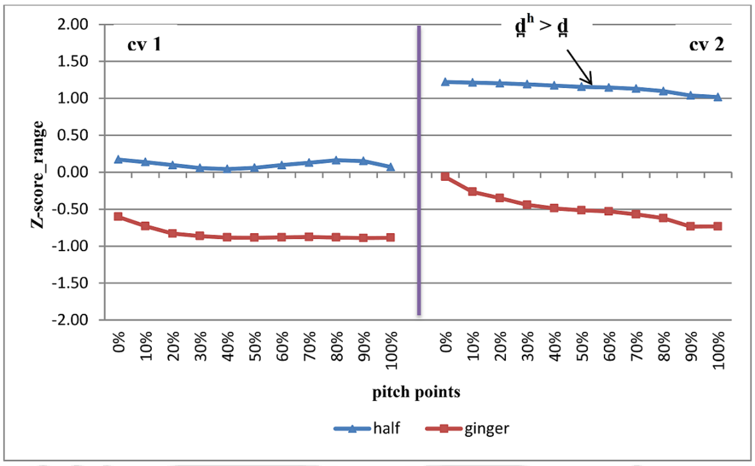

Figure 3.15: Normalized pitch tracks for [**a da**](n=21, 7 speakers * 3 iterations each).

To confirm the speculation of tonal contrast observed through visual examination, we conducted one way Repeated measure ANOVA using the data of disyllabic words. As expected and observed for monosyllabic words too, the value of acoustic components realized in terms of duration and intensity Mauchly's test indicated that the assumption of sphericity had been met χ^2 (2)= .96, p = .19 when compared the duration values of the first syllable from all the speakers and for all the iterations of the contrasting words. The results show that there wasn't any significant effect of duration on tone types for both the syllables- CV1: F (2, 167) = 0.77, p = 0.92, and CV2: F (2, 89) = 0.83, p = 0.7. Similarly, the intensity was also not found to be a significant factor in Sylheti; CV1: F (2, 132) = 0.97, p = 1.2, and CV2: F (2, 117) = 0.73, p = 0.71.

However, a highly significant interaction is observed between the tone types and all the f_0 related measurement: mean f_0 - F (3.87, 126) = .47, p = 0.00(CV1), and F (1.67, 114) = .47, p = 0.02(CV2); maximum f_0 - F (1.17, 113) = .23, p = 0.03 (CV1), and F (1.23, 102) = .62, p = 0.00 (CV2); minimum f_0 - F (1.04, 86) = .49, p = 0.00 (CV1), and F (1.37, 156) = .87, p = 0.03 (CV2); f_0 at vowel mid-point- F (2.63, 126) = .47, p = 0.00 (CV1), and F (1.64, 76) = .57, p = 0.00 (CV2). This issue of tone alignment in morphological derived condition is analysed and re-examined in chapter 5 of this dissertation.

3.5 A third tone in Sylheti?

During our pilot study, two of our older informants (we could not include their voice for any of our production experiments largely due to their physiological problems) provided us list of homophonous words representing multiple-way minimal sets (some had up to four to five ways minimal sets). After consulting the native speakers those who participated in the production experiment, we reduced the number of such pairs to 9, each representing three-way minimal sets. These words occur frequently in their daily communication purposes and all the speakers were well aware of each of their lexical meanings. Due to noise intervention we were forced to leave out a few tokens. Each of these 9 words (Table 3-5) was segmented and acoustic measurements were taken following the same procedure adopted for other monosyllabic and disyllabic words.

Sylheti words	**Gloss**	**Sylheti words with or without a history of underlying aspiration**	**Gloss**	**Sylheti words with a history of underlying aspiration**	**Gloss**
[ɸata]	'grinding stone'	[ɸata]	'goat'	[ɸata]	'send'
[sira]	'flatten rice'	[sira]	'short tempered'	[sira]	'torn'
[mura]	'heel'	[mura]	'stool'	[mura]	'rotate

Table 3.5: The dataset displaying the list of disyllabic words (with three-way lexical contrast).

Following the z score normalization procedure adopted in this study, we converted the non-normalized Hz values (measured at 11 equidistant points) into normalized z scores. The z score values of each iteration of all the 9 words were averaged using the values of each speaker separately and individual pitch contours of each syllable were reconstructed using a spread-sheet. The visual inspection of each speaker's data indicated the presence of a third tone which appeared to be an underlying low-rising tone. Consider the pitch tracks represented in Figure 3-16 and 3-17.

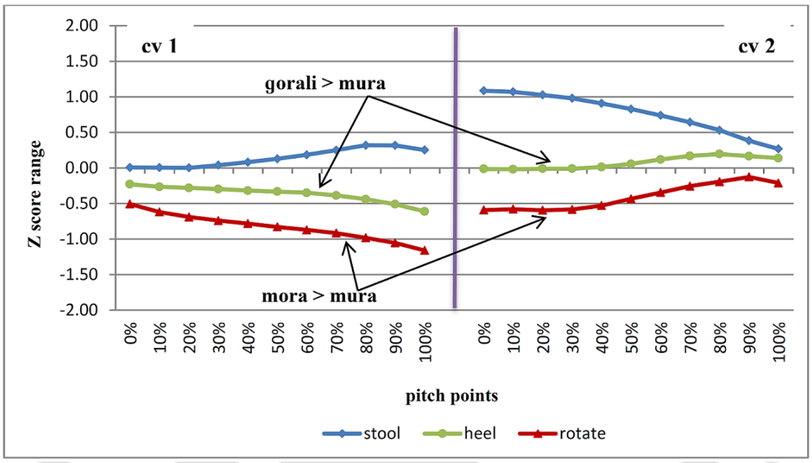

Figure 3.16: Normalized pitch tracks for [**mura**](n=22 for each word).

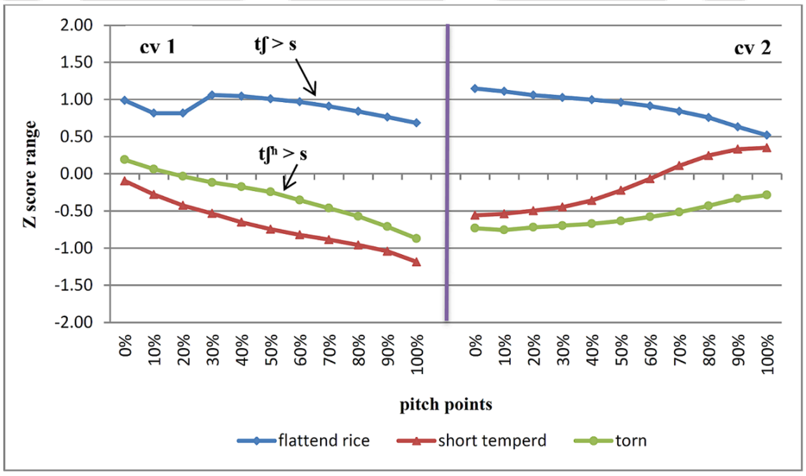

Figure 3.17: Normalized pitch tracks for [**sira**](n=18, for each word).

The tonal contrast between the high and low tone appeared to be certain in both the series of [zal] and [sira]. During the data collection, a few native speakers informed us that they were not familiar with all the words denoting three-way lexical contrast, which subsequently reduced out data. Since sample data were minimum in number, we decided not to conduct any statistical test for these words. However, after listening of the recorded sound files from atleast six speakers for a considerable number of times, we are to some extent sure about the presence of 3rd tone in Sylheti. The normalized average pitch contours also indicate that there might be a three-way tonal contrast in Sylheti, though it requires an expensive and rigorous experiment with more number of words having three (or more) lexical contrast to claim anything extensively.

3.6 Discussion

Dutta (2007), while analysing the properties of Hindi stops[3], argued that non-aspirated voiced stops are known to lower the f_0 of the adjacent vowel while the aspirated counterparts further lowered the f_0 of the neighbouring vowels. The production of breathy voiced stop requires the opening of the glottis at the rear end, thus allowing a higher airflow, resulting in a kind of murmur with mixed voicing and lead to a low tone (Kingston 2007, Ladefoged 1996). Since the breathiness is produced with abducted arytenoid cartilages, therefore, it often lowers the larynx and hence correlates with a low pitch. The observations by Dutta (2007), Kingston (2007) and Ladefoged (1996) do not seem to hold in Sylheti where the loss of the feature [+spread glottis] gives rise to a high tone. There is also the possibility that once this (underlying) breathiness property [+spread glottis] is lost, as it is the case in Sylheti, the higher airflow initially required for producing the breathy voiced stop is directed towards the adjacent vowel largely due to the hypo-correction tendency of native speakers, resulting in a high tone. In chapter two, with durational measurements we showed that Sylheti stops do not contrast in terms of underlying breath voice property. The phonetic factor responsible for a high tone in Sylheti thus seems to be a by-product of the loss of the phonological property

[3]Hindi has four-way stop contrasts

[+spread glottis]. We conclude that the lexical tonal contrast emerged through series of historical developments and phonologized so as to every lexical entry is assigned with an underlying tone. To understand the differences in vowel qualities of contrastive tones in terms of their phonation properties (viz., breathy, modal and creaky), we conducted another acoustic experiment and measured the spectral properties of the contrastive vowels occurring in the monosyllabic words. The details experimental procedure and the findings of this experiment are discussed thoroughly in the fourth chapter of this dissertation.

Chapter 4

Co-relation between tone and voice qualities of Sylheti vowels

General Introduction

The foundation of canonical theories of tonogenesis is driven by the interaction between consonantal features and predicable pitch patterns of the following vowel. Generally voiced obstruents are expected to lower the f_0 of the following vowels whereas their voiceless counterparts may even raise it (Yip 2002, Hombert 1978, Maddieson 1977). However, such predictions do not seem to work in case of Sylheti where the loss of (underlying) breathiness contrast of Sylheti obstruents (both voiced and voiceless) have been compensated with higher f_0 in the adjacent vowel. This chapter therefore attempts to address the phonologized correlation between tone and voice qualities of vowels carrying contrasting tones. A rigorous experiment has been conducted to understand and analyze various acoustic components related to voice qualities of vowels. The findings of the experiment showed that the vowels associated with contrastive tones are indeed different in terms of their voice quality.

4.1 Phonation and tone

In order to maintain phonemic contrasts, the state of the glottis is exploited not only to produce voicing distinction (voiceless and voiced), but also to produce various degrees of phonation or voice quality. Phonation or voice quality refers to the production of speech sounds by the vibration of vocal folds (Ladefoged 1971, 1996, Gordon 2001, Wayland and Jongman 2002, Esposito 2010a, 2011, Huffman 1987). Ladefoged (1971) is of the opinion that depending on the variations of the glottal constriction continuum, degrees of phonation types could be determined. In his model he proposed that the size of the glottis might range from voiceless (when the vocal folds are held furthest apart), through breathy voice (where the glottis is held more open), to regular (modal voicing), to creaky (produced with a constricted glottis) and lastly to glottal closure (when the vocal folds are held closest together, hence no vibration and without phonation). Since the glottal constriction is a continuum, the degrees of creakiness, breathiness or even the modal voice might vary from more constricted to more open. The range of glottal positions, thus, represents three most common contrasting phonation[1] types - breathy, modal and creaky (Ladefoged 1971, Gordon and Ladefoged 2001, Esposito 2010). Following Ladefoged (1971) and Gordon and Ladefoged (2001) the model is reproduced in Figure 4-1:

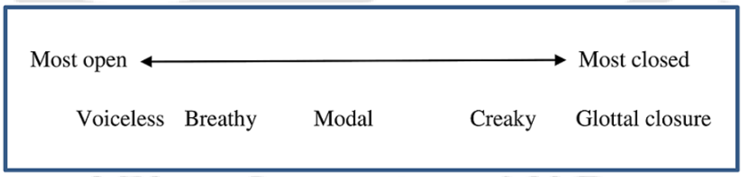

Figure 4.1: Continuum of Phonation types (after Ladefoged, 1971), reproduced from Gordon and Ladefoged (2001).

[1]Ladefoged (1971) further added that these phonation categories are the outcome of the controllable variations produced by exploiting the state of glottis, either by individual idiosyncratic efforts or due to unintentional pathological conditions and are not absolute; hence these might differ somewhat across languages and even across speakers of the same language. So for example, what counts as breathy voice for one speaker might count as modal for another in the same language.

Thus, voice quality distinctions or phonation types among phonemes are exploited in many languages of the world to preserve lexical contrasts. Languages can employ the voice quality distinctions on either consonants (obstruents) as in Hindi (Ohala 1983, Dixit 1989, Dutta 2007), or Maithili (Yadav 1984) or on vowels as in many Zapotec languages (Jones and Knudson 1977, Esposito 2010b, Esposito and Khan 2012). A very few languages maintain phonation contrast in both breathy vowels and voiceless aspirated consonants (as in Jalapa Mazatec, an Otomanguen language spoken in Mexico (Krik *et al.* 1993, Gordon2001), Suai (Abramson and Luangthongkum 2001)) and rarely in both breathy vowels and breathy voiced aspirated consonants as in White Hmong (Esposito 2012) and Gujarati (Dave 1967, Fischer 1967, Esposito and Khan 2012).

4.1.1 Acoustic Correlates of Phonation

Even though the basic acoustic display of waveforms and spectrograms can easily differentiate between a modal and nonmodal such as creaky (irregularly spaced glottal pulses and reduced intensity compared to modal) and/or breathy (lower f_0 relative to modal and by substantial aperiodic or noisy energy) (Gordon 2001), researchers have proposed a variety of acoustic properties suitable for measuring various phonation types. Among those, spectral measurements have been the most popular and consistent method of phonation used in various languages. such as Hmong which distinguishes breathy and modal vowels (Huffman 1987, Gordon 2001), Mazatec that differentiates between breathy, modal and creaky vowels (Silverman *et al.* 1995, Blankenship 1997), !Xóõ which also maintains a three way contrast among breathy, modal and strident (a third type of phonation) vowels (Ladefoged 1983, Ladefoged *et al.* 1988), Khmer that contrasts breathy and clear vowels (Wayland and Jongman 2002) and so on. Tilt is determined by the shape of glottal pulse and measures the difference between the amplitude of the first harmonic and the second or prominent harmonic of the first and a higher formants (such as (H1-H2), (H2-H4), (H1-A1), (H1-A2), and (H1-A3)). Among various tilt measurements, the difference between the amplitudes of the first and second harmonics (H1-H2) and the difference between the amplitudes of the first harmonics and third formants (H1-A3) have been used to draw the distinction in phonation types among various phonations in languages such as English (Stevens and Hanson 1995), Krathing Chong (Blankenship 2002), Takhian Thong Chong (DiCanio 2009), and so on. In the higher frequencies, spectral properties are expected to be weaker for breathy voice than that of a modal voice; however, it is stronger for a creaky voice. On the other hand, during the production of a breathy voice, the opening of the glottis might create an intense audible noise which is capable of distinguishing it from other phonation types.

In !Xóõ (Ladefoged 1983) the tilt component of H1-A1 effectively differentiates the phonation types, whereas, in Krathing Chong H1-A2 distinguishes the levels of phonations (along with the values generated from H1-H2 and H1-A3). Studies on phonation qualities that successfully distinguish the difference between a breathy, modal or creaky vowels noticed higher values (in dB) for breathy voiced, an intermediate values (in dB) for modal voice and relatively lower values (in dB) for creaky voice phonation types (Esposito 2006, 2010, Wayland and Jongmong 2003, Esposito and Khan 2012).

Thus it emerges from the literature that various measurements of spectral tilt that distinguish phonation types do not behave uniformly in any single language (Gordon and Ladefoged 2001, Esposito and Khan 2012). Other studies like that of Blankenship (1997) while investigating the correlation between contrastive tones and phonation types in Mpi found that the values of H1-H2 consistently distinguish phonation types on high tone vowels than on mid or low tone vowels. Esposito (2010b) found that in Santa Ana del Valle Zapotec, H1*-H2* (* marks represents the corrected amplitude values in the literature[2]) effectively distinguishes between breathy, modal and creaky phonation in female speech (H1*-H2* were found to be greater for breathy voiced (in dB), intermediate for modal (in dB) and a relatively lower value (in dB) for creaky vowels) but that did not show any consistency for the male speakers (instead H1*-A3*

[2]Hanson (1995) first proposed that while calculating spectral tilt or spectral balance, the amplitudes of harmonics can be corrected or normalized for the effects of the frequencies and bandwidths of adjacent formants. The concept was later developed and extended by many researchers such as Iseli *et al.* (2007), Shue *et al.* (2011) and so on. The corrected amplitude values are generally represented by an asterisk (*) mark (Esposito 2010, Esposito and Khan 2012).

[or the overall spectral tilt] showed highest dB values for breathy vowels, followed by modal and negative values for creaky vowels, thus distinguished each phonation types in expected directions). Esposito (2010b) further reported that the correlation between f_0 and phonation types do not always behave uniformly, instead it is lower and mid-range f_0 that control the phonation qualities of the vowels. Studies also showed that H1*-H2* was positively correlated with f_0 only for the speakers whose pitch was lower than 175 Hz (Iseli *et al.* 2007, Esposito and Khan 2012).

Gordon and Ladefoged (2001) further noted that non-modal phonation types are generally associated with lowering of fundamental frequency. For example, in a language such as Mam (England 1983) and many Northern Iroquoian languages such as Mohawk, Cayuga and Oneida (Chafe 1977, Michelson 1983, Doherty 1993), the creaky phonation is associated with a lower f_0 (relative to modal phonation). This lowering effect of creaky voice, however, is not universal across languages. Hombert *et al.* (1979) showed that the process of glottalization can be associated with high tone in the historical development of some of the Athabaskan languages, while the same glottalization process is associated with a low tone in closely related languages (Leer 1979, Gordon and Ladefoged 2001, Kingston 2009). Breathy phonation, on the other hand is more consistently connected with a lower f_0 in the majority of languages (Gordon and Ladefoged 2001). To understand phonation related qualities of the vowels associated with contrastive tones, an experiment has been conducted to examine various spectral components. In the following section we will discuss the details of the experimental design, the methods adopted for the acoustic analysis, and the details of statistical tests and procedure employed for the current experiment.

4.2 Experimental Procedure

Speech material and Subjects

The results of the production experiment discussed in the previous chapter of this dissertation revealed that the loss of (underlying) breathy voice contrast [+spread glottis] among the obstruents (both voiced and voiceless) in Sylheti gave birth to the high tone ([d̪an] ‘paddy’ [d̪^{h}an > [d̪] and [d̪an] ‘donate’, [bat̪] ‘rice’ ([b^{h} > b]), and [bat̪] ‘arthritis’)- thus creating a two way lexical contrast in Sylheti.For the current experiment,twenty monosyllabic words (comprising the vowel [a]) were considered (Table 4-1). Note that these words were also part of our production experiment (along with 4 other additional words). However, since various studies have shown that high first formant of the vowel [a] minimizes the effects on first and second harmonics (Bickley 1982, Ladefoged 1983, Huffman 1987, Ladefoged *et al.* 1988, Krik *et al.* 1993, Blankenship 1997, Gordon 2001), we decided to include the homophonous pairs containing the vowel [a] only. In this experiment we have considered data from 9 native speakers of Sylheti (7 male, 2 female) from Dharmanagar district of north Tripura. Note that same data was also used for analyzing tone in Sylheti. The research question addressed here is that, if these vowels carrying contrastive tones at all differ and if they do, then do they differ in terms of their voice quality, and if so, how do their distinct phonation quality correlate with the contrastive tones.

Sylheti words	Gloss	Sylheti words with breathiness	Gloss
[zào]	'gruel'	[záo]	'tamarisk'
[gà]	'body'	[gá]	'wound'
[gài]	'cow'	[gái]	'stroke'
[bàt̪]	'arthritis'	[bát̪]	'rice'
[dàx]	'roaring of cloud'	[dáx]	'drum'
[bàn]	'tie'	[bán]	'pretend'
[d̪àn]	'donate'	[d̪án]	'paddy'
[xàl]	'skin'	[xál]	'channel/drain'
[xuà]	'well'	[xuá]	'jackfruit cell'
[zàl]	'net'	[zál]	'chilly hot'

Table 4.1: The dataset displaying the list of monosyllabic words considered for the experiment

4.3 Measurement

To understand the tonal property of Sylheti tones a production experiment was conducted (see Chapter 3) and all the target words. The same .wav files along with their TextGrid files were used in the current experiment. Acoustic measurements of the target words over time were taken automatically using the software VoiceSauce . The software is run as an extension through Matlab. Corrections of the harmonic amplitudes were done automatically in VoiceSauce by using measured formants frequencies and bandwidths estimated by those frequencies. The correction procedure in VoiceSauce[3] is similar to the method used by Hansen (1995) which was later extended by Iseli *et al.* (2007). All the spectral measurements of every vowel were estimated at a 30 ms window at 2 ms intervals. The following spectral measurements were considered for the current experiment:

[3]Shue *et al.* 2011.

1. Amplitude of the first harmonic (H1*);
2. the difference between the amplitude of first and second harmonic (H1*- H2*);
3. amplitude of the second harmonic minus the amplitude of the fourth harmonic (H2*- H4*);
4. difference between the amplitude of the first harmonic and the amplitude of the first formant peak (H1*- A1*);
5. difference between the amplitude of the first harmonic and the amplitude of the second formant peak (H1*- A2*), and;
6. difference between the amplitude of the first harmonic and the amplitude of the third formant peak (H1*- A3*);

Mean value of each of these acoustic components of each subject was used to plot the bar diagrams for visual inspection of the words with contrastive tones. Separate statistical measurements were conducted using SPSS (version 20) to test the significance of each of the acoustic parameters for each subject. Altogether, 540 tokens were examined for the current experiment (20 words * 3 repetitions * 9 subjects).

4.4 Statistical tests

It is reasonable to assume the null hypothesis that the sample distribution is not different from normal distribution of the population. In order to verify this assumption, we conducted a Shapiro-Wilk test on the data of each of the acoustic parameters and for every speaker separately. The Shapiro-Wilk test confirmed that our data is not normally distributed (alternative hypothesis), and as such it rejects the null hypothesis. In order to account for the absence of a normal distribution, we conducted a non-parametric Maan-Whitney U test separately for each subject and for each acoustic component to compare and test the significance of various spectral measurements of the vowels with contrastive tones. Normal Q-Q plots are shown below for H1*- A1*

[Figure 4-2 and Figure 4-3] and H1*- H2* [Figure 4-4 and Figure 4-5]. These figures show that our data sample is not normally distributed.

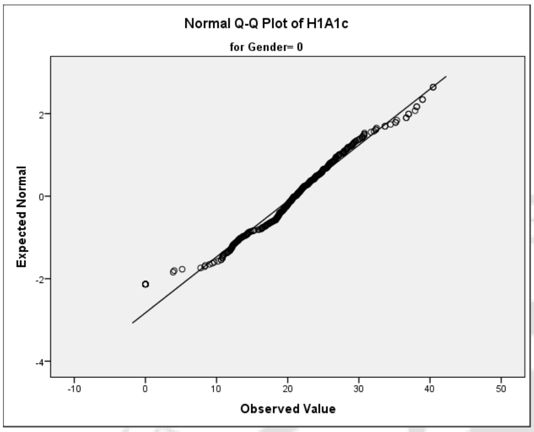

Figure 4.2: Female speaker

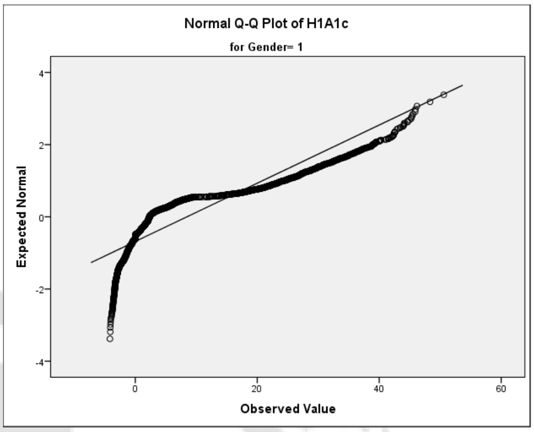

Figure 4.3: Male speaker

Figure 4-2 and Figure 4-3: Normal Q-Q plot showing the distribution of the values of the H1*-A1*. Gender= 0 ($p < .05$) represents the data of a female subject, whereas Gender=1 ($p < .05$) stands for the data of a male subject. H1A1c is the corrected value generated by VoiceSauce which is normally represented with an asterisk (*) such as H1*- A1*.

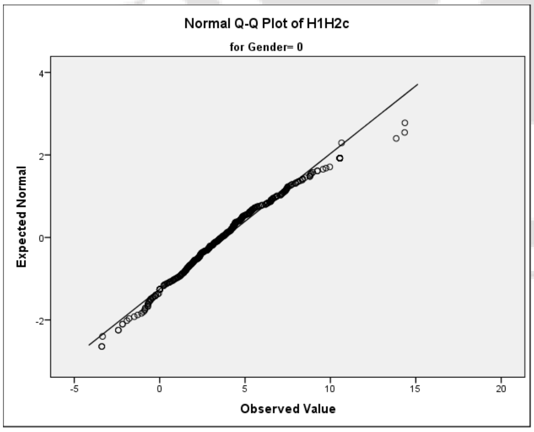

Figure 4.4: Female speaker.

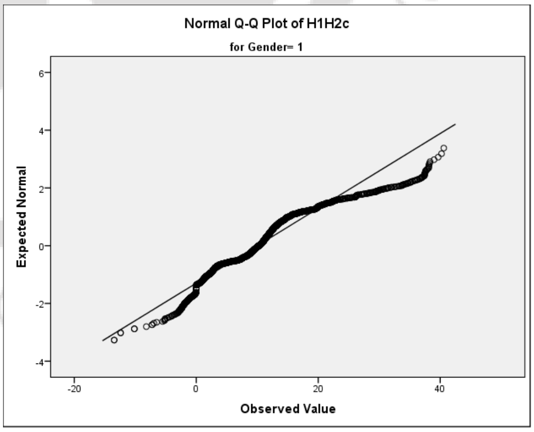

Figure 4.5: Male speaker.

Figure 4-4 and Figure 4-5: Normal Q-Q plot showing the distribution of the values of the H1*-H2*. Gender=0 ($p < .05$) represents the data of a female subject, whereas Gender=1 ($p < .05$) stands for the data of a male subject. H1H2c is the corrected value generated by VoiceSauce.

4.5 Results

As the Shapiro-Wilk test confirmed non-normalized distribution of a large amount of data(since all the spectral measurements of each vowel were estimated at a 30 ms window at 2 ms intervals and for each speaker separately),a non-parametric Mann-Whitney U test has been adopted for each subject separately. Mann-Whitney U test has been conducted using tone-type as factor and various tilt components such as *H1, H1*- H2*, H2*- H4*, H1*- A1*, H1*- A2*, H1*- H3* as dependent variables to examine the significance difference (if any) among the vowels carrying contrastive tones. Since we have a relatively large population, we considered mean values of the token instead of median values (this was also done due to the fact that, we observed identical values of mean and median for the sample of individual speakers).

(a) The amplitude of the first harmonic (H1*) The mean amplitude values of the first harmonic (H1*, measured in dB) of the vowels associated with contrastive tones are shown in Figure 4-6. The values of H1*is shown in the vertical axis, while x-axis represents each subject (M= male, F=female). Results indicate that the vowels carrying L tone do bear higher H1* values (in dB)compared to the vowel associated with H tone. This trend is observed across all the speakers.The difference of H1* values between the vowels carrying contrastive tones is significant for speaker 1 (U = 258, Z = -4.06, p = 0.00 [two-tailed]), speaker 2 (U = 9014, Z = -6.02, p = 0.00 [two-tailed]), speaker 5 (U = 3574, Z = -.71, p = 0.05 [two-tailed]), speaker 6 (U = 3373, Z = -2.88, p = 0.01 [two-tailed]), speaker 7 (U = 4240, Z = -6.68, p = 0.00 [two-tailed]), speaker 8 (U = 8531, Z = -2.31, p = 0.02 [two-tailed]), and speaker 9 (U = 1752, Z = -6.84, p = 0.00 [two-tailed]) respectively. However, the data from speaker 3 and speaker 4 do not show any statistical differences.

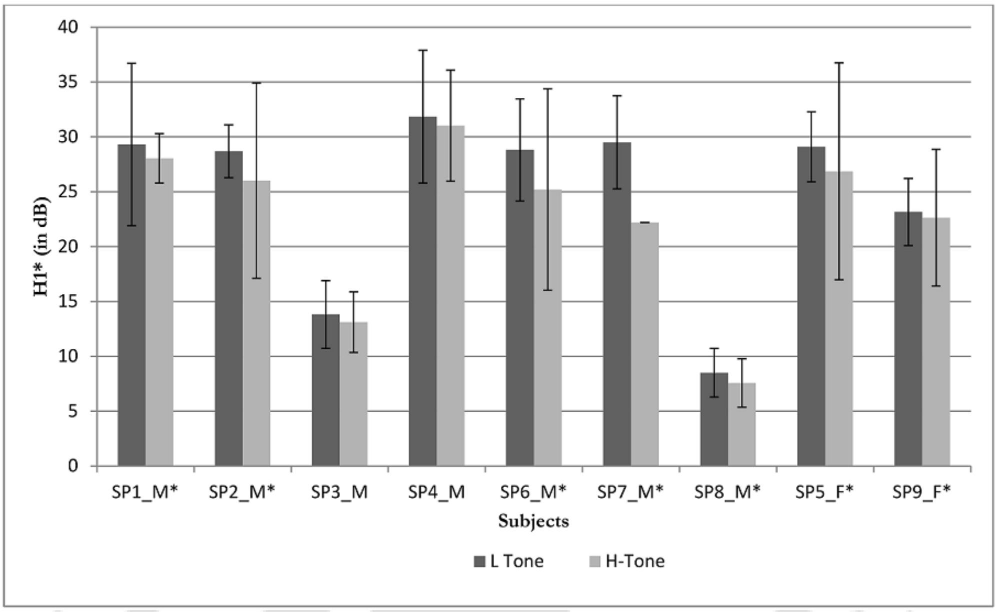

Figure 4.6: Shows the comparison of the mean amplitude values (with standard deviation as error bars) of first harmonic (H1*) (measured in dB) of the vowels carrying contrastive tones,measured for each speaker separately (M=male, F= female). The lightgray scale shows the values of the vowels associated with high tone (H Tone) and the dark gray scale shows the values of the vowels associated with low tone (L Tone).Significance among the contrastive tones for each speaker is marked with an asterisk.

(b) The difference between the amplitude of first and second harmonic (H1 - H2*) Like the values of H1*, the difference between the amplitude of the first and second harmonics (H1*- H2*) was also found to be higher for the vowel associated with a lower fundamental frequency (Figure 4-7). According to Mann-Whitney U test this difference is significant for speaker 2 (U = 12515.5, Z = -2.35, p = 0.02 [two-tailed]), speaker 3 (U = 12737, Z = -.74, p = 0.02 [two-tailed]), speaker 4 (U = 11821, Z = -5.34, p = 0.00 [two-tailed]), speaker 5 (U = 3489, Z = -.96, p = 0.03 [two-tailed]), speaker 7 (U = 4228, Z = -6.7, p = 0.00 [two-tailed]), speaker 8 (U = 76842, Z = -2.61, p = 0.00 [two-tailed]), and speaker 9 (U = 2006, Z = -6.15, p = 0.00 [two-tailed]) respectively. No significant difference of H1*- H2* values was observed among the vowels with contrastive tones for speaker 1.Interestingly, data from speaker 6 presents a reverse pattern where the vowels carrying high tone bear higher values than the vowels associated with low tone and the difference was found to be significant (U = 3618, Z = -2.24, p = 0.03 [two-tailed]).

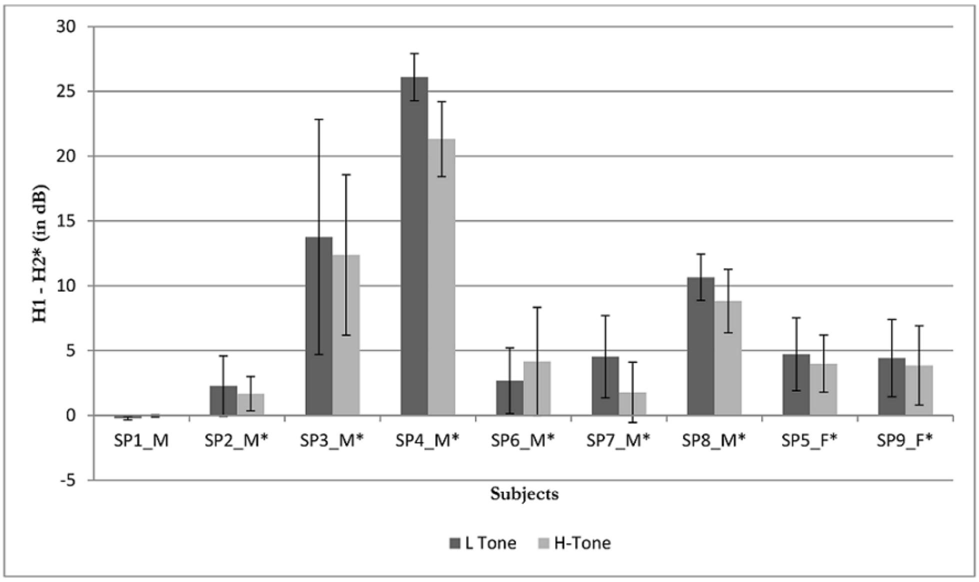

Figure 4.7: Shows the comparison of the mean values (with standard deviation as error bars)of (H1*- H2*) (measured in dB) of the vowels carrying contrastive tones,measured for each speaker separately (M=male, F= female). The lightgray scale shows the values of the vowels associated with high tone (H Tone) and the dark gray scale shows the values of the vowels associated with low tone (L Tone).Significance among the contrastive tones for each speaker is marked with an asterisk.

c) The difference between the amplitude of second and fourth harmonic (H2*- H4*) The difference between the amplitude of second and fourth harmonics also shows higher values for the vowels carrying low tone (for all the speakers) than the vowels associated with high tone (Figure 4-8). Mann-Whitney U test confirms this difference to be significant for speaker 1 (U = 301, Z = -3.55, p = 0.00 [two-tailed]), speaker 3 (U = 9947, Z = -3.15, p = 0.02 [two-tailed]), speaker 4 (U= 8122, Z = -8.89, p = 0.00 [two-tailed]),speaker 6 (U = 2180, Z = -5.95, p = 0.00 [two-tailed]), speaker 7 (U = 6447, Z = -3.14, p = 0.02 [two-tailed]), and speaker 8 (U = 87386, Z = -2.62, p = 0.01 [two-tailed]) respectively. No significant difference of H2*- H4* was observed among the vowels with contrastive tones for speaker 2, speaker 5, and speaker 9(even though the vowel with L tone exhibited higher H1*- H2* values (in dB)).

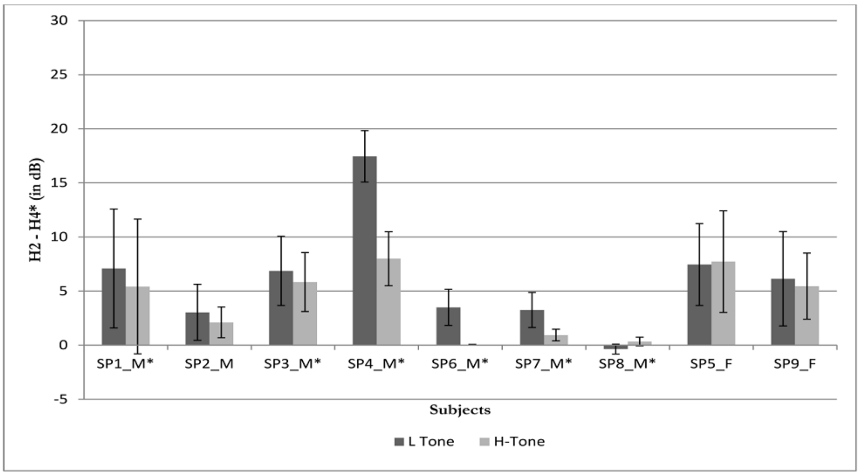

Figure 4.8: Shows the comparison of the mean values (with standard deviation as error bars)of (H2*- H4*) (measured in dB) of the vowels carrying contrastive tones,measured for each speaker separately (M=male, F= female). The lightgray scale shows the values of the vowels associated with high tone (H Tone) and the dark gray scale shows the values of the vowels associated with low tone (L Tone).Significance among the contrastive tones for each speaker is marked with an asterisk.

d. Difference between the amplitude of the first harmonic and the amplitude of the first formant peak (H1*- A1*)

The difference between the amplitude of the first harmonic and the amplitude of the first formant peak do not show a uniform pattern (Figure 4-8). The data also indicates that the from majority of the subjects (speaker 2, speaker 5, speaker 6, speaker 7, speaker 8, and speaker 9) have higher values associated with the vowels carrying low tone. According to Mann-Whitney U test this difference is significant for speaker 5 (U = 2995, Z = -2.42, p = 0.01 [two-tailed]),speaker 7 (U = 5478, Z = -4. 7, p = 0.00 [two-tailed]), speaker 8 (U = 28810, Z = -18.09, p = 0.00 [two-tailed]), and speaker 9 (U = 4147, Z = -.264, p = 0.04 [two-tailed]) respectively. On the other hand, the value of H1*- A1* seems to be higher for the vowel associated high tone for speaker 1 (U = 386, Z = -2.54, p = 0.01 [two-tailed]), speaker 2 (non-significant), speaker 3 (non-significant), and speaker 4 (non-significant).

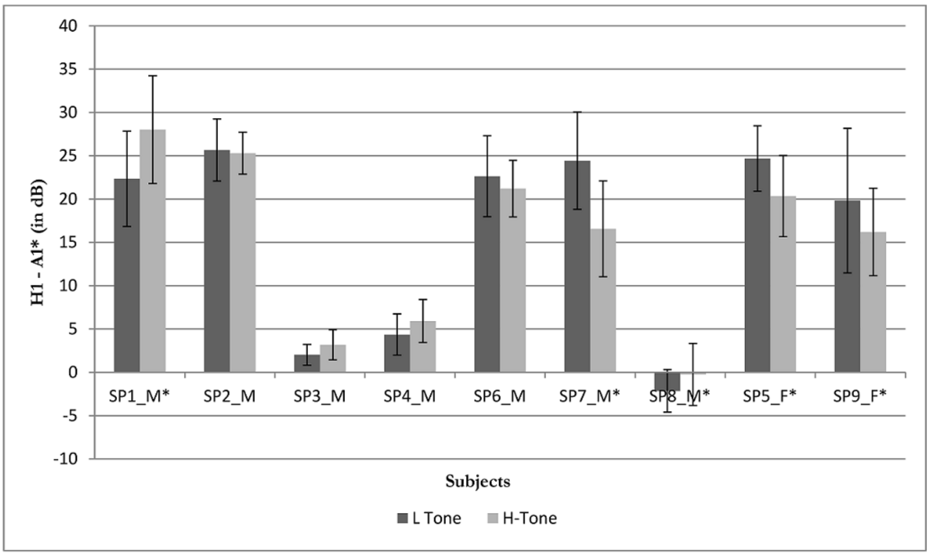

Figure 4.9: Shows the comparison of the mean values (with standard deviation as error bars)of (H1*- A1*) (measured in dB) of the vowels carrying contrastive tones,measured for each speaker separately (M=male, F= female). The lightgray scale shows the values of the vowels associated with high tone (H Tone) and the dark gray scale shows the values of the vowels associated with low tone (L Tone).Significance among the contrastive tones for each speaker is marked with an asterisk.

e. The difference between the amplitude of the first harmonic and the amplitude of the second formant peak (H1*- A2*)

The values of H1*- A2* also show a similar trend, that is higher values for low tone vowels (for most of the subjects) than their contrastive counterparts (Figure 4-9). Mann-Whitney U test confirms this difference to be significant for speaker 1 (U = 79411, Z = -4.72, p = 0.00 [two-tailed]) speaker 2 (U= 9142, Z = -5.89, p = 0.00 [two-tailed]),speaker 5 (U = 2692, Z = -3.32, p = 0.00 [two-tailed]), speaker 6 (U = 3234, Z = -3.24, p = 0.01 [two-tailed]), and speaker 7 (U = 2413, Z = -1.27, p = 0.00 [two-tailed]) respectively. Data from speaker 4 (Significant: U = 12333, Z = -4.84, p = 0.00 [two-tailed]),and speaker 8 (non-significant) show a reverse pattern, that is, higher values for the vowels carrying high tones.

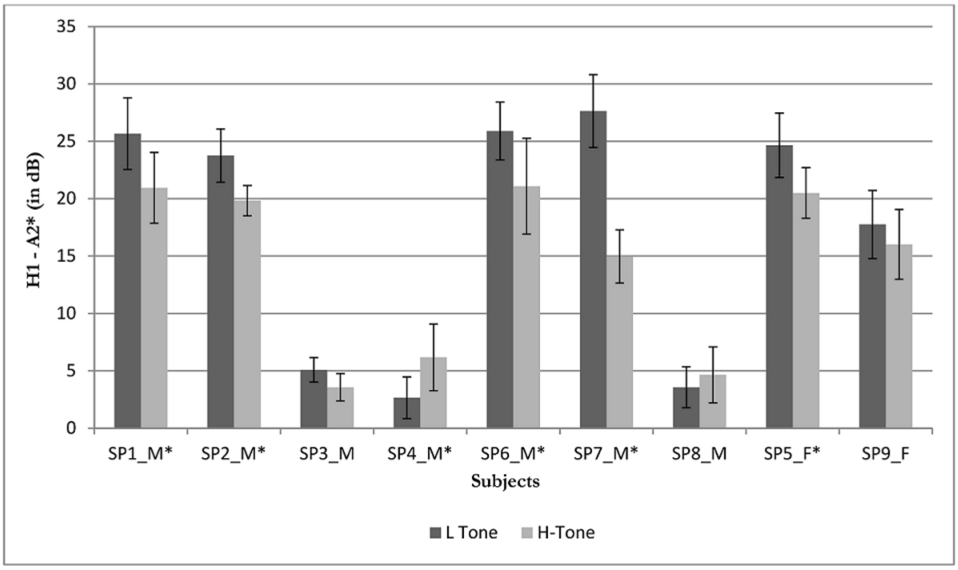

Figure 4.10: Shows the comparison of the mean values (with standard deviation as error bars)of (H1*- A2*) (measured in dB) of the vowels carrying contrastive tones,measured for each speaker separately (M=male, F= female). The lightgray scale shows the values of the vowels associated with high tone (H Tone) and the dark gray scale shows the values of the vowels associated with low tone (L Tone).Significance among the contrastive tones for each speaker is marked with an asterisk.

f. The overall spectral tilt (H1*- A3*)

The overall spectral tilt shows higher values for the vowels carrying low tone for all the speakers (Figure 4-10). According to Mann-Whitney U test this difference is significant for speaker 2 (U= 14372, Z = -.397, p = 0.69 [two-tailed]), speaker 3 (U = 8467, Z = -4.42, p = 0.00 [two-tailed]), speaker 5 (U = 2074, Z = -5.15, p = 0.00 [two-tailed]), speaker 7 (U= 7620, Z = -1.23, p = 0.04 [two-tailed]), speaker 8 (U = 57195, Z = -10.59, p = 0.00 [two-tailed]), and speaker 9 (U = 2797, Z = -3.97, p = 0.00 [two-tailed]) respectively. Data from speaker 1, speaker 4, and speaker 6 do not show any significant difference.

The findings of acoustic components examined to relate the phonation qualities suggest that the vowels carrying contrastive tones are undeniably different so far as their voice qualities is concerned. The results of the current experiment (as shown above) reveal positive and higher values for the vowels carrying low tones. On the

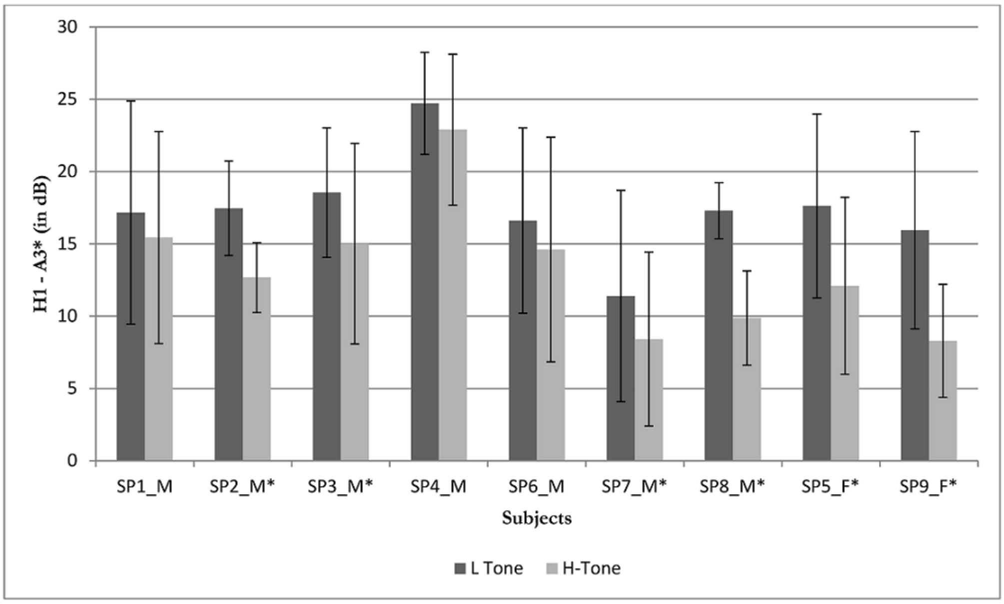

Figure 4.11: Shows the comparison of mean values (with standard deviation as error bars)of (H1*- A3*) (measured in dB) of the vowels carrying contrastive tones,measured for each speaker separately (M=male, F= female). The lightgray scale shows the values of the vowels associated with high tone (H Tone) and the dark gray scale shows the values of the vowels associated with low tone (L Tone). Significance among the contrastive tones for each speaker is marked with an asterisk.

other hand, the vowels carrying high tone are associated with relatively low and positive values. The literature on phonation types generally associate the greater and positive values (in dB) with breathy vowels, an intermediate value for modal vowels and less and often negative values for a creaky vowels generated from various spectral measurements (Stevens and Hanson 1995, Blankenship 2002, DiCanio 2009, Waylad and Jongman 2002, Esposito 2010, Esposito and Khan 2012). However, the data recorded from the 9 native speakers of Sylheti do not suggest a clear evidence of breathy voice qualities of the vowels carrying contrastive tones (we did not hear any breathiness); a few tokens from three native speakers seem to be slightly creaky, even though the acoustic signal did not suggest much. The presence of consistent higher spectral values associate with the vowels bearing low tones (and also due to the absence of probable audible noise indicating the absence of actual 'breathiness'), confirmed that the vowels bearing low tones have slightly 'laxer' voice qualities than those bearing high tones (which seem to have 'tenser' voice qualities).

4.6 Discussion: Sylheti tonogenesis

Many researchers have discussed the relationship between tonogenesis and consonant types (Hombert 1978, Ladefoged 1971, Thurgod 2002, Gordon and Ladefoged 2002, Gordon 2002, Svantesson and House 2006, Kingston 2011, and so on). In this regard, Hombert (1978) argued that an intended feature of a sound may have intrinsic effects on neighbouring sounds. Thus, in course of the historical development of a language when a particular feature of a specific sound is lost, the speakers tend to reinterpret (or rather compensate) the lost feature (or the original intended feature) as intrinsic feature in the neighbouring sound. When the intrinsic features are reinterpreted as the main features and the original feature is lost (the most widely cited cases are that of Vietnamese, where the loss of voicing distinctions led to tone (check) and Punjabi where the loss of aspiration led to the development of tone), the result is likely to be the usual source of tonogenesis (Hombert 1978, Bhaskararao 1998).

In order to elucidate the mechanism of 'sound change', Ohala (1993) proposed two phonetically motivated mechanism- *hypo-correction* and *hyper-correction*. Sound changes related to hypo-correction such as regressive assimilation (Ohala 1990) or tonogenesis (where tonal contrast surfaces due to prior and gradual loss of voicing contrast) occurs when a listener fails to normalize the perturbations in the speech signal and ends up taking/judging the signal at face value. This process might lead to a difference in the conception of that speech signal by the listener from the intended (original) one. Hyper-corrective sound change (such as labialization, retroflexion, glottalization), on the other hand refers to contexts where a listener not only fails to normalize deviation in production but may also overanalyze and erroneously attribute an intended phonetic cue as contextual (Ohala 1990).

In this context, let us recall the process of tonal development in Punjabi from breathy voiced consonants. In Punjabi, the BHV sequence became voiceless unaspirated leaving a low tone on the vowel. Notice that a two-way loss of the breathy voiced consonant is compensated in the neighbouring vowel, resulting in a low tone. Other studies have also shown that breathy voiced consonants are stronger f_0 depressor (Ladefoged 1971, Kingston 2011). However the depression of f_0 following the loss of breathy voiced consonant is hardly common let alone a universal phenomenon.

In case of Sylheti, we observed two way tonal contrasts (the status of third tone requires significant evidence in terms of phonetic and phonological properties): the emergence of high tone in the vowels following the loss of an intrinsic (breathy voice) property [+spread glottis], and a low tone elsewhere.

To account for a similar kind of tonal phenomena i.e. the emergence of a high tone following the feature [+spread glottis] in NakhornSithammarat Thai, Kingston (2011) drew a distinction between the features [spread glottis] and [constricted glottis]- "In Yung-chiangKam, higher tones have developed after [constricted glottis] than [spread glottis] consonants, while in NakhornSithammarat Thai it is the reflexes after [spread glottis] consonants that are higher. These outcomes are referred to henceforth as "Constricted High" and "Spread High" splits" (Kingston 2011). He has argued that "either [spread glottis] or [constricted glottis] consonants first induced higher tone reflexes in the three way splits (as found in languages such as Yung-chiangKam, NakhornSithammarat and Szu ta chai), and the higher reflexes later exchanged values with the lower tones" (as explained in Kingston 2011). Since both [spread glottis] and [constricted glottis] consonants are capable of either raising or lowering f_0, then, it is left with the choice of the native speakers of the individual language which feature to retain. The development of high tone as a reflection of glottal constriction (or due to a process of glottalization) is reported in many Athabaskan languages (Kingston 2011, Gordon and Ladefoged 2002).

Under the purview of the above discussion, we therefore propose a two-fold evolution of Sylheti tonogenesis- first, the historical development of the loss of intended feature [+spread glottis] associated with (voice) obstruents is indeed reinterpreted and readjusted with a perturbed f_0 in the following vowel(s) (in order to maintain the lexical distinctions among the likely homophonous words); secondly, due to hypo-correction-a linguistic phenomenon which may occur in historical processes as proposed by Ohala (1993), where the listeners probably fail to normalize the coarticulatory effects such as the effects of [+spread glottis] on f_0, the vowels following the obstruents (with the history of an underlying breathiness property)might have been compensated with a perturbed f_0. Since the breathiness property [+spread glottis] of the consonants (obstruents) was readjusted (hence reinterpreted) on the adjacent vowels, the vowels might have acquired a 'tenser' phonation quality in order to maintain the lexical distinction of homophonous words. An apparent consequence of the loss of intended property (underlying breathiness) led to the development of a high tone (since the conditioning environment wasn't completely lost that resulted in hypo-correction). Once the conditioning environment (the underlying breathiness property [+spread glottis] of the obstruents) was lost, the f_0 patterns were phonologized for the homophonous words with contrastive lexical tones. Therefore, the development of tone in Sylheti is not entirely phonetic but perceptual[4]! The schematic representation of Sylheti tonogenesis is represented below (Figure 4-12):

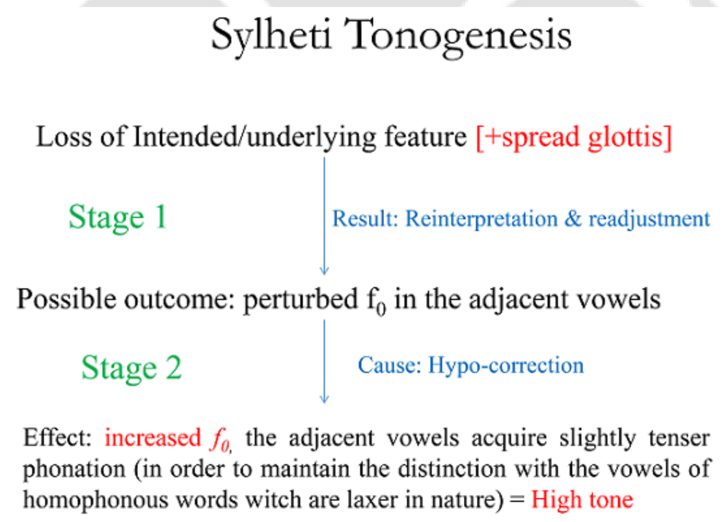

Figure 4.12: Schematic representation of Sylheti tonogenesis

There is an alternative argument too: perhaps f_0 was initially raised after voiceless

[4]Thus the results indicate that the environment that conditioned the tonal contrast is retained to some small but measurable extent in the phonation of the tone bearing vowel.

aspirated stops, and once that raising was consistent, it was extended to the voiced [+spread glottis] consonants, so that the entire class behave alike. This alternative argument is supported by the fact that the voiceless aspirated stops have undergone more changes, spirantization as well as loss of [+spread glottis], while the voiced aspirated stops have just lost [+spread glottis] specifications.

It is interesting to note that Sylheti continues to have a long history of coexisting with other Tibeto-Burman languages (such as various dialects of Kok-Borok, Reang which are tonal in nature). Even though there is no clear evidence of direct borrowing of lexical items from those tonal languages into Sylheti, there is still a narrow-window of possibility that the emergence of Sylheti tones is due to an areal feature as the indigenous speakers of Tibeto-Burman group of languages by and large use Sylheti as a common medium for interaction. It is also possible that in the course of historical language shift many speakers of tonal languages started using Sylheti for wider communication thereby introducing tonal distinctions. All of these are only speculations about the development of tone in Sylheti. During our multiple field visits we have observed that the current generation of Sylheti native speakers however, tends to attach an underlying tone with each and every lexical entry. What it means is that the realization of high tone in Sylheti is not due to (limited/restricted) the loss of underlying breathiness property [+spread glottis] only- a high tone (or contrastive tone) can also appear in a homophonous pair with similar phonological changes For example, in the monosyllabic series of [**ɸɔr**] meaning 'read' and 'guard' exhibited similar phonological changes; that is, the onset voiceless bilabial stop [p] is spirantized to voiceless bilabial fricative [**ɸ**] ([p > **ɸ**]). Thus it could be guessed that both these words would be realized with similar pitch height. However, the normalized pitch contours of these words ('read' and 'guard') exhibited the presence of distinctive tonal contrasts (Figure 4-12). Similarly, in the case of the monosyllabic pair of [xal] series meaning 'drain' [k^h > x] and 'skin' [$tʃ^h$ > x]; with almost similar phonological changes demonstrate the presence of two way tonal contrasts. In case of the first word 'drain' the voiceless aspirated velar stop [k^h] in word initial position exhibits two way of phonological development (loss of underlying breathiness property and spirantization) and is marked with a high tone, similarly the voiceless aspirated velar affricate [tʃʰ] in the word initial position also may have undergone two phonological developments- loss of underlying breathiness property and deaffrication and as a result of which it has surfaced with a contrastive low tone (Figure 4-13). We have also noticed instances of contrastive tones on homophonous pair with no change (historical and/or phonological) at all as could be seen in the disyllabic pair of [mela] series (Fig 4-15) meaning 'fair' (specified with a low tone) and 'much/many' (specified with a high tone).

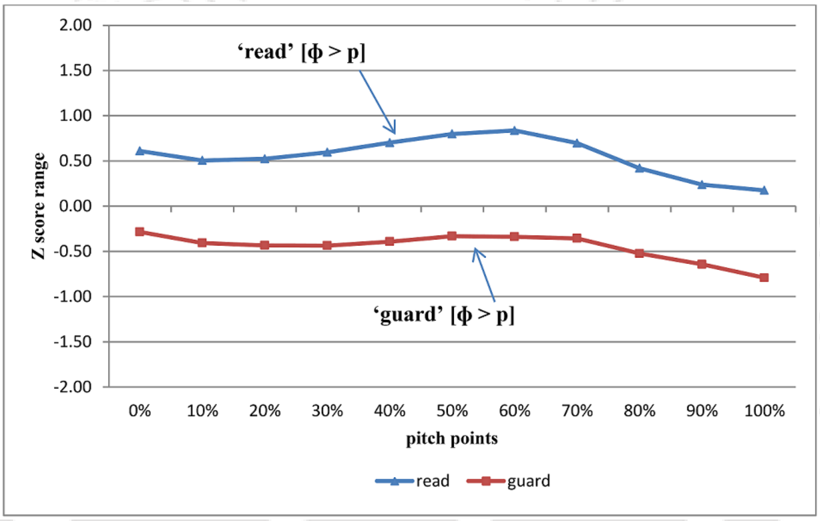

Figure 4.13: Normalized pitch tracks for [ɸɔr] series (n=18 for each word)

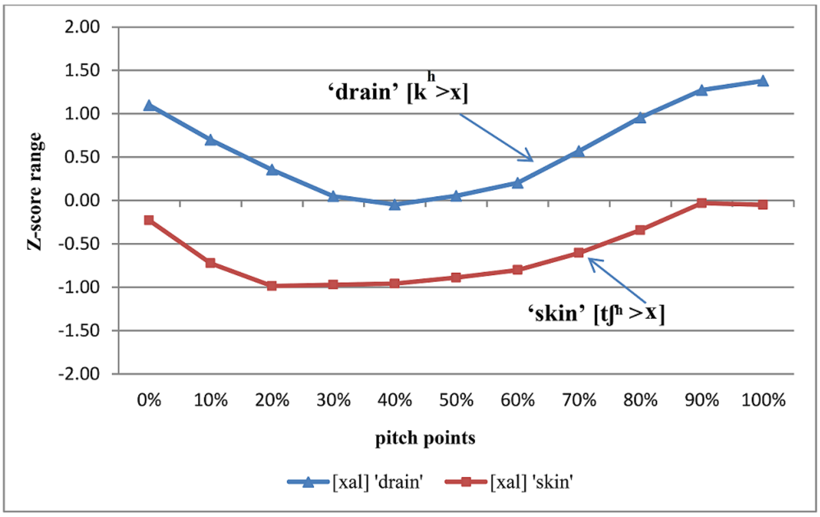

Figure 4.14: Normalized pitch tracks for [xal] series (n=24 for each word)

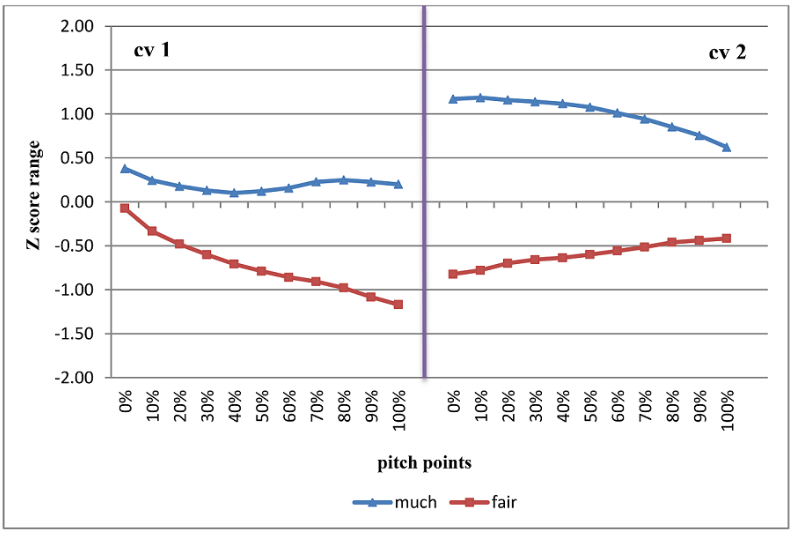

Figure 4.15: Normalized pitch tracks for [me.la] series (n=24 for each word)

We conclude that the lexical tonal contrast is emerged through series of historical developments and phonologized so as to every lexical entry is assigned with an underlying tone. In the next chapter, with the help of synthesized stimuli we will examine how the native speaker of Sylheti perceives tone.

Chapter 5

Perception of Lexical Tones

Introduction

Tone as a phonological category is capable of distinguishing the meanings of two lexical words (or utterances) by varying the pitch height and/or contour differences. The findings of the production experiment on Sylheti tone as discussed in the third chapter of this dissertation have confirmed the existence of distinct tonal contrast in this language. A series of historical developments such as the loss of breathy voice contrast predominantly led to a high tone in Sylheti. To understand the phonation related qualities of the vowels associated with the contrastive tones, we examined various spectral components of the vowels (Chapter 4). The findings of acoustic components examined to determine the phonation qualities suggest that the vowels carrying contrastive tones are undeniably different as per their voice qualities are concerned- the vowels associated with high tone are in the continuum of modal to creakiness; whereas, the vowels carrying a low tone are indeed modal in nature. It is argued that the nature of tonal system of a given language must 'converge with fundamental processes associated with the production and perception of tones' (Gandour 1979). Keeping that in mind, the present chapter aims to build upon the results of the production experiment using a systematic perception experiment.

5.1 Perception of tone

Generally it is assumed that listeners do employ and exploit phonetic and/or phonological features associated with a speech signal (at least in part) in order to perceive and distinguish that signal from another one. According to Studdert-Kennedy (1975), "perception entails the analysis of the acoustic syllable, by means of its acoustic features, into the abstract perceptual structure of features and phonemes that characterize the morpheme". There have been numerous attempts by many researchers to understand and determine the nature of perceptual components employed by speakers to distinguish speech sounds vowels (Abramson 1979, Hillenbrand and Nearey 1999, Hillenbrand *et al.* 1995, Verbrugge *et al.* 1976, Terbeek 1977, Miller 1989, Johnson 1997), consonants (Miller and Liberman 1979, Harris 1958, Nittrouer 2002, Jongman 1989, Kurowski and Blumstein 1984, Repp and Svastikula 1988, Singh 1975), and tones (Gandour and Harshman 1978, Gandour 1978, 1979, Francis *et al.* 2003, Lin and Repp 1989, Chang 2012).

Researchers however, often do vary in their opinions and analysis about the way tone is perceived by the native speakers of a particular language. While it is widely accepted that the consonants (especially the stop consonants) are mostly perceived in categorical manner (Liberman *et al.* 1961, Liberman *et al.* 1957), the perception of vowel contrasts does not seem to follow the same trend (Abramson 1976, Stevens *et al.* 1969). The idea of *'categorical perception'* was proposed by Lieberman et.al (1957). The concept explains the way listeners deal with the perception of equally spaced stimuli with many-to-one mapping. The categorical perception of speech supports the idea that the perception of two tokens with comparable acoustic variations depends on whether these tokens are heard as the members of same or different categories. Thus, categorical perception is dependent on listeners capability of identifying or discriminating a particular token belonging to a specific category of speech sounds- "when stimuli are perceived categorically, equivalent acoustic differences between two tokens are treated differently, depending on whether the two tokens are heard as members of the same category or as members of different categories. Two members of one category are less discriminable than are two tokens from two different categories with an equivalent acoustic difference between them. The idea is that, through experience with a given language, listeners learn the location of specific category boundaries along various acoustic continua. By increasing discrimination accuracy across these boundaries and/or reducing it within boundaries, listeners improve their ability to hear two acoustically similar but not identical members of one category as *the same* and, conversely, improve their ability to hear two acoustically similar members of distinct categories as *different*" (Francis *et al.* (2003). Wang (1976), while studying the perception of Mandarin Chinese observed that native speakers do differentiate high rising tone (Tone 2) from a high level tone (Tone 1) in a categorical manner. Abramson (1979), on the other hand, reported that Thai level tones are not perceived categorically. Interestingly, Francis *et al.* (2003) observed that contour tones in Cantonese are perceived in categorical manner, however, the level tones in Cantonese do not follow the same pattern. This chapter examines the properties of perceptual cues of Sylheti tones and investigates whether tones in Sylheti are perceived in a categorical manner or not.

In a tone language, it is equally important for a hearer to perceive the f_0 fluctuations (and the phonetic and phonological properties) associated with distinct speech signals that distinguish the lexical meanings (Yip 2002, Francis *et al.* 2003). To qualify as distinct tones, speech signals of homophonous words must contain large enough f_0 fluctuations between them so that the hearer can use the pitch difference as a cue to distinguish the lexical meanings of the homophones. The foremost acoustic element that distinguishes tones is their f_0 differences; duration, amplitude and the voice quality of the vowels are also considered asother important (secondary) acoustic cues in a tone language (Lin and Repp 1989, Yip 2002, Gandour 1974, 1978, Gandour and Harshman 1978).

5.2 Perception experiment

The outcome of the instrumental experiment of Sylheti tone (examined in Chapter 3) have confirmed and identified the distinct f_0 of homophonous words as the (only) acoustic element distinguishing the lexical meaning (for both monosyllabic and disyllabic words). No substantial difference was observed in terms of duration and intensity values of the homophonous pairs (both monosyllabic and disyllabic pairs). A perception experiment was therefore designed to understand and determine whether listeners (really) use the f_0 fluctuations (only) associated with the vowels (the rhyme or the voiced part of a syllable) of contrastive tones in identifying and/or discriminating the lexical meaning of contrastive pairs of lexical items. Since no significant difference was found for duration (for example, $p > 0.05$, [$F(1, 414) = 1.21$, $p = 0.27$ [one-way ANOVA of monosyllabic words]) and intensity (for example, $p > 0.05$, [$F(1, 414) = 2.19$, $p = 0.14$ [one-way ANOVA of monosyllabic words]) of the syllables carrying contrastive tones, we decided to exploit only the f_0 properties of the contrastive pairs instead. Further, while selecting and preparing the synthesized stimulus, we ensured the presence of matching values in terms of duration and intensity properties associated with the contrastive tones. The following section describes the methodology adopted for the perception experiment of Sylheti tones.

5.2.1 Stimulus

For the production experiment on tone(Chapter 3), we collected data from 8 native speakers of Sylheti. The Target words were embedded in a fixed sentence frame of “I am saying **X**” [ami **X** xɔiar], **X** being the target word. A total of 24 monosyllabic words and 27 disyllabic words were examined in that experiment. For the current experiment on tone perception, out of these 24 monosyllabic words, 8 were carefully chosen (comprising the target words listed in Table 5-1). Six of those target words (embedded in the fixed sentence frame) were uttered by a female speaker, the remaining two were spoken by a male speaker. While selecting the target words, we ensured vowels (the rhyme or the voiced part of a syllable) of the contrastive pairs have equal durational and intensity values (so that these do not play any role in tone perception). The duration and intensity values of the selected words were manually cross checked in Praat.

Sylheti words	Gloss	Sylheti words with a history of underlying aspiration	Gloss
[ga]	'body'	[ga]	'wound'
[ban]	'tie'	[ban]	'pretend'
[d̪an]	'donate'	[d̪an]	'paddy'
[zal]	'net'	[zal]	'chilly hot'

Table 5.1: The dataset displaying the list of monosyllabic words considered for the perceptual experiment

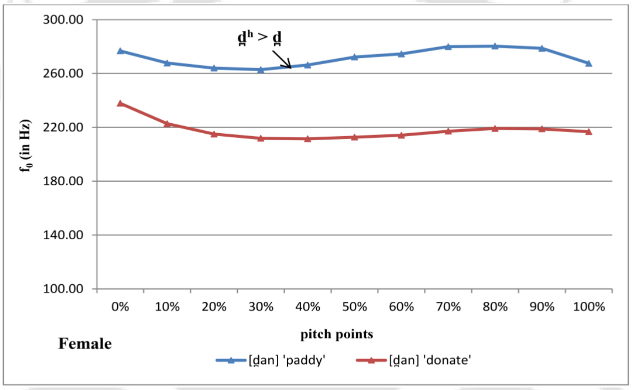

Figure 5.1: Non-normalized averaged pitch track for [**d̪an**] (n=3) is drawn using the percentage wise pitch values produced by the female speaker consider in this experiment.

To generate the artificial stimulus, we first marked the pitch onset and pitch offset of the target words and measured the duration values of the contrastive pairs. The pitch contour of each target word was then divided into 5 parts (at every 20% of the vowels (the voiced part of a syllable/rhyme) by using the following formula:

[Total duration (of the of the vowels- the voiced part of a syllable/rhyme) * X/ 100]

where X is the point of calculation (such as 20%, 40%, ..).

This was done manually and each durational point was marked separately. To prepare the synthetic stimulus, pitch contour of each vowel (the voiced part of a syllable/rhyme) was shifted up or down to match a particular point (say 20% or 40% and so on) of the pitch contour of their contrastive counterparts. So for example, stimulus 1 was created by manipulating and exchanging the first 20% of a high tone contour with the first 20% (starting from the onset till the first 20% of the rhyme) of its low tone counterparts; the remaining portion of the target contour (from 21% till the offset [100%] of the low tone contour) remained unchanged. Thus, 5 stimulus (stimulus 1- starting from the onset till first 20% of the low tone contour, stimulus 2- starting from the onset till 40% of the low tone contour, stimulus 3- starting from the onset till 60% of the low tone contour, stimulus 4- starting from the onset till 80% of the of the low tone contour, and stimulus 5- the entire low tone contour (starting from the onset till the offset), of each target words were resynthesized by exchanging the exact points (for example, 20%, 40%, 60%, 80% and entire high tone contour- 100%) of the contrastive high tone contour (Figure 5-2).

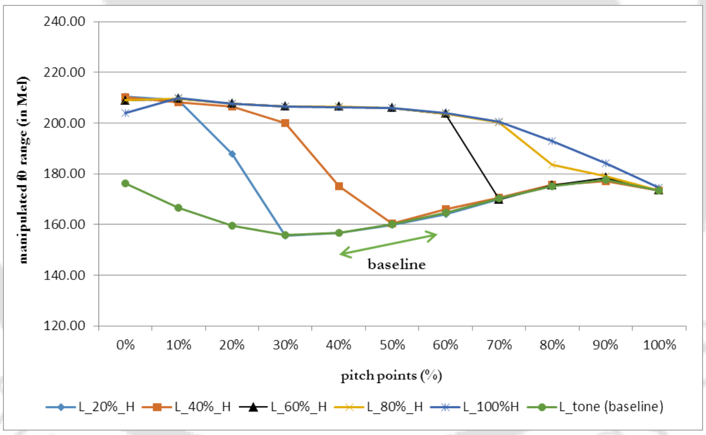

Figure 5.2: Synthesized contours of first five stimulus set. The original baseline or low tone contour is also shown. The contours are averaged across all the tokens (3 token for each word) of all the low tone words (4 words) considered in this experiment.

The mean f_0 of the original baseline/low tone was found to be 195.56 Hz (169.72 Mel). The artificial stimulus was thus gradually increased (at every 20%) to match the high tone counterparts. For example, mean f_0 of stimulus 1 was 12.70 Hz (8.88 Mel) higher than the original low tone contour. The details of f_0 differences between the synthesized stimuli and the original baseline (contours of the original low tone or high tone) are discussed in section 5.2.4 (Results and Discussion).

Following the same procedure 5 other stimuli were also created by manipulating the contours of high tone words into a systematic low tone ((Figure 5-3). All the manipulations were done manually by using the Praat pitch manipulation function and the manipulated (target) stimuli were resynthesized to create the sound files used in this experiment.

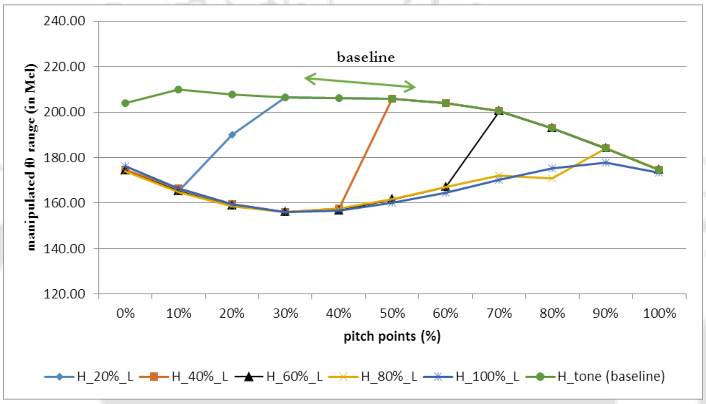

Figure 5.3: Synthesized contours of final five stimulus set. The original baseline or high tone contour is also shown. The contours are averaged across all the tokens (3 token for each word) of all the low tone words (4 words) considered in this experiment.

Note that only the contours of the target words (Table 5-1) were manipulated and resynthesized, the remaining portion of the entire (embedded) sentence [ami X xɔiar], including the portion of the onset consonant(if any) of the target words were kept unchanged.

5.2.2 Subjects

Twelve native Sylheti speakers, (6 male and 6 female) participated in this perceptual experiment. For this experiment we deliberately wanted to involve both the younger generation speakers (the data for production experiment were taken mostly from the younger generation speakers) as well as the older generation speakers to have a better understanding about the status of lexical tone in Sylheti. The first six subjects, aged between 40 and 63 years (3 male and 3 female), represent the middle aged to old generation, while the remaining 6 subjects aged between 21 and 29 years (3 male and 3 female) represent the younger generation. Out of these twelve subjects, five of them were also part of the production study. All the participants reported that they had no history of speaking or hearing disability.

5.2.3 Experimental Procedure

The target stimulus embedded in the natural sentence carrier frame [ami **X** xɔiar], 'I am saying **X**' (**X** being the target word), were played on a notebook connected with a headphone. Each stimulus was embedded with three options - the real meaning, the contrastive meaning, and the third option being *"NOT SURE"*. Each participant heard 3 repetitions of each stimulus (which were randomized) for a total of 120 tokens (8 words * 3 repetitions * 5 manipulation [each rhyme of those 8 words was manipulated at every 20% of the total duration]). Altogether 1440 responses from 10 stimuli were examined. All the subjects were instructed to listen and respond to each item and choose one of the three options displayed on the computer screen that best represents the meaning of the target word/stimulus that they have heard. Subjects were allowed to listen to each stimulus up to three times (if required)[1].

Note that a categorical perception test requires both identification and discrimination task. In an identification task a listener (of a native language) is expected to label or identify and perceive two members of one category with identical acoustic differences as members of the same category. Discrimination task on the other hand requires a speaker's ability to hear the difference between a pair of sounds (based on any distinct acoustic change) belonging to two different categories. Ideally, all stimuli in a given category should be perceived as the same, whereas stimuli from different categories, no matter how close on the continuum, should be perceived as different.

In our experiment, instead of following the model of categorical perception directly we have modified our experiment and left the option with the native speakers (listeners) whether to identify or discriminate a particular token belonging to a specific tone. We

[1]In general, categorical perception experiments have been most successful in showing voice as a categorical feature.

argue that, in a given context (in our case it is the exploitation of f_0 of the rhyme) if a native speaker identifies a particular word 'X' as 'A' (say for example a word with a higher f_0) and not as 'B' or 'C', then, 'A' is identified as belonging to one particular category (here as high tone). On the other hand, when that particular context of the same word 'X' is changed or altered (lowering the f_0 of the rhyme to a certain extent) and the resultant word is perceived as 'B' over 'A' or 'C', then option 'B' is being discriminated as belonging to the other category ('A'- high tone). This process also provides the opportunity to examine whether these stimuli are perceived categorically or not. We hypothesized that stimulus with less discrimination might be perceived as the token belonging to the same category, i.e., a token (say a low tone contour) with a manipulation of only first 20% of its contour (into first 20% of contrastive high tone) might not be perceived as a different word (since the manipulated contour will have a very similar shape of that of a low tone), similarly, a token with large discrimination (for example a manipulation of 80% of its total contour) would be perceived as a different word. Thus, the more the fluctuation is, the more it is likely to be perceived as contrastive counterpart.

5.2.4 Results and Discussion

Manipulation of low tone contour into high tone contour (at every 20% of the contour)

Stimulus 1: To create stimulus 1 the first 20% of a low tone contour is resynthesized and substituted with the first 20% of contrastive high tone counterpart, the remaining portion of the low tone contour (from the 21% till the offset of the contour, see figure 5-2 and 5-3 for example) remained unchanged. On average the mean f_0 value of the first 20% (manipulated portion) of the synthesized stimulus (245.66 Hz or 202.39 Mel, first 20%, is 49.3 Hz or 35.02 Mel higher than the first 20% of the original baseline or low tone contour.

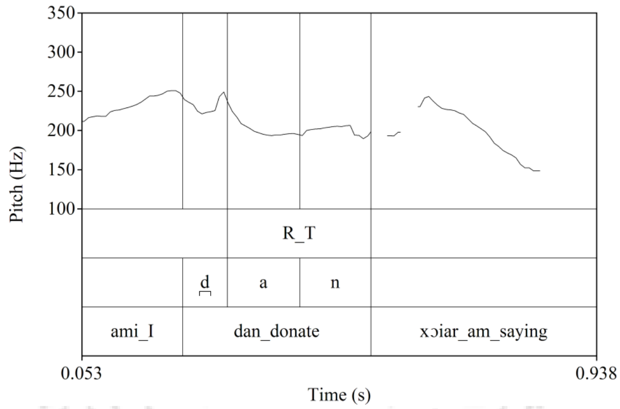

Figure 5.4: Original pitch-track (contour) of a low tone word [**d̪an**] 'donate' produced by a female speaker in a fixed sentence frame "I am saying **donate**" [ami **d̪an** xɔiar]. Portion [R_T] represents the vowel (the voiced part of a syllable/rhyme) of the target word.

The statistical results obtained from the production experiment on tone(as discussed in chapter 3) showed a significant effect of tone on f_0 at the first 20% (from the onset till the first 20% of the the rhyme/voiced part) of the vowels carrying contrastive tones (for example, $p < 0.05$, [$F\ (1,\ 414) = 622.58$, $p = 0.00$)], measured in Mel and then converted into normalized z score [mono-syllabic words]). On average, the synthesized stimulus (the entire contour of stimulus 1) is 11.57 Hz (8.43 Mel) (mean f_0) higher than the original baseline or low tone contour.

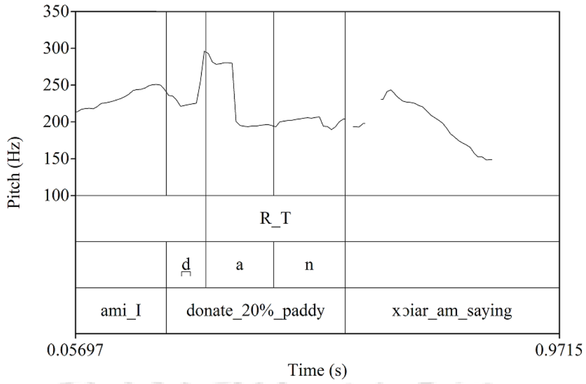

Figure 5.5: An example of stimulus 1 where first 20% of a low tone word (contour) [**d̪an**] 'donate' is raised and exchanged with the f_0 value of the first 20% of its high tone counterpart [**d̪an**] 'paddy' produced by the same female speaker in the same fixed sentence frame "I am saying **paddy**" [ami **d̪an** xɔiar]. Portion [R_T] represents the vowel (the voiced part of a syllable/rhyme) of the target word.

Responses to stimulus 1 (Figure 5-6) gathered from all the subjects (across all age groups) show that the manipulation of first 20% of the total contour of a low tone was still perceived as a low tone word[2]. On average, 88.89% of the total responses identified stimulus 1 as a low tone word, whereas only 9.72% of the total responses from all the speakers identified it as a high tone word. 1.39% of the total response was marked as *'Not Sure'*.

[2]It is almost certain that the overall f_0 difference of the manipulated stimulus (created on the first 20% of vowel (the rhyme or the voiced part of a syllable) wasn't large enough (the synthesized stimulus was only 11.57 Hz higher than the original low tone word) for the native speaker to be able to hear stimulus 1 differently.

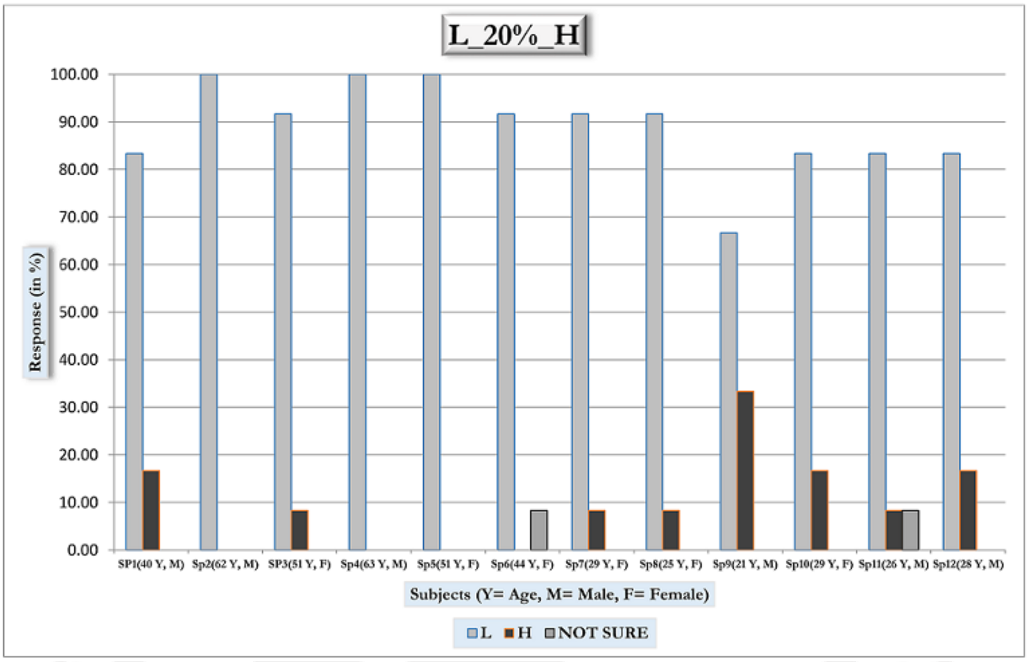

Figure 5.6: Responses (in % shown in Y axis) of the speakers (shown in X axis) to Stimulus 1 (SP= Speaker/Subject, Y= Age, M=Male, F=Female). Total Responses considered = 144 (4 words * 3 repetitions * 12 subjects, L= (identified as) low tone word, H= (identified as) high tone words).

Stimulus 2: Stimulus 2 represents the second manipulated f_0 contour where the first 40% of a low tone contour is resynthesized and substituted with first 40% of contrastive high tone contour, the remaining portion of the low tone contour (from the 41% till the offset of the contour) remained the same. The mean f_0 value of the first 40% (manipulated portion) of the synthesized stimulus (243.03 Hz or 200.01 Mel) is 51.80 Hz or 37.07 Mel higher than the first 40% of the original baseline or low tone contour (190.23 Hz or 162.94 Mel. On the other hand, when the entire contour of the synthesized stimulus and the original baseline or low tone contour are compared, it is observed that on average the synthesized stimulus (stimulus 2) is only 23.34 Hz or 16.99 Mel (mean f_0) higher than the original baseline or the vowel (the voiced part or the rhyme) carrying low tone.

Responses to stimulus 2 (Figure 5-7) gathered from all the subjects (across all age groups) show that the manipulation of first 40% of the total contour of a low tone into the same range (first 40%) of their respective high tone counterparts resulted in mixed responses. While a few subjects (subject no 5, 8, 3, and 11) perceived these stimulus mostly as low tone words, a few others (subject no 1, 4, 7, and 9) gave mixed responses. Subject 12 surprisingly identified these tokens mostly as high tone words (nearly 70% of his total response identified stimulus 2 as high tone word). On average, 68.06% of the total responses identified Stimulus 2 as low tone word, whereas 31.25% of the total responses from all the speakers identified it as high tone counterpart. 0.69% of the total response was marked as *'Not Sure'*.

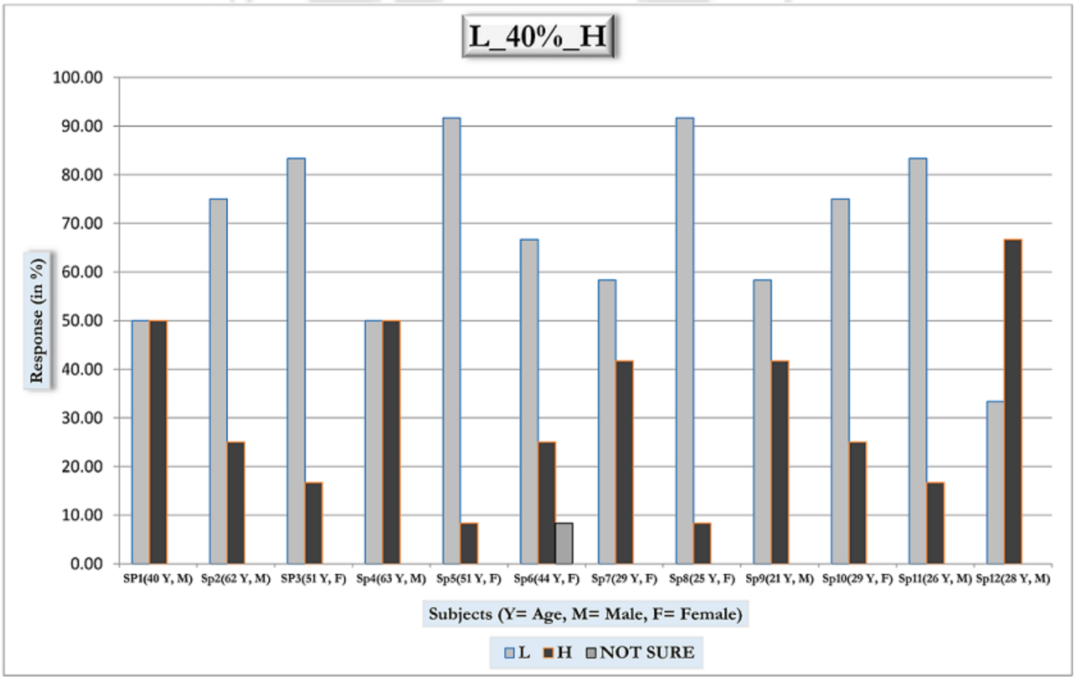

Figure 5.7: Responses (in % shown in Y axis) of the speakers (shown in X axis) to Stimulus 1 (SP= Speaker/Subject, Y= Age, M=Male, F=Female). Total Responses considered = 144 (4 words * 3 repetitions * 12 subjects, L= (identified as) low tone word, H= (identified as) high tone words).

Stimulus 3: Stimulus 3 is the third manipulated contour where the first 60% of a low tone contour is substituted with the first 60% of contrastive high tone contour. On average, the manipulated portion (first 60% of the contour) of the synthetic stimulus (stimulus 3) is 57.11 Hz or 44.19 Mel higher than the original baseline or the voiced part of the syllable/rhyme representing the low tone, whereas when the entire contour (starting from the onset till the offset of the voiced part of the syllable/rhyme) of the synthesized stimulus (stimulus 3) and the original baseline (low tone contour) are compared, it is observed that the average f_0 of the synthesized stimulus is 37.34 Hz or 26.94 Mel higher than the original low tone contour(mean f_0).

Responses to stimulus 3 gathered from all the subjects (across all age groups) indicate that manipulation of first 60% of a low tone into the same range (first 60%) of their respective high tone counterparts created large enough f_0 fluctuations for the native speakers to perceive this stimulus as high tone. While most of the subjects (subject no 1, 4, 5, 7, 8, 10, and 11) have identified these tokens (4 words * 3 repetitions) mostly as high tone words, a few subjects perceived these tokens mostly as low tone word (speaker no 2 and 3) [more than 50% and 70% respectively]. Subject 6, 9 and 12 on the other hand gave mixed responses (see Figure 5-8).

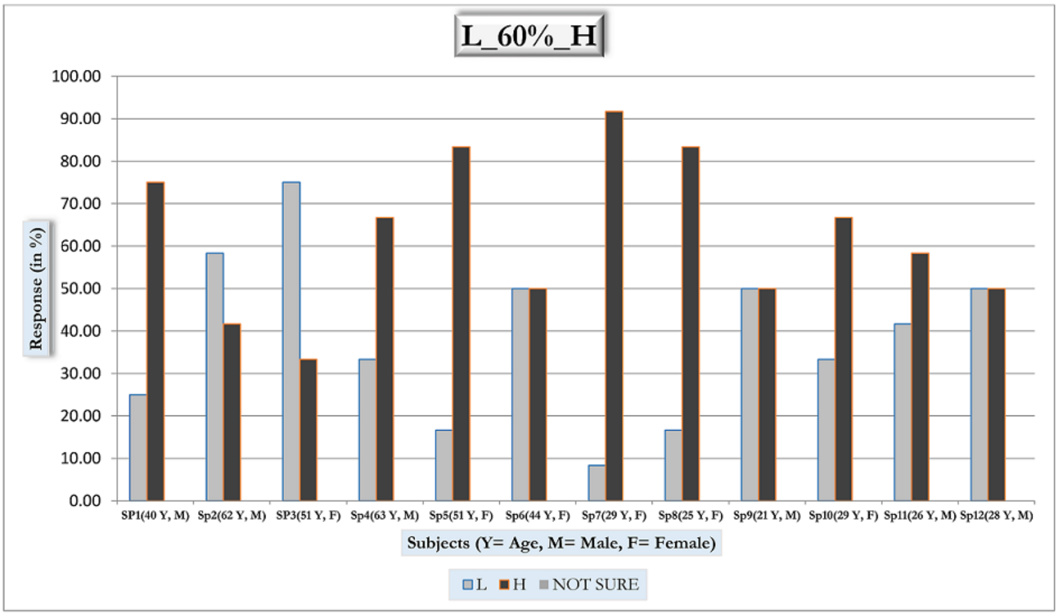

Figure 5.8: Responses (in % shown in Y axis) of the speakers (shown in X axis) to stimulus 3 (SP= Speaker/Subject, Y= Age, M=Male, F=Female). Total Responses considered = 144 (4 words * 3 repetitions * 12 subjects, L= (identified as) low tone word, H= (identified as) high tone words).

On average, 38.19% of the total responses perceived stimulus 5 as low tone, whereas as much as 62.5% of the total responses from all the speakers perceived it as high tone token. These responses indicate that the pitch differences of the (contrastive) contours starting from 50% (and above) of the rhyme/voiced part of the syllable might play a vital role in perceiving the tonal distinctions.

Stimulus 4: To create the fourth synthesized stimulus of this perception experiment we first measured and identified the first 80% contour of low tone words used in this experiment. Similarly, we also measured the first 80% contour of the contrastive high tone pairs. Finally, the synthesized stimulus is created by exchanging the f_0 contour of the high tone words with their low tone counterparts. The final 20% of the low tone words remained unchanged. Thus, the first 80% of stimulus 4 bears the f_0 contour of high tone while the final 20% is that of the low tone counterpart (see Figure 5-9 and Figure 5- 10 for an exemplary token). Altogether 4 such tokens are created. The synthesized stimulus is 46.23 Hz (33.2 Mel) [mean f_0 of the entire contour] higher than the original low tone baseline/contour. We hypothesized that since the contours of these low tone tokens are transformed into (almost) identical high tone counterparts, these tokens would be perceived mostly as the high tone tokens.

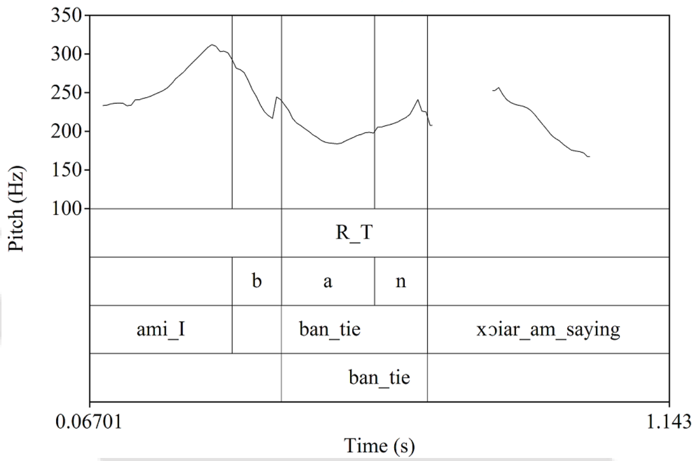

Figure 5.9: Original pitch-track (contour) of a low tone word [**ban**] 'tie' produced by a female speaker in a fixed sentence frame "I am saying **tie**" [ami **ban** xɔiar]. Portion [R_T] represents the voiced part or the rhyme of the target word.

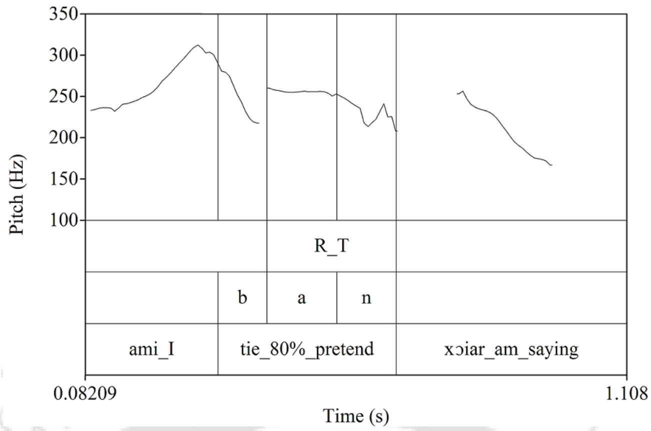

Figure 5.10: An example of stimulus 4 where first 80% of a low tone word(contour) [**ban**] 'tie' is raised and exchanged with the f_0 value of the same range (from onset till the 80% the total rhyme/voice part of the target syllable) of its high tone counterpart [**ban**] 'pretend' produced by the same female speaker in the same fixed sentence frame "I am saying **pretend**" [ami **ban** xɔiar]. Portion [R_T] represents the voiced part or the rhyme of the target word.

As expected, the manipulation of first 80% of a low tone contour into the same range (first 80%) of their respective high tone counterparts have been perceived mostly as high tone tokens. On average, only 13.19% of the total responses perceived stimulus 7 as low tone word, whereas as much as 86.11% of the total responses from all the speakers identified it as high tone token. 1.39% of the total responses were marked as *'Not Sure'* (Figure 5-11).

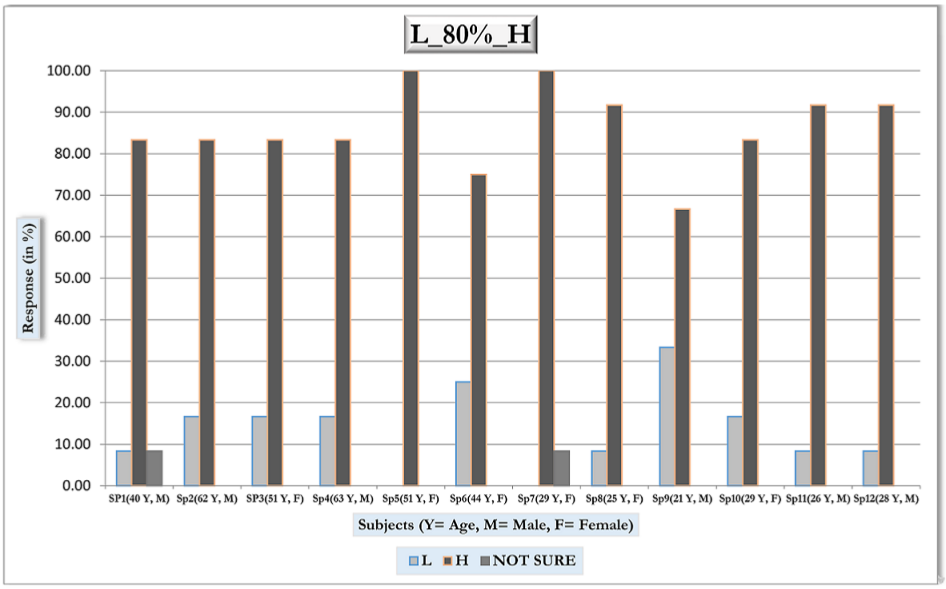

Figure 5.11: Responses (in % shown in Y axis) of the speakers (shown in X axis) to Stimulus 7 (SP= Speaker/Subject, Y= Age, M=Male, F=Female). Total Responses considered = 144 (4 words * 3 repetitions * 12 subjects, L= (identified as) low tone word, H= (identified as) high tone words).

Stimulus 5: The entire contour (100%) of a low tone is exchanged and resynthesized with the contour of the contrastive high tone. As per our hypothesis, most of these tokens should be perceived as high tone words. On average, the synthesized stimulus is 47.86 Hz (34.42 Mel) [mean f_0] higher than the original baseline or low tone contour.

The overall results (Figure 5-12) indicate that the manipulation of the entire low tone contour into high tone counterpart (the entire contour, from the onset [0%] till the offset [100%] of the rhyme/voiced part of the syllable) has been perceived mostly as high tone words. On average, only 10.42% of the total responses identified stimulus 5 as low tone word, whereas 89.58% of the total responses were in favour of high tone word.

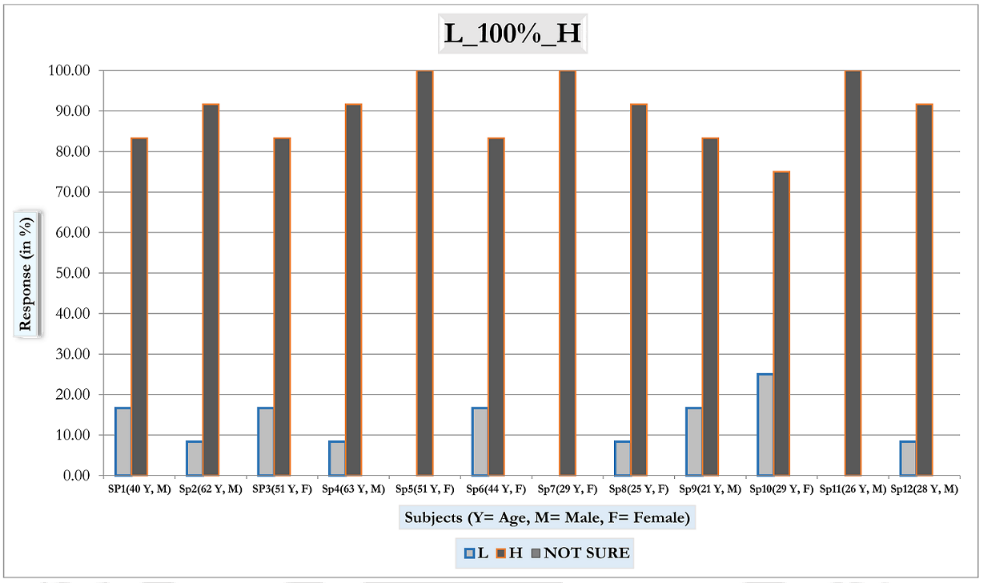

Figure 5.12: Responses (in % shown in Y axis) of the speakers (shown in X axis) to Stimulus 5 (SP= Speaker/Subject, Y= Age, M=Male, F=Female).Total Responses considered = 144 (4 words * 3 repetitions * 12 subjects, L= (identified as) low tone word, H= (identified as) high tone words).

Manipulation of high tone contour into low tone contour (at every 20% of the contour)

Stimulus 6: The first 20% of a high tone contour is resynthesized by lowering and exchanging that portion with the first 20% of a contrastive low tone contour. The synthesized portion (first 20%) of stimulus 6 is 37.40 Hz or 30.64 Mel (mean f_0) lower than the original baseline or high tone contour. However, the overall synthesized contour (starting from the onset [0%] till the offset [100%] of the voiced part or the rhyme of the syllable) is 11.17 Hz) (7.84 Mel) [mean f_0] lower than original baseline or high tone contour (see Figure 5-13 and Figure 5-14 for an exemplary token of stimulus 6).

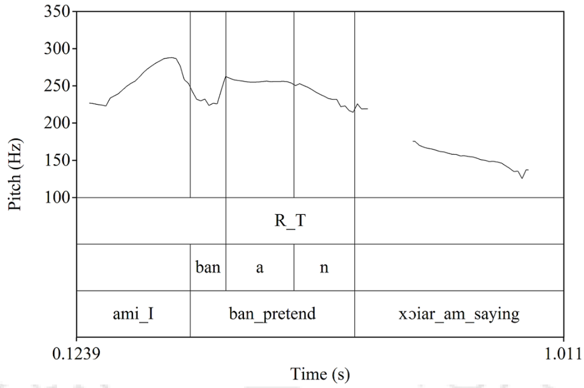

Figure 5.13: Original pitch-track (contour) of a high tone word [**ban**] 'pretend' produced by a female speaker in a fixed sentence frame "I am saying **pretend**" [ami **ban** xɔiar]. Portion [R_T] represents the voiced part or the rhyme of the target word.

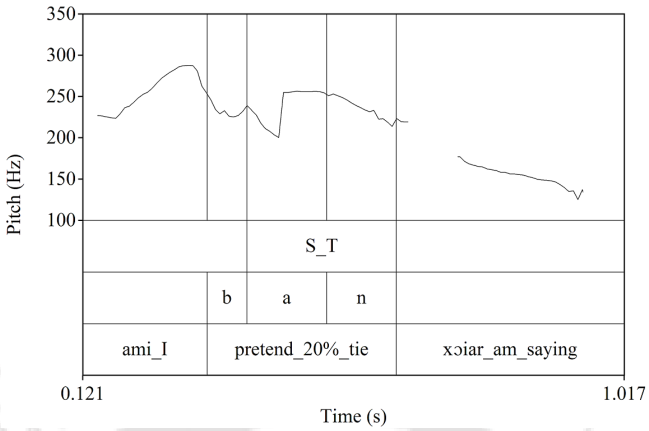

Figure 5.14: An example of stimulus 6 where the first 20% of a high tone word (contour) [ban] 'pretend' is lowered and exchanged with the first 20% of its low tone counterpart [**ban**] 'tie' produced by the same female speaker in the same fixed sentence frame "I am saying **tie**" [ami ban xɔiar]. Portion [R_T] represents the voiced part or the rhyme of the target word.

We assumed that the f_0 fluctuation between the synthesized stimulus (stimulus 6) and original baseline of high tone contour might not be large enough to be perceived as different words. Responses to stimulus 6 gathered from all the subjects (across all age groups) show that most of the listeners heard stimulus 6 as a high tone word(Figure 5-15).

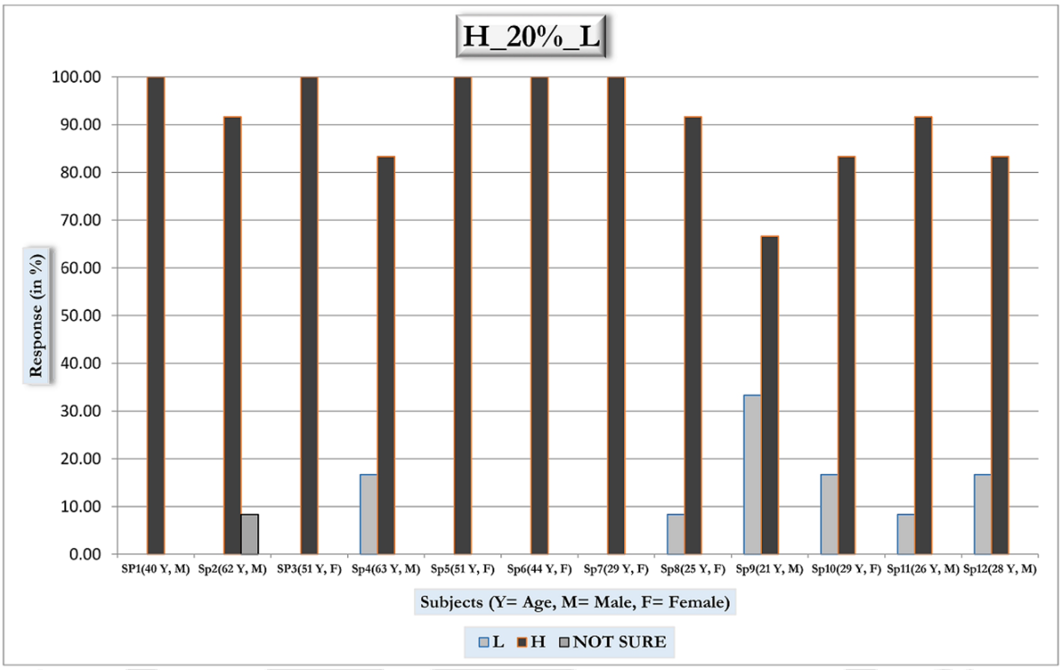

Figure 5.15: Responses (in % shown in Y axis) of the speakers (shown in X axis) to stimulus 6 (SP= Speaker/Subject, Y= Age, M=Male, F=Female).Total Responses considered = 144 (4 words * 3 repetitions * 12 subjects, L= (identified as) low tone word, H= (identified as) high tone words).

On average, 90.97% of the total responses perceived stimulus 6 as a high tone word, whereas only 8.33% of the total responses from all the subjects identified it as a low tone word. A total of 0.69% of the responses was marked as *'Not Sure'*

Stimulus 7: To create stimulus 7, the first 40% of a high tone contour is exchanged with the first 40% of contrastive low tone contours. The remaining portion (from 41% till the offset (100%) of the voiced part or the rhyme of the syllable) of the high tone contour was unchanged. On average, the synthesized portion (first 40%) of stimulus 7 is 52.11 Hz (43.98 Mel) [mean f_0] lower than the original baseline or the high tone contour; however, the entire contour (from 0% till 100% of the voiced part or the rhyme of the syllable) is only 25.54 Hz (18.04 Mel) [mean f_0] lower than the original high tone contour.

Responses to stimulus 7 (Figure 5-16) gathered from all the subjects (across all age groups) indicate that most of the listeners have heard stimulus 7 mostly as high tone word. Subject 9 appears to be a bit confused and gave mixed responses (half of his responses were marked as low tone word). Overall 78.47% of the total responses perceived stimulus 7 as a high tone word, whereas 20.83% of the total responses from all the speakers perceived it as a low tone word. 0.69% of the total response was marked as *'Not Sure'*.

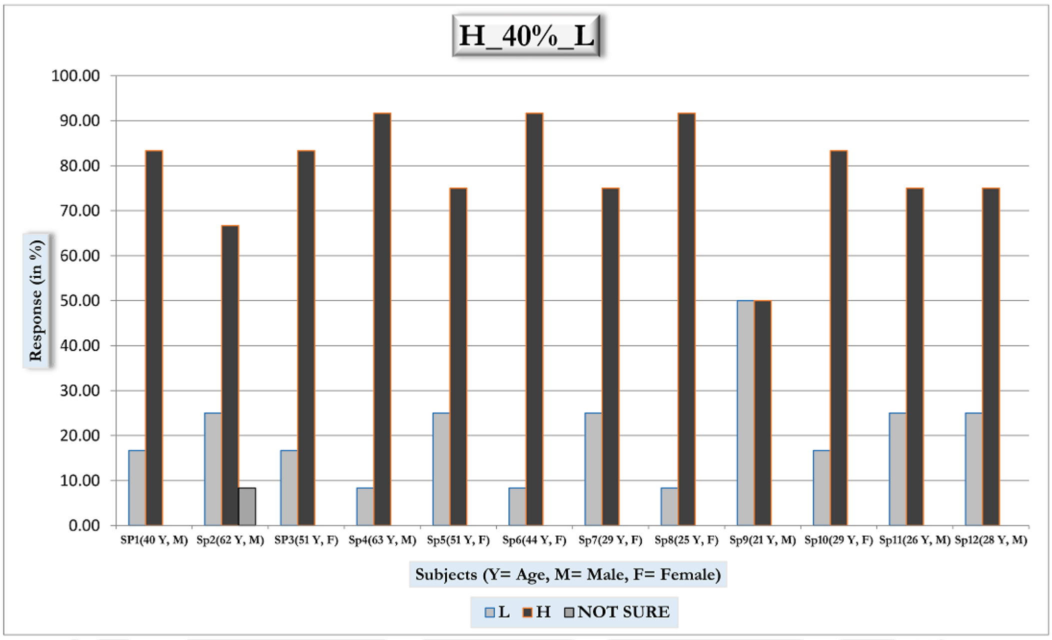

Figure 5.16: Responses (in % shown in Y axis) of the speakers (shown in X axis) to Stimulus 4 (SP= Speaker/Subject, Y= Age, M=Male, F=Female).Total Responses considered = 144 (4 words * 3 repetitions * 12 subjects, L= (identified as) low tone word, H= (identified as) high tone words).

Stimulus 8: The first 60% of a high tone contour is lowered and resynthesized to match the first 60% of a low tone counterpart; the remaining portion (from the 61% till 100%) of the high tone contour remained unchanged. The synthesized portion (first 60%) of stimulus 8 appears to be 54.36 Hz or 43.46 Mel [mean f_0] lower than the original baseline or high tone contour whereas the entire synthesized contour (from 0% till 100%) is 37.32 Hz (26.54 Mel) [mean f_0] lower than the original baseline.

The overall results (Figure 5-17) indicate that the f_0 fluctuation generated in the synthesized contour of stimulus 8 was large enough for most of the listeners (subject no 5, 7, 8, 9, and 11)to hear these tokens mostly as low tone words.On the other hand, a few listeners heard these tokens mostly as high tone word (subject 2, 6 and 10) [more than 50%]. Subject 1, 4, and 12 gave mixed responses (50% each). On average, 58.33% of the total responses heard stimulus 8 as a low tone word, whereas 41.67% of the total responses from all the subjects heard stimulus 8 as a high tone word.

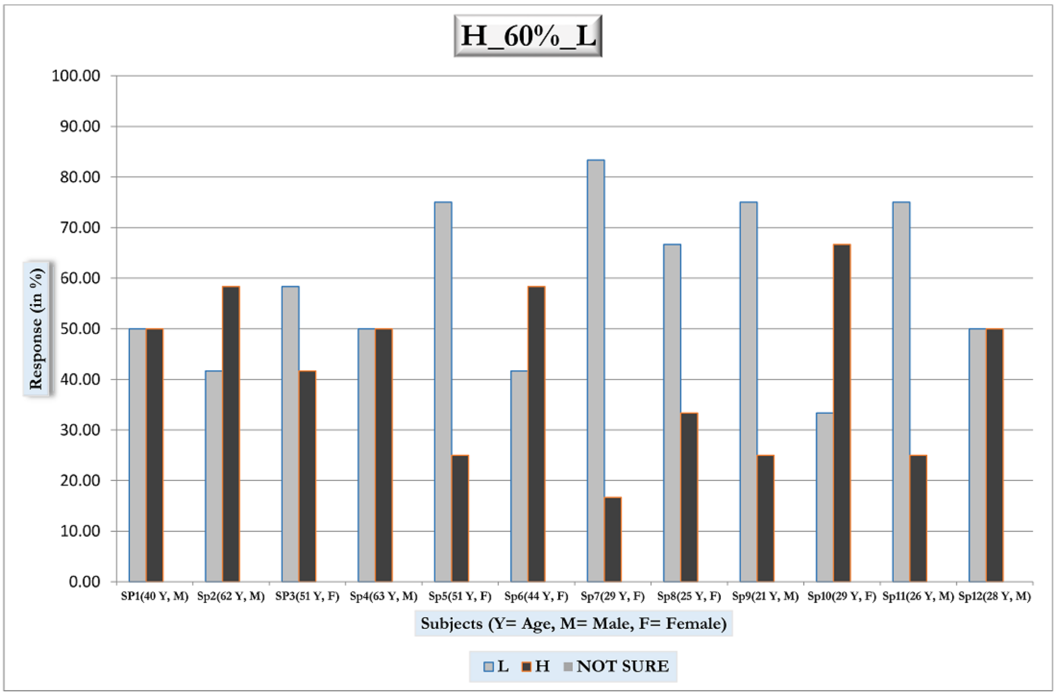

Figure 5.17: Responses (in % shown in Y axis) of the speakers (shown in X axis) to Stimulus 6 (SP= Speaker/Subject, Y= Age, M=Male, F=Female).Total Responses considered = 144 (4 words * 3 repetitions * 12 subjects, L= (identified as) low tone word, H= (identified as) high tone words).

Stimulus 9: The first 80% of a high tone contour is manipulated and exchanged with the first 80% of contrastive low tone counterparts; the final 20% (from the 81% till the offset of the voiced part or the rhyme of the syllable) of the high tone contour remained unchanged (see Figure 5-18 and Figure 5-19 for an exemplary token of stimulus 9). The synthesized portion (first 80%) of stimulus 9 is 52.43 Hz (39.36 Mel) [mean f_0] lower than the original baseline. However, the entire contour (from 0% till 100% voiced part or the rhyme of the syllable) of stimulus 9 is 46.11 Hz (33.02 Mel) [mean f_0 of the entire contour] lower than the original high tone baseline, thus making the synthesized contour a low tone contour. Hence, we can assume that most of these tokens would be heard as representing the low tone word.

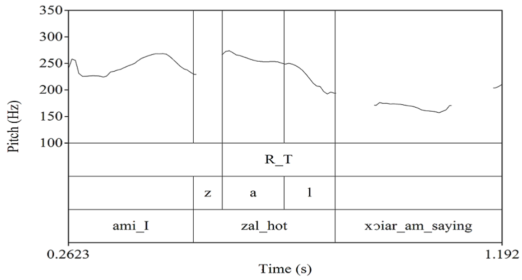

Figure 5.18: Original pitch-track (contour) of a high tone word [**zal**] 'hot/spicy' produced by a female speaker in a fixed sentence frame "I am saying **hot/spicy**" [ami **zal** xɔiar]. Portion [R_T] represents the voiced part or the rhyme of the target word.

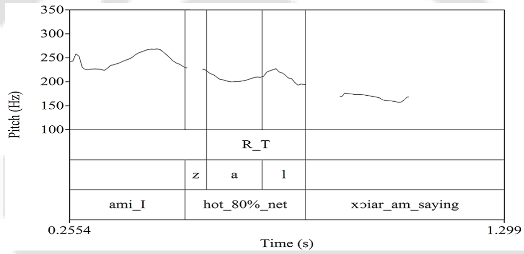

Figure 5.19: An example of stimulus 8 where the first 80% of a high tone word(contour) [**zal**] 'hot/spicy' is exchanged with the first 80% of its low tone counterpart [**zal**] 'net' produced by the same female speaker in the same fixed sentence frame "I am saying **net**" [ami **zal** xɔiar]. Portion [R_T] represents the voiced part or the rhyme of the target word.

As expected the overall results indicate that all the subjects heard stimulus 9 mostly as low tone words (Figure 5-20). On average, 92.36% of the total responses perceived stimulus 9 as low tone words, whereas only 6.94% of the total responses from all the speakers perceived it as high tone words. 0.69% of the total responses were marked as *'Not Sure'*.

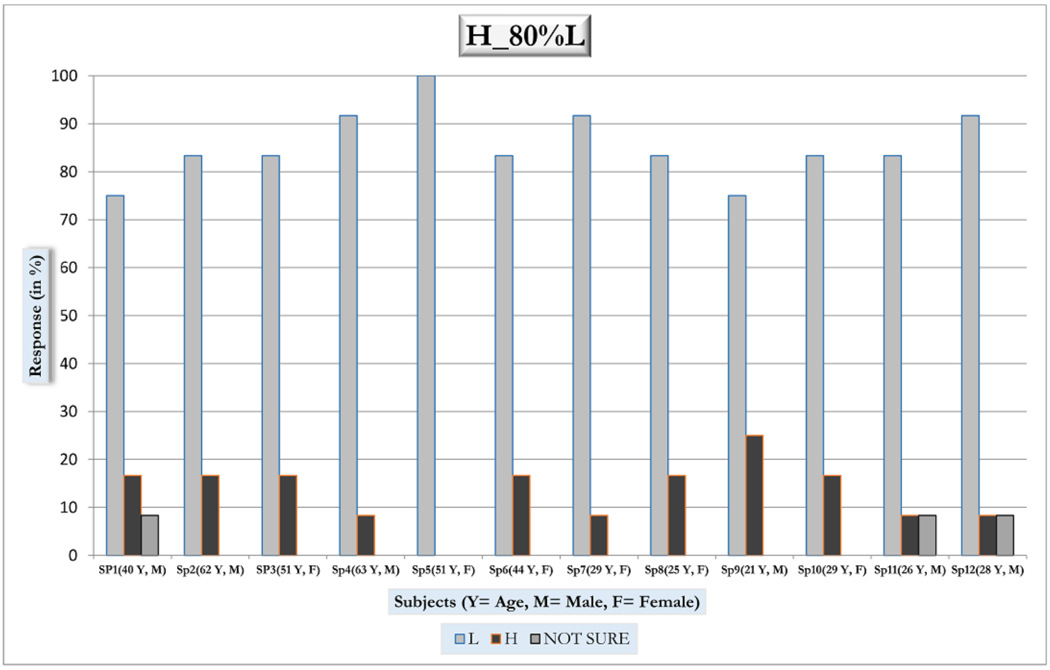

Figure 5.20: Responses (in % shown in Y axis) of the speakers (shown in X axis) to Stimulus 8 (SP= Speaker/Subject, Y= Age, M=Male, F=Female).Total Responses considered = 144 (4 words * 3 repetitions * 12 subjects, L= (identified as) low tone word, H= (identified as) high tone words)

Stimulus 10: The entire contour (from the onset (0%) till offset (100%) of the voiced part or the rhyme of the syllable) of a high tone word is exchanged and resynthesized with the entire contour (100%) of contrastive low tone words. The synthesized contour of stimulus 10 is 47.86 Hz (34.42 Mel) [mean f_0 of the entire contour] lower than the contrastive high tone counterpart. We assume that the tokens of stimulus 10 would be heard mostly as low tone words. As expected, all of the responses to stimulus 10 gathered from all the subjects (across all age groups) suggest that these tokens were heard (mostly) as low tone words. On average, 92.36% of the total responses perceived stimulus 10 as a low tone word, whereas only 6.94% of the total responses of all the subjects perceived it as a high tone word. 0.69% of the total response was marked as *'Not Sure'*.

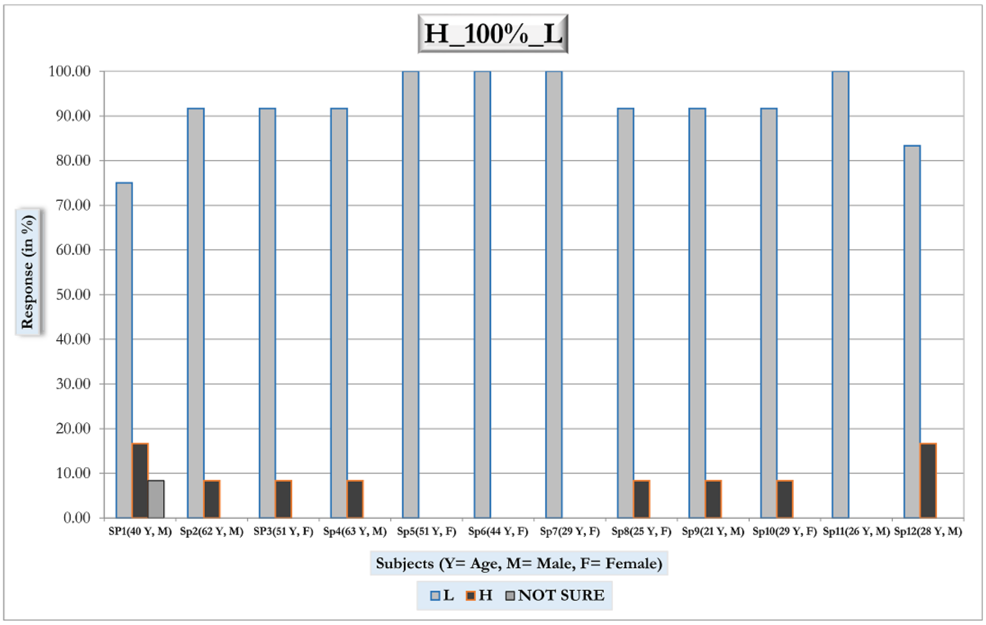

Figure 5.21: Responses (in % shown in Y axis) of the speakers (shown in X axis) to Stimulus 2 (SP= Speaker/Subject, Y= Age, M=Male, F=Female).Total Responses considered = 144 (4 words * 3 repetitions * 12 subjects, L= (identified as) low tone word, H= (identified as) high tone words).

From the results discussed above it is reasonably clear that the f_0 fluctuation of the contrastive words bears the most prominent perceptual cue in identifying contrastive tones in Sylheti. Further, the trend of perception of contrastive tones also suggest that tones in Sylheti is perceived (mostly) in categorical manner. The results also suggest that listeners do also rely on f_0 fluctuation made till a certain portion of the total voiced part or the rhyme of the syllable- manipulations made till 60% (and above, Figure 5-22 and Figure 5-23) of the total contour are likely to be perceived as contrastive tone; thus also confirming the importance of the length of the voiced part or the rhyme of the syllable. We termed this boundary as 'categorical point'.

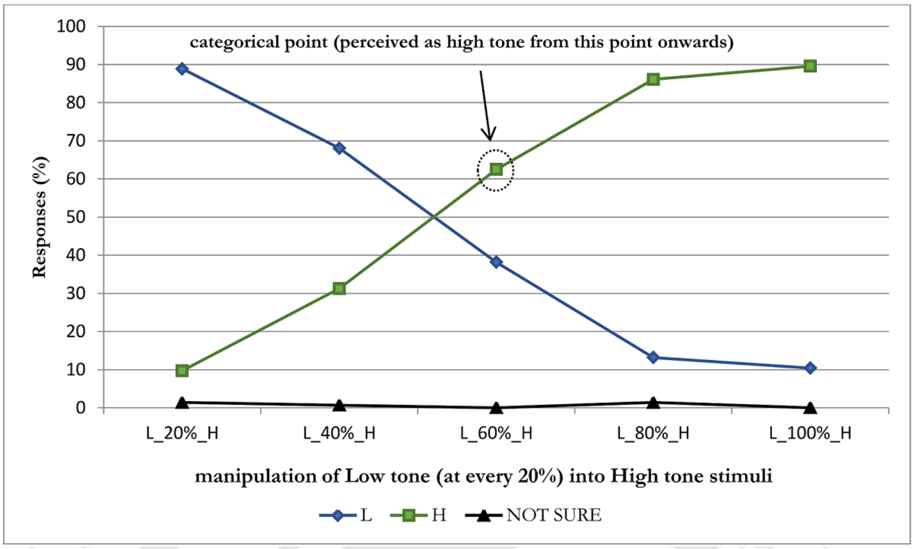

Figure 5.22: The average response of (manipulated) low tones in to a (synthesized) high tone. The systematic manipulation was made at every 20% of the original baseline (that is voiced part or the rhyme of target syllable). Total responses considered = 720 (4 words * 3 repetitions * 12 subjects * 5 stimulus).

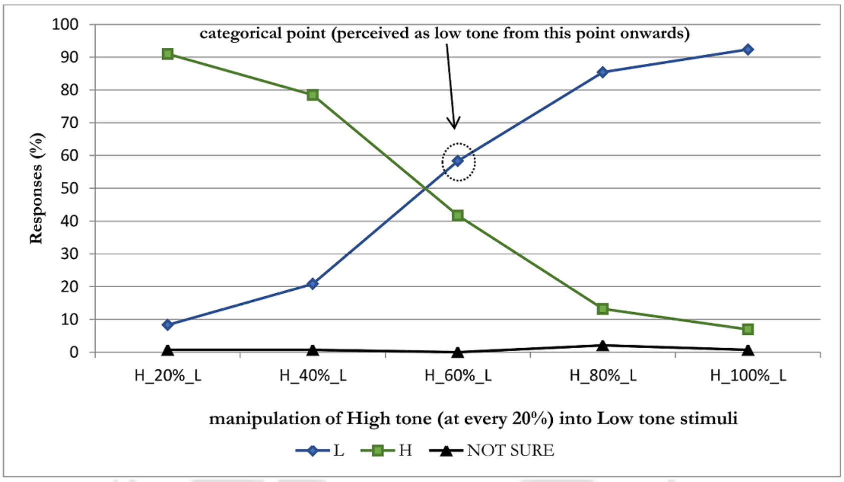

Figure 5.23: The average response of (manipulated) high tones in to a (synthesized) low tone. The systematic manipulation was made at every 20% of the original baseline or high tone (that is voiced part or the rhyme of target syllable). Total responses considered = 720 (4 words * 3 repetitions * 12 subjects * 5 stimulus).

In chapter three of this dissertation (production of tone) we have observed in general, the difference between the traces is greater in the middle of the vowel/voiced part of the rhyme. We therefore conclude that, the contrastive contours (the entire rhyme or the voiced part of a target syllable: starting from the onset [0%] till the offset [100%] must contain a minimum f_0 fluctuation of 37.32 Hz (26.94 Mel) [mean f_0] and the range of f_0 fluctuation must be continued (at least) till 60% of the total rhyme/voiced part of target syllable) to be able to be perceived as different words. In chapter six we shall discuss the complexities of tone assignment in morphologically derived environments.

Chapter 6

Morphophonology in Sylheti

General Introduction

The results of the perception test discussed in the previous chapter of this dissertation confirmed that Sylheti native speakers by and large depend on f_0 fluctuation (along with the length of TBU to some extent) to be able to perceive tonal contrast. In this chapter we will examine the complexities of tone assignments in morphologically derived environments. In addition to that we will also assess the nature and underlying tonal properties of a number of different suffixes that take part in various morphological processes in Sylheti.

6.1 Morphophonology of tone

Yip (2002) proposes to divide 'tonal morphemes' into two types based on the origin of surface tone in a word. It is observed that prefixes and suffixes sometimes bear the tone specified in the root word. This is achieved by assimilation or spreading rule- those tonal morphemes belong to type one. On the other hand, there are some other tonal morphemes that overwrite the (underlying) tone of the root (if any) replacing them entirely. This process of tone reversal is accomplished by deletion or dissimilation rule. According to Yip (2002), those tonal morphemes belong to type two.

Furthermore, in a few language tones of the suffixes or prefixes appear to be certainly predictable in the sense that those suffixes or prefixes bear a tone that is opposite of the neighbouring tone (foot); i.e., words (roots) that end in H take L affixes, and words (roots) that end in L takes H affixes. This process is termed as 'polarity' (Yip 2002, Antilla 2001, Cahill 1999). The process of polarity could be seen in many African languages including the Bantu languages. Many Gur languages such as Dagaare and Knni (Cahill 2004, Antilla and Bodomo 2000) are perhaps best known for this phenomenon.

In our final production experiment, we examined (common) noun roots with two types of suffixes- the definite classifier [- ta] and the inanimate plural marker [- tain]. The definite classifier [- ta] is also used as a general classifier; however it is restricted to inanimate objects only. There is another classifier available in Sylheti [-xan] which is restricted to inanimate count nouns (we haven't considered this classifier in this experiment. In addition to the inanimate plural suffix [- tain], there are two other plural suffix available in Sylheti- [-ra] and [-t̪ara] which are restricted to animate nouns. In our experiment we have used the inanimate plural suffix [- tain] only.

Further, verb forms with suffixes such as- [-sɛ] '3rd person perfective', [-i] '1st person present indefinite', and [-ɛ] '3rd person present indefinite' were also examined. The primary objective of this chapter is to evaluate the underlying tonal nature of suffixes mentioned above and to arrive at a proper understanding of tonal alignment in Sylheti. The following section describes the methodology adopted for data collection, digitizing them, and the acoustic analysis of speech data.

6.2 Experimental procedure

6.2.1 Methodology adopted for data collection

In chapter three of this dissertation we have shown that the pitch movement of the rhyme in a syllable shows that Sylheti possesses two lexical tones, viz. high and low. To understand and evaluate the interaction of these lexical tones with various suffixes used in Sylheti, the current experiment is designed to understand the following two factors-

- to evaluate the nature of contrastive tones (high or low) in the presence of the same suffix, and
- to evaluate the nature of tone of a particular suffix in the presence of contrastive tones (high or low).

To achieve these goals nine native speakers(7 males and 2 females and aged between 17-28 years of age) from the Dharmanagar district of north Tripura of Sylheti were asked to produce various scripted sentences following the same procedure adopted in other production experiments; i.e., a priming sentence was recorded first, followed by the target words (with various suffixes attached to the target word) in a fixed sentence frame of [ami **X** xɔiar], 'I am saying **X'** (**X** being the target word with the distinct suffix attached to those target words). A total of 24 words (12 noun roots and 12 verb roots, Table 6-1 and Table 6-2) bearing contrastive tones were chosen for the current experiment.

Noun (roots)	Gloss	Suffix	Gloss
[àd̪à]	'ginger'	[-ʈain]	'inanimate plural marker'
[bàlà]	'bracelet'	[-ʈain]	'inanimate plural marker'
[bálá]	'good'	[-ʈain]	'inanimate plural marker'
[bàr̈i]	'home'	[-ʈain]	'inanimate plural marker'
[bárí]	'heavy'	[-ʈain]	'inanimate plural marker'
[d̪án]	'paddy'	[-ʈain]	'inanimate plural marker'
[ɖáx]	'drum'	[-ʈain]	'inanimate plural marker'
[zàl]	'net'	[-ʈain]	'inanimate plural marker'
[d̪án]	'paddy'	[-ʈa]	'definite classifier'
[ɖáx]	'drum'	[-ʈa]	'definite classifier'
[gà]	'body'	[-ʈa]	'definite classifier'
[zàl]	'net'	[-ʈa]	'definite classifier'

Table 6.1: The dataset comprising the noun roots and the suffixes considered in this experiment.

Verb (roots)	Gloss	Suffix	Gloss
[ɖéx]	'to see'	[-s ɛ]	'3rd person perfective'
[ɖál]	'to pour'	[-i]	'1st person present indefinite'
[gà]	'to sing'	[-i]	'1st person present indefinite'
[gà]	'to sing'	[-ɛ]	'3rd person present indefinite'
[lɛ́x]	'to write'	[-i]	'1st person present indefinite'
[màr]	'to beat'	[-i]	'1st person present indefinite'
[màr]	'to beat'	[-ɛ]	'3rd person present indefinite'
[ʃùn]	'to listen'	[-i]	'1st person present indefinite'
[ʈáx]	'to stay/live'	[-i]	'1st person present indefinite'
[xɛ́l]	'to play'	[-i]	'1st person present indefinite'
[xɔ̀r]	'to do'	[-ɛ]	'3rd person present indefinite'
[xɔ̀r]	'to do'	[-i]	'1st person present indefinite'

Table 6.2: The dataset comprising the verb roots and the suffixes considered in this experiment.

Post recording, individual sound files of target words (with individual suffix) were manually segmented, and two tier Praat . Textgrid files were created- the first tier representing individual phonemes of each word, and the second tier representing the vowel (voiced part of the rhyme). A Praat script was written to measure the pitch contour at every 10% of the total duration of each rhyme (voiced part of the vowel). Pitch was thus calculated at 11 consecutive points- starting from the onset ('start-pitch' [0%]) till the offset ('end-pitch [100%]) across the duration of each rhyme of the syllable; each point representing 10% of the total length of the pitch track. This was done with a pitch floor of 40 Hz and pitch ceiling of 600 Hz with a default time step of 10 milliseconds. Finally, z-score normalization procedure was adopted to avoid inter/intra speaker and token variation (as was also done for the in the production experiment on Sylheti tones).

6.2.2 Subjects and recording

Nine native Sylheti speakers, (7 males and 2 females) aged between 17-28 years of age from the Dharmanagar district of north Tripura produced the scripted sentences designed for this experiment. The experiment was carried out in a quiet environment ensuring zero to minimal noise distortion. Four out of these nine subjects were also part of the previous production study conducted to understand the property of Sylheti tone as discussed in chapter three of this dissertation. All the participants reported that they had no history of speaking or hearing disability.

Data was recorded using a unidirectional head-worn microphone connected to a recorder (Tascam DR- 100MKII) via xlr jack. Subjects were asked to repeat each sentence 4 times and the best three iterations were considered for the final analysis. Altogether 702 tokens 926 words * 3 iterations * 9 subjects) were examined.

6.3 Results and Discussion

6.3.1 The inanimate plural suffix [- tain]

Plural number in Sylheti is marked by adding the inanimate plural suffix [- tain]. This suffix was added to four types of words; viz., underived monosyllabic words with underlying low tones (for example: [zal] 'net'), underived disyllabic words with underlying low tone (for example: [ad̪a] 'ginger', [bala] 'ornament', [bari] 'home'), underived monosyllabic words with underlying high tone (for example [d̪an] 'paddy', [dax] 'drum') and underived disyllabic words with underlying high tone (for example: [bala] 'good', [bari] 'heavy').

A careful observation of the pitch track of noun roots with (underlying) contrastive tonal specification (combined with the inanimate plural suffix [- tain]) suggest that the plural suffix [- tain] is basically toneless, and joins the root to form 'tonal complexes' proposed by Akinlabi and Liberman (2000). The tone of the toneless inanimate plural suffix [- tain] is such that it appears with a high tone after (both monosyllabic and disyllabic) roots which bears a low tone and vice versa, that is, it appears with a low tone when the preceding root is specified with a high tone (Figure 6-1 and Figure 6-2):

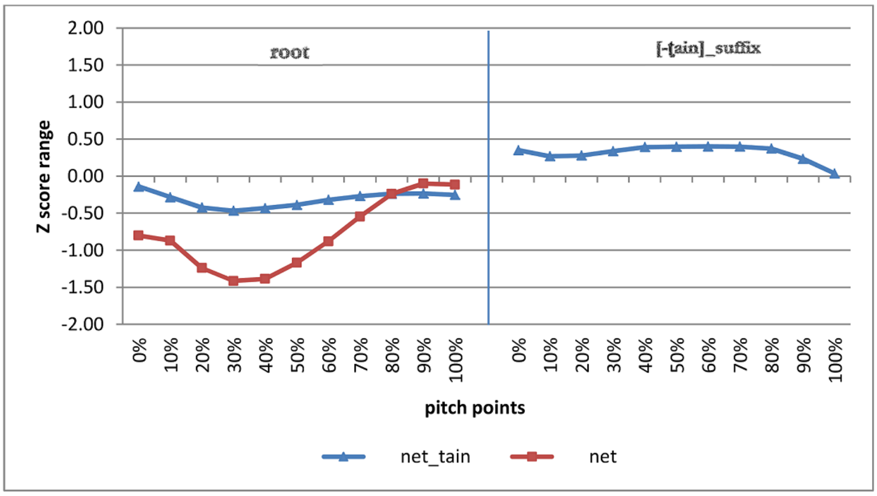

Figure 6.1: Averaged normalized pitch tracks of the word [**zal**] 'net' is shown when produced with and without the plural suffix [- tain]. The underived form of [zal] 'net' is specified for an underlying low tone. The pitch track with rectangular line represents the underived root [zal] without the inanimate plural suffix [- tain], whereas the pitch track with triangular line represents the root with the plural suffix [- tain], (n=27, 9 speakers * 3 iterations each. The suffix [- tain] acquires a high tone after the root with (underlying) low tone (thus presenting an LH contours).

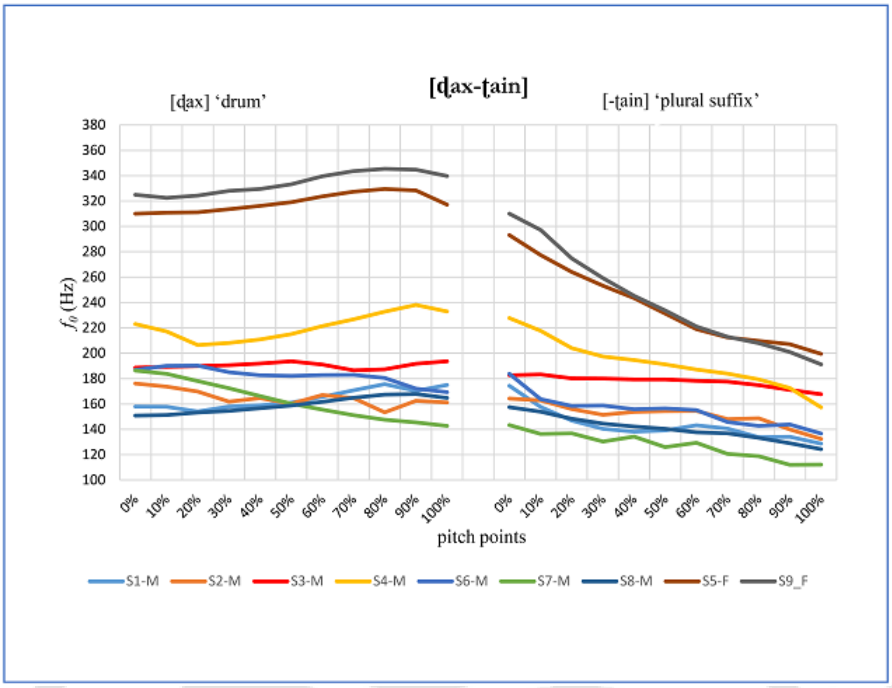

Figure 6.2: Non-normalized pitch tracks of all the speakers for the word [**dax**] ‘drum’, is shown when produced with the inanimate plural suffix [- tain]. The underived form of [dax] ‘drum’ is specified for an underlying high tone [d^h > d] (n = 3 for each speaker).

A similar pattern was also observed in disyllabic words. The tonal property of the inanimate plural suffix [- tain] is the opposite of the one specified for the root; that is, an L tone appears after the high tone root and an H tone appears after the low tone root. Consider the following figures:

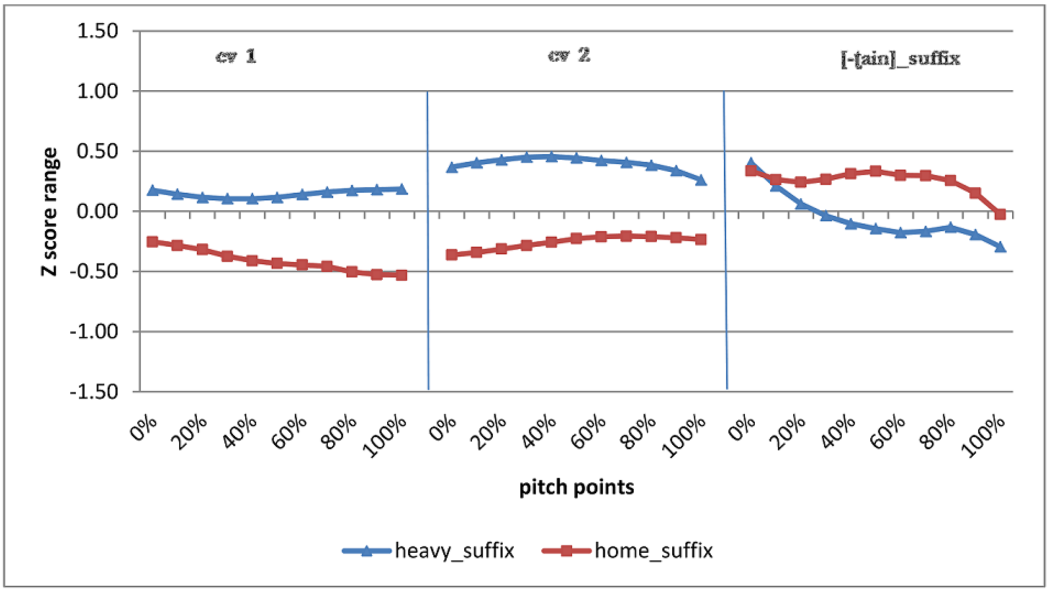

Figure 6.3: Averaged normalized pitch tracks of [**bari**] (both high and low) syllables produced with the inanimate plural suffix [- tain]. The underived root [bari] 'home' is specified for an underlying low tone, whereas the underived root [bari] 'heavy' ([b^h > b] is specified for an underlying high tone (n=27 each, 9 speakers * 3 iterations each. These pitch tracks clearly indicate the process of tone reversal.

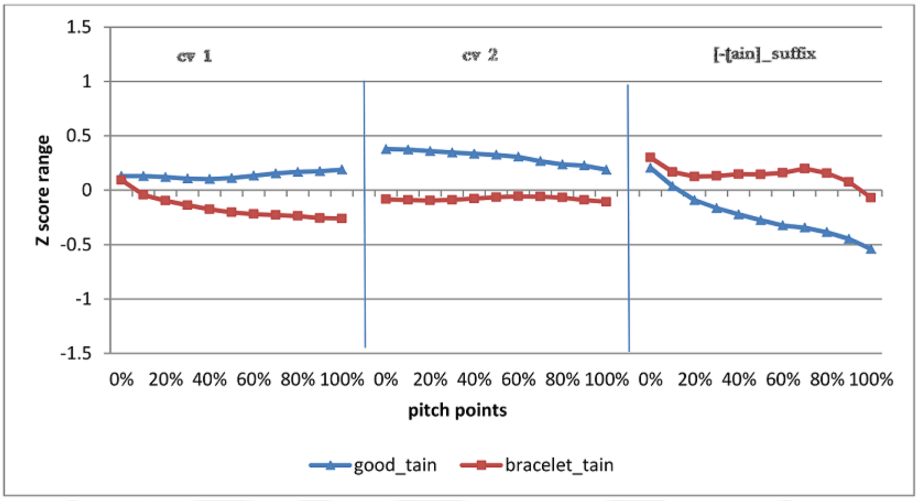

Figure 6.4: Averaged normalized pitch tracks of [**bala**] (both high and low) syllables produced with the inanimate plural suffix [- tain]. The underived root [bala] 'ornament' is specified for an underlying low tone, whereas the underived root [bala] 'good' ([b^h > b] is specified for an underlying high tone, (n=27 each, 9 speakers * 3 iterations each. These pitch tracks clearly indicate the process of tone reversal.

6.3.2 The inanimate definite clarifier [- ta]

The [- ta] suffix in Sylheti is attached after (common) noun roots as a definite classifier. The suffix was added with underived noun roots with (underlying) contrastive tonal specification. Once again it was observed that like the inanimate plural suffix [- tain], the classifier [- ta] is also toneless, and demonstrated a case of tone polarization; that is, a high tone is realized on the suffix after a low tone root and conversely a low tone is realized after a high tone root (Figure 6-5 and Figure 6-6).

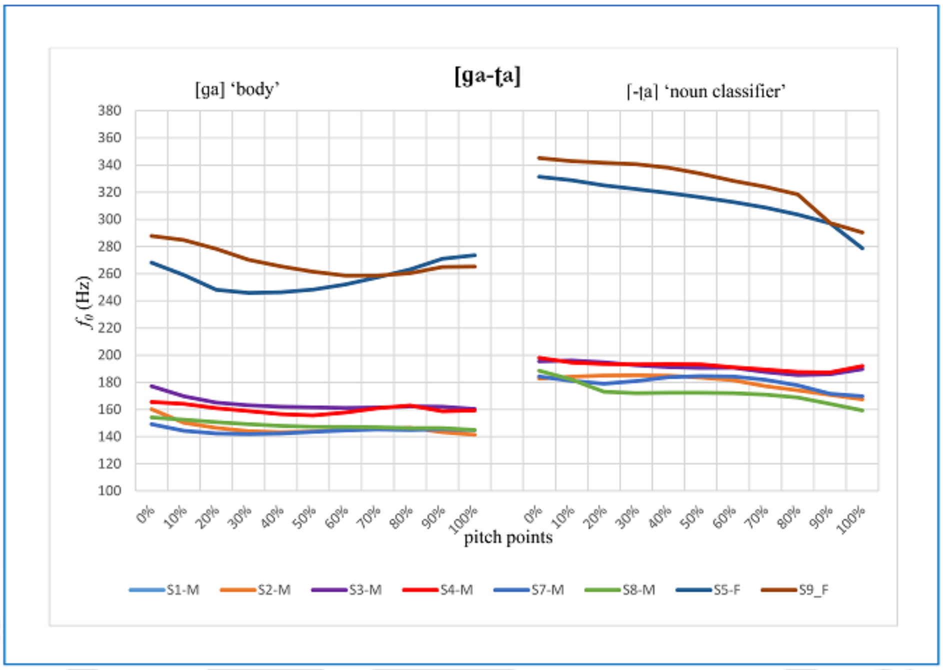

Figure 6.5: Non-normalized pitch tracks as produced by 8 speakers of the word [**ga**] 'body' with the definite classifier [- ta]. The underived form of [ga] 'body' is specified for an underlying low tone (n=3 for each speaker).

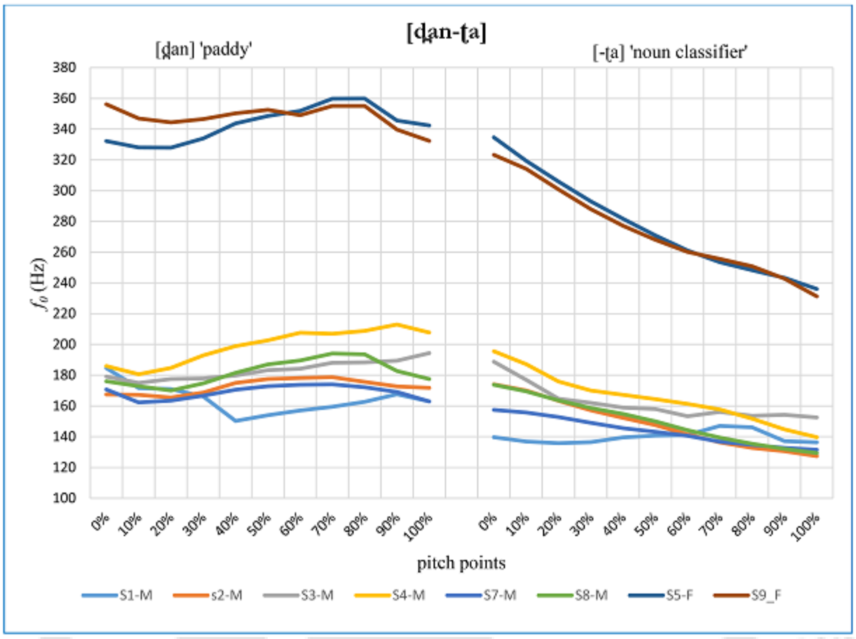

Figure 6.6: Non-normalized pitch tracks as produced by 8 speakers of the word [**d̪an**] 'paddy', with the definite classifier [- ta]. The underived form of the monosyllabic root [d̪an] 'paddy' [$d̪^h$ > d̪] is specified for an underlying high tone. (n=3 for each speaker).

The visual inspection of the figure shown above clearly suggests that the suffix [- ta] is realised with a high tone. However when the same suffix was added with a high tone root, the tonal property of the suffix was changed to a low tone.

The data examined above clearly reveal that the suffixes (both the classifier and the inanimate plural suffix) attached to the noun roots are basically toneless. Thus they take the tone opposite to the tone of the preceding noun roots they are attached to, displaying a clear instance of tonal polarity. When the same suffix (be it the inanimate plural suffix or the definite classifier) was attached to the underived noun roots with (underlying) contrastive tones, we observed that the suffix attached to the root with (underlying) low tone was almost 37 Hz or 26 Mel (on average) higher than when attached to a root with (underlying) high tone property (Figure 6-7).

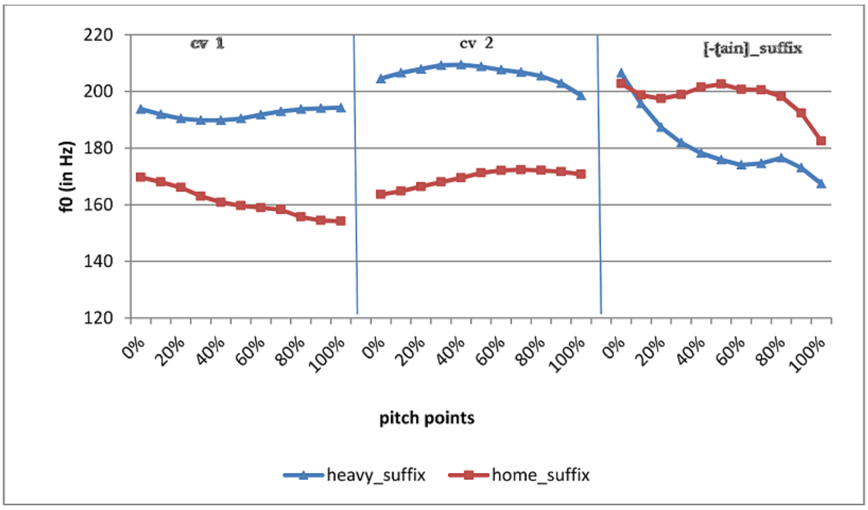

Figure 6.7: Averaged non-normalized pitch tracks of [**bari**] (both high and low) syllables produced by a female speaker with the inanimate plural suffix [- tain]. The underived root [bari] 'home' is specified for an underlying low tone, whereas the underived root [bari] 'heavy' [b^h > b] is specified for an underlying high tone n=3, 3 iterations for each word.

The phenomenon of tonal polarity has mostly been analysed as an upshot of Obligatory Contour Principle (OCP). For example, while analysing the case of tonal polarity in Moore, Kenstowicz et al. (1988) argued that to be a case of tonal dissimilation. In Moorean underlying H tone of a suffix appears to be an L tone on the surface if the stem carries an initial H-tone (that is, if there is an OCP violation) (Kenstowicz et al. 1988, Lee 2013). Contrary to this, other studies claim that tonal polarity surfaces from syllables with toneless input (Anttila and Bodomo 2000, Lee 2013). In the case of Sylheti, we can clearly argue that the inanimate plural suffix and the classifier are clear instances of tone polarity and are due to Obligatory Contour Principle (OCP) which states that adjacent identical elements are prohibited (on the same tier). So we can conclude that the suffixes investigated so far in this experiment do not bear an underlying tone. These suffixes are specified with the property of bearing the tone opposite to the tone in the root (- tone). We will see in the following section that suffixes in Sylheti can also be specified with their own tonal property.

6.3.3 Verb roots with various suffixes

We examined a few verb forms which appeared with the suffixes such as- [-sɛ] '3rd person perfective', [-i] '1st person present indefinite', and [-ɛ] '3rd person present indefinite'. We observed a similar pattern, that is, the verbal suffixes appear with a high tone when attached to a low tone stem,the root still maintains its own (low) tonal property (Figure 6-8).

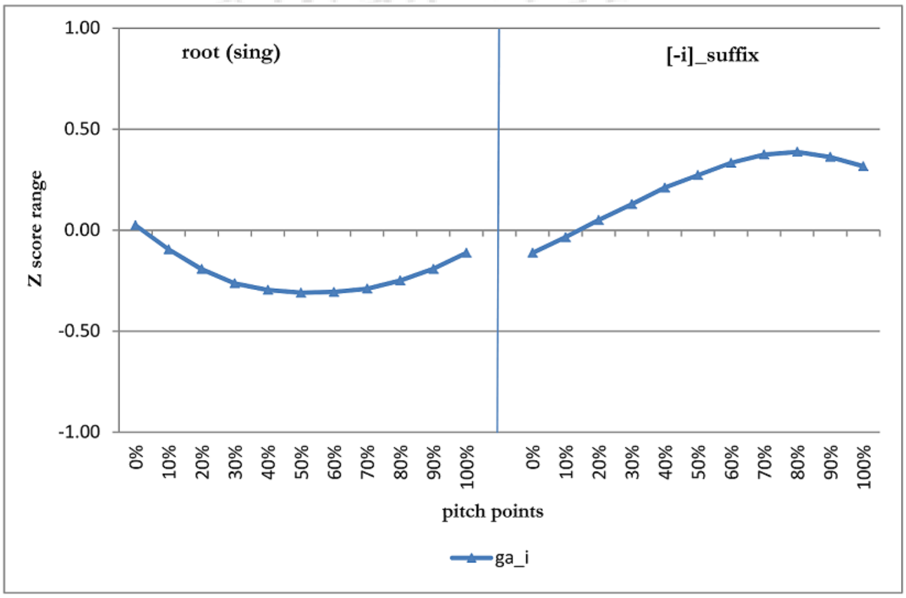

Figure 6.8: Averaged normalized pitch track of the verb form [**ga-i**] 'I/we sing'. The root of the verb is specified with an underlying low tone. (n=27, 9 speakers * 3 iterations each.

The same pattern, that is, the presence of high tone suffix when attached to a low tone stem could also be observed for other suffixes as well. For example, when the 3rd person present indefinite suffix [-ɛ] is attached to the same low stem [a] 'to sing', we can observe a similar tonal contour (Figure 6-9) as seen in Figure 6-8 above.

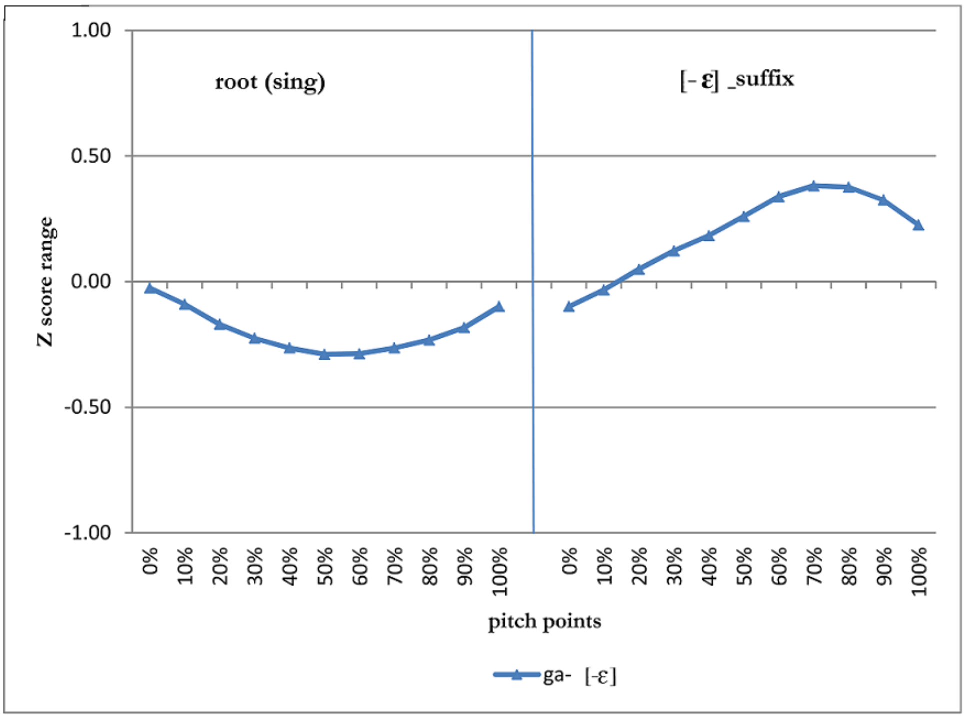

Figure 6.9: Averaged normalized pitch tracks of the verb form [ga-ɛ] 'he sings'. The stem of the verb [ga] carries an underlying low tone. (n=27, 9 speakers *3 iterations each.

Similarly, when the same suffix [-ɛ] is attached to another verb stem [mar] 'beat' with underlying low tone, the suffix appears with an H tone (Figure 6-10).

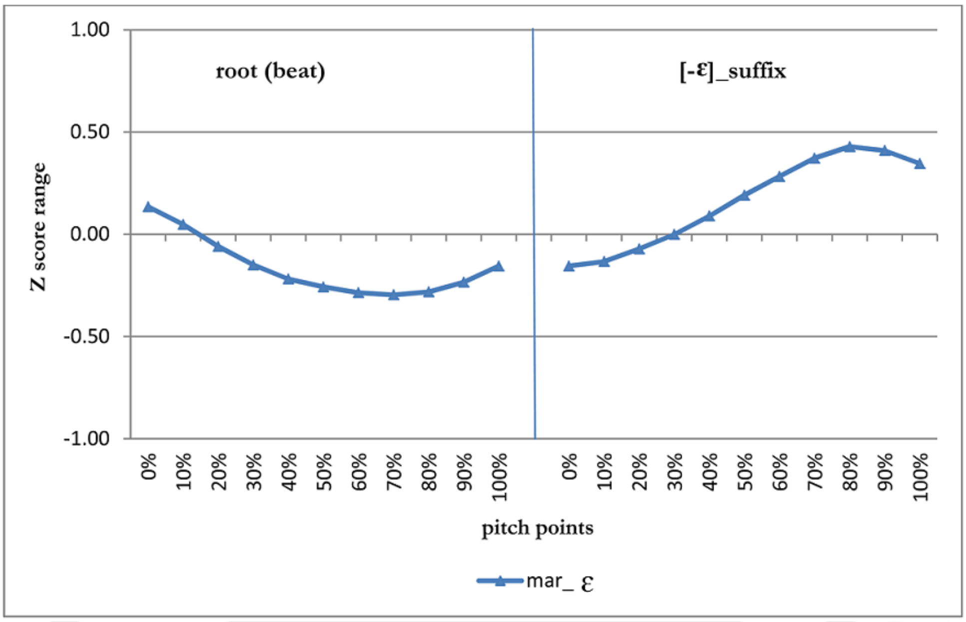

Figure 6.10: Averaged normalized pitch tracks for verb [**mar-ɛ**] 'he beats'. The stem of the verb is specified with an underlying low tone. (n=27, 9 speakers *3 iterations each.

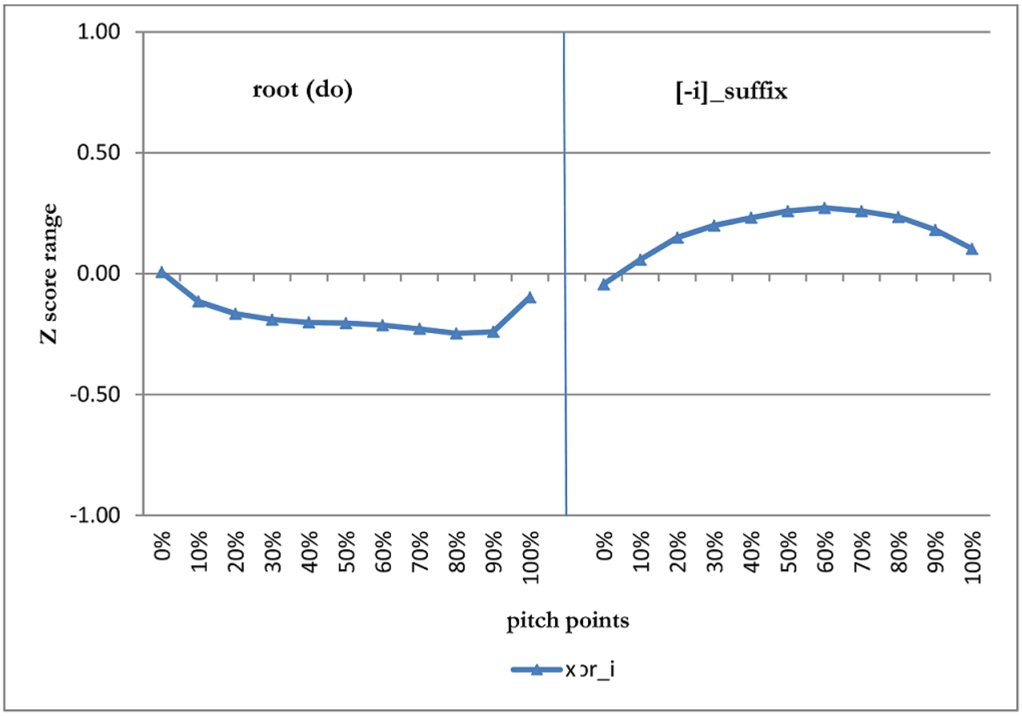

Figure 6.11: Averaged normalized pitch tracks for verb [**xɔr-i**] ‘I/we do’. The stem of the verb is specified with an underlying low tone. (n=27), 9 speakers * 3 iterations each.

Likewise, when the suffix [-i] ‘1st person present indefinite’ is attached to another low tone stem [xɔr] once again the tonal contour of LH pattern could be seen (Figure 6-11).

So far, from all the verb forms we considered it is apparent that these suffixes appear with a high tone when the root carries (an underlying) low tone. This can indicate two possibilities: 1) that this is a case of tonal polarity and 2) these suffixes are specified for underlying high tone.

Tone reversal in some verb forms

When the same suffixes are attached to high tone verb forms, we can notice that these verbal suffixes appear with a high tone- the (underlying) tonal specification of the root is changed to low; the suffix maintains its (underlying) high tone (Figure 6-12 and Figure 6-13). However, no such changes could be seen when the root ends in a Low tone, i.e., the suffix still carries its underlying (high) tone (as could be seen from the tonal contours shown above).

Figure 6.12: Averaged normalized pitch track of the verb form [**t̪ax-i**] 'I/we live/stay'. The stem of the verb is specified for an underlying high tone following the loss of underlying breathiness property of the onset consonant [t̪ʰ > t̪]. (n=27), 9 speakers *3 iterations each.

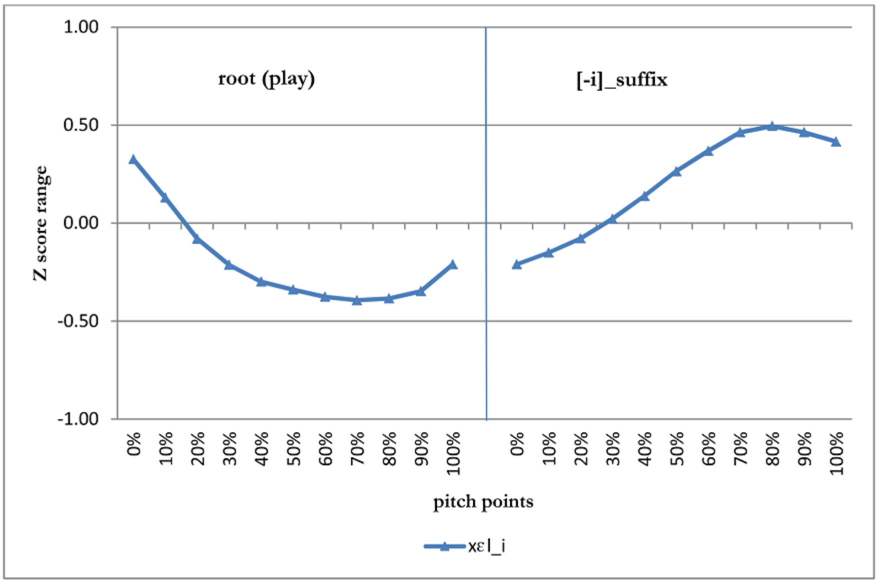

Figure 6.13: Averaged normalized pitch tracks for verb [**xɛl-i**] 'I/we play'. The stem of the verb is specified for an underlying high tone following the loss of underlying breathiness property of the onset consonant, the onset consonant is further spirantized to [x]- [k^h > x]. (n=27), 9 speakers *3 iterations each.

The figures shown above clearly suggest that the suffix [-i] carries its underlying high tone on the surface; the underlying high tone of the verb form is changed to low tone. A similar tonal contour could also be seen for other suffixes as well. For example, when the 3rd person perfective suffix [-se] ($tʃ^h$ > s) is attached with a high tone verb stem [d̪ex] 'see' [k^h > x], the high tone of stem is reversed to low tone, and the suffix [-se] is specified with its underlying high tone, thus presents an LH contour (Figure 6-14).

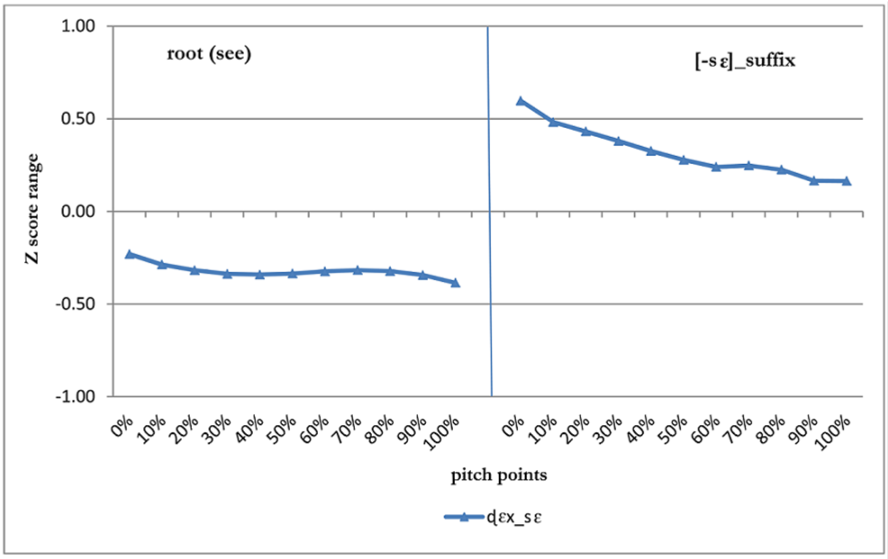

Figure 6.14: Averaged normalized pitch tracks of the verb form [**d̪ex-se**] 'he/she sees'. The stem of the verb is specified for an underlying high tone following the loss of underlying breathiness property of the onset consonant [d̪^h > d̪]. (n=27, 9 speakers, 3 iterations each.

6.4 Discussion/Conclusion

From the discussion presented above we can argue that suffixation in Sylheti presents a case of tonal complexity where 1) in case of noun suffixes (both inanimate plural suffix and the definite classifier) are toneless, these suffixes are specified for the opposite tone of the adjacent stem that they are attached to, and 2) the verbal suffixes are specified for underlying high tone[1].

A similar yet slightly different feature was also recorded in two of the Gur languages such as Dagaare and Kɔnni (Cahill 2004, Antilla and Bodomo 2000). In Dagaare, for example, (noun) suffixes are specified for underlying high tone. The H tone suffix dissimilates to L when next to noun root that ends in an H tone; however, no tone change takes place when the root ends in a low tone (Anttila 2001). Unlike Dagaare,

[1]Prof. Aditi Lahiri (University of Oxford) suggested a different approach altogether (personal communication). She assumes that since Sylheti is a verb final language (SOV pattern), therefore all the underived verb form might be realized with an underlying high tone (default) (effect of prominence as seen in many other Indo Aryan languages), and there is no inherent tonal property attached to the verbal suffix (just like the noun suffixes as examined in this chapter). These toneless verbal suffixes whenever attached to a verb root (with a default underlying high tone), the underlying high tone of the root is shifted to the suffix and the roots are surfaced with a low tone (to avoid identical elements-OCP). However, before the data collection we consulted two of our primary subjects (native speakers of Sylheti) to understand and examine the (underlying) tonal property of the (underived) verb roots considered in this experiment. For this purpose a small production experiment was also conducted using two different contexts- the target verb roots in the middle position of a fixed carrier sentence [I X said] where 'X' is the target verb root; and their (verb roots) occurrence in isolation. After hearing the subjects several times and examining the pitch contours of all the verb roots we were convinced that these verb roots are specified for contrastive tone- the underived form of verb roots where the obstruents lost their breathiness property [+spread glottis] (either in onset or coda position in a syllable) appeared to have higher f_0. For example, the average f_0 of the (underived) verb root [**t̪ax**] 'to stay/live' [t̪ʰ > t̪] was found to be 30 Hz higher than the (underived) verb root of [**xɔr**] when produced by the same female speaker using the same contexts as mentioned above. We understand that the underlying high tone of the verbal suffixes that we have proposed requires robust phonological justification (the origin of the high tone in the suffix). The idea proposed by Prof. Lahiri might open a new direction towards this which we wish to examine in future. We sincerely thank Prof. Lahiri for her valuable suggestions and feedback.

we observed a reverse process in Sylheti- in case of verbal suffixes those are specified for underlying high tone does not change to L, instead they delete the underlying tonal property (of the verb form) into low tone when attached to the roots with underlying high tone. However, no such reversal process could be seen when they are attached to low tone root. Thus, Sylheti verbal suffixes could be argued to belong to the 'dominant category' as proposed by Kiparsky (1982c, 1984a) and Inkelas (1998). The 'dominant suffixes' idiosyncratically transform or delete the structure of tone or stress of the base they are attached to, often but not necessarily substituting a new pattern in place of the deleted structure (Kiparsky 1982c, 1984a, Inkelas 1998).

Furthermore, be it tonal polarity or tone reversal, Sylheti tone assignment in morphologically derived words show that they obey a constraint which requires the root and the suffix to have different tones. Since there are only two tones, if the suffix has a dominant high tone as in the verbal suffixes, the root is bound to change its tone to meet the requirement of a tonal opposition. However, in the case of noun roots, where the tone of the suffix is recessive and therefore it can appear with a low tone if the root is specified for underlying high tone.

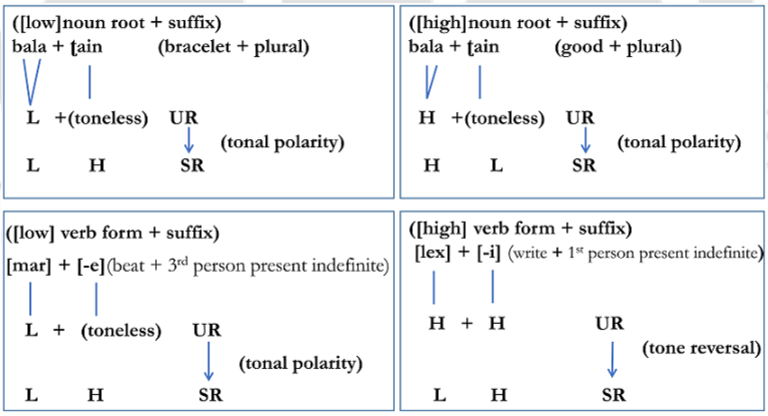

Figure 6.15: Process of (tonal) suffixation in Sylheti. Noun (root) + suffix = tonal polarity ('recessive suffix'), verb (form) + suffix = tone reversal ('dominant suffix').

Such tonal complexity was not predicted for Sylheti which is supposed to have developed tone at a later stage of the language. It obviously throws open a lot of questions for tonogenesis as there is no obvious reason for the development of such tonal complexes as this cannot be traced to the restructuring of the phoneme inventory in Sylheti. It means that Sylheti has phonologized tone not only where there has been loss of phonation but also in various unpredictable ways, such as developing morpho-phonemic tonal processes as the ones just described: tonal dissimilation and tone reversal. Such processes also have not been recorded for Tibeto-Burman languages spoken in the vicinity nor in other Indo-Aryan languages which have tone, e.g. Punjabi.

Chapter 7

Conclusion, Implications and future research

Major findings

In this study we have been able to show the way historical development of sound changes viz. loss of underlying breathiness property is compensated with the phonological realization of lexical tone. The major goal of this dissertation was to explore the tonal inventories of Sylheti and to provide a detailed phonetic and phonological account of Sylheti tonogenesis. These are achieved through detailed acoustic and statistical means. Further, this work also attempted to understand the way tone is perceived by the native speakers of Sylheti. In addition to that this research also addresses the issue of tone assignment in a morphologically derived process such as the process of suffixation. In the following sections the major findings of this study are summarized. Towards the end of this chapter the limitations of this study and scope of future work have been discussed.

7.1 Phoneme inventories in Sylheti

One of our primary objectives was to explore the phoneme inventory of Sylheti. The general observation of consonant reduction due to the phonological process of consonant weakening involving the loss of underlying breathiness contrast, spirantization and deaffrication has been examined with the aid of systematic production experiments. VOT experiment on Sylheti voiced obstruents confirmed the loss of underlying breathiness contrast in Sylheti. We did not observe any noise following the burst of underlying voiced stops. As expected, the (underlying) voiced aspirated stops did not show any difference with their unaspirated counterparts ($p > 0.05$, [$F(1, 359) = 0.095$, $p = 0.76$)]. The processes of spirantization and deaffrication have been discussed with the assistance of acoustic waveforms. Further, the acoustic experiment conducted on Sylheti vowels confirmed the reduction of vowel inventory- unlike Standard Colloquial Bengali which has seven vowels (Chatterjee 1926) we report on the presence of only five vowels ([i], [ɛ], [a], [ɔ] and [u]) in Sylheti. High vowels [i] and [u] were found to be significantly shorter than the rest of the vowels. Formant values were measured at the mid-point and the normalized F1 and F2 values were used to plot the vowel diagrams in Sylheti. 'Euclidean distances' between vowel pairs in Sylheti vowel inventory were calculated separately for male and female speakers. As expected, we observed the biggest differences among the peripheral vowels [i], [u], and [a]. The pattern was found to be same in both male and female speakers' data. The pair of [i] and [ɛ] showed the smallest difference for both male and female speakers. The average Euclidean distance and overall vowel space is found to be less for male speakers (108 Mels) as compared to female speakers (168 Mels).

7.2 Tone inventories

The principal thrust of this dissertation was to understand the tonal property of the homophonous pair (borne out due to the loss of either underlying breathiness property [+spread glottis] or due to spirantization/deaffrication). This has been examined and explored through various acoustic components such as f_0 (measured at various points), duration and intensity, followed by ample statistical evidences, this study has established the presence of lexical tone in Sylheti. Since high tone realization from the loss of the breathiness property is not predicted, we conducted extensive spectral measuremnts on the tone bearing vowels. The results showed that there is indeed some remnant differences between the spectral properties of vowels bearing high and low tones which is discussed in section 7.3.

As far as the number of tones is concerned, this study concludes that there are at least two tones- the phonological realization if high tone is achieved through the loss of (underlying) breathiness property of Sylheti obstruents (regardless of voicing contrast) and a low tone elsewhere. However, this predictability does not seem to work always in Sylheti- as such every homophonous pair is realized with contrastive tone. So for example, if the conditioning environment of the homophonous pair is somewhat similar (for example, loss of breathiness property of the onset obstruent as in the pair of [xál] 'drain/channel' [k^h > x], and [xàl] 'skin' [$tʃ^h$ > x]), then one of the homophonous pair would be realized with a clear low tone. In other words, a high tone can also be observed following a voiceless obstruent (as could be seen in the homophonous pair of ([ɸɔ́r] 'read' [p > ɸ], and [ɸɔ̀r] 'guard' [p > ɸ]. A contrastive tone can also be observed if both the homophonous words bear (similar) identical (onset) obstruent as could be seen in the pair of [mélá] 'much/many', and [mèlà] 'fair').

In case of underived disyllabic words, both the syllables are specified for a singleton tone. So for example, if the first syllable of a disyllabic word starts with a higher f_0, the same pattern of high f_0 would be spread to the following syllable, similarly if the first syllable starts with a lowered f_0, then the same pattern is maintained in the succeeding syllable. The distinct pitch tracks representing high and low contour is maintained throughout the entire rhyme or voiced part of a syllable. This is observed for both monosyllabic and disyllabic words.

However, the(possible) presence of a (phonological) third tone in Sylheti remains somewhat speculative and requires further investigation. When we carried out statistical measurement, the factor tone appears to carry a strong significant effect on all acoustic measures related to f_0(measured at vowel mid, mean f_0, max f_0, and min f_0) in the Repeated Measure ANOVA. We did not observe any significant effect of vowel duration and intensity as a function of tone.

7.3 Phonation qualities of the vowels carrying contrastive tone

The phonation qualities of vowels carrying contrastive tones are explored by examining various acoustic components which essentially led to the understanding of Sylheti tonogenesis. The acoustic properties associated with vowels carrying contrastive tones confirm that the vowels carrying high tones have slightly 'tenser' voice quality than the vowels carrying the low tones (which have 'laxer' phonation quality). Following that, we propose a two-fold evolution of Sylheti tonogenesis- first, the historical development of the loss of intended feature [+spread glottis] associated with (voice) obstruents is indeed reinterpreted and readjusted with a perturbed f_0 in the neighbouring vowel(s); secondly, due to hypo-correction where the listeners probably fail to normalize the coarticulatory effects such as the effects of [+spread glottis] on f_0, the vowels following the obstruents (with the history and underlying breathiness property) might have been compensated with a perturbed f_0. Since the breathiness property of the consonants (obstruents) was readjusted on the adjacent vowels, the vowels might have acquired a property similar to 'tenser' phonation in order to maintain the lexical distinction of homophonous words. Thus the syllable of those words (with the history of underlying breathiness property either in onset or coda position) is surfaced with an increased f_0. Once the conditioning environment (for example, the feature [+spread glottis]) was lost, the f_0 patterns were phonologized for the homophonous words with contrastive lexical tones. We will have to conduct more research on other supposed dialects of Bengali spoken in the contiguous neighbouring areas to see if our speculation regarding the emergence of tone in Sylheti bears out. That is, the presence of perturbed f_0 due to the loss of breathiness contrast in those dialects may indicate that a similar process is under way in those dialects as well.

7.4 Tone perception

As the tonal system of a given language must 'converge with fundamental processes associated with the production and perception of tones' (Gandour 1979), an experiment was conducted on 12 native speakers using synthesized stimuli in a fixed sentence frame- the same sentence frame was also used for the production experiment of tone (as discussed in the third chapter of this dissertation). Eight monosyllabic homophonous pairs were carefully chosen to create the synthesized stimuli. The stimuli were prepared by varying (shifted up or down) the pitch contours of the contrastive tones at every 20% of the rhyme and were embedded in a natural sentence frame. Experimental findings confirm that Sylheti native speakers exploit f_0 based information to identify contrastive meaning of the homophonous pairs. It was further observed that the categorical manner adopted by the native speakers to perceive lexical meaning requires f_0 to be maintained at least till 60% of the total rhyme/voiced portion of a syllable.

7.5 Tone assignment in morphologically derived words: the process of suffixation in Sylheti

Our final goal was to understand the nature of tone alignment in morphologically derived environment. To meet that goal and understand the interaction of lexical tones with various suffixes available in Sylheti we conducted our final production experiment. Following the experimental findings we argued that suffixation in Sylheti presents a case of tonal complexity where- 1) in case of noun suffixes (both plural inanimate suffix and the definite classifier) are toneless, these suffixes are specified for the tone opposite to tone of adjacent stem they are attached to- thus displaying a clear instance of tonal polarity (avoid adjacent tone), and 2) the verbal suffixes are specified for underlying high tone- where the dominant high tone of the (verbal) suffix overwrites the underlying high tone of the root- leading to tonal dissimilation.

7.6 Implications and future research

The findings of this study is expected to enrich the concept of tonal phonology in general and Sylheti in particular. The tonogenetic property leading to the development of high tone following the loss of underlying breathiness property of Sylheti obstruents is certain to contribute to the existing theoretical knowledge. The findings also reveal that the general predictions of consonant tone interaction and the possible outcome ([aspirated] voiceless obstruent = high tone, [aspirated] voiced obstruent = low tone) as claimed in the canonical literature of tonogenesis needs to be amended. Contrary to the claims made in the literature of tonogenesis, we observed a high tone following the loss of underlying breathiness contrast of Sylheti obstruents. Even the voiceless obstruents can generate a low tone as we have observed in this language.

In addition to that, this study also provides the first acoustic analysis of phoneme inventory of a previously undocumented language. We also examined the phonation qualities of vowels carrying contrastive tones in detailed. The methodology adopted for examining tone perception in Sylheti provided crucial information- Sylheti native speakers exploit f_0 perturbations and perceive the lexical meaning of contrastive pairs in categorical manner. Amidst the above mentioned implications, this study also has certain limitations and requires further investigations. We have already mentioned that the phonological realization of the third tone requires an elaborate study with more homophonous pairs with three (or more) way lexical contrast. The morpho-phonology of Sylheti involving the process of prefixation, reduplication, tone assignment in compound words, have also not been discussed in this study. Further, the process of suffixation leading to tonal dissimilation as observed in case of verbal suffixes and the historical origin of underlying high tone in the verbal suffixes requires phonological justification[1]. Considering the distinct phonetic and phonological properties of Sylheti along with the tonal complexities as explained in this study makes it clear that this language is very different than Standard Colloquial Bangla. Further, the (long) existence of its own script in *Sylhoti Nagari* puts a question mark whether Sylheti should

[1]Sincere thanks to Prof. Aditi Lahiri for her valuable feedback (personal communication).

really be considered as a variety of Bangla.

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. Abramson, A. S. (2004). *The Plausability of Phonetic Explanations of Tonogenesis. in Gunnar Fant et al. (eds), From Traditional Phonology to Modern Speech Processing: Festschrift for Professor Wu Zongji's 95th Birthday* (pp. 17-29) Beijing: Foreign Language Teaching and Research Press.

3. Akinlabi, A. (1985). Tonal Underspecification and Yoruba Tone. PhD Dissertation, Universityof Ibadan.

4. Akinlabi, A., and M. Liberman.(1995). On the phonetic interpretation of the Yoruba tonalsystem. *Proceedings of the International Congress of Phonetic Sciences 1995.* Stockholm, Sweden, pp. 425.

5. Akinlabi, A & M. Liberman. (2000a). The Tonal Phonology of Yoruba Clitics. In B. Gerlach and J. Grijzenhout (eds.) *Clitics in Phonology, Morphology and Syntax. Amsterdam: Benjamins.* Pp. 31 62.

6. Akinlabi, A & M. Liberman. (2000b). Tonal complexes and tonal alignment.Paper given at the Tone Symposium, University of Tromso. To appear in the **Proceedings of NELS** 31.

7. Alves, Mark. (1995). Tonal Features and the Development of Vietnamese Tones. *Working Papers in Linguisticsi (Vol. 27: 1-13).* Department of University of

Hawaii. Manon.

8. Anderson, S. (1974). *The Organization of Phonology.* New York: Academic Press.

9. nderson, S. (1978). Tone features. In V. Fromkin (ed.), *Tone: A Linguistic Survey.* New York: Academic Press, pp. 13376.

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

11. Antilla, A., and A. Bodomo. 1996. Stress and tone in Dagaare. MS, Stanford University.Also ROA # 1691296.

12. Anttila, Arto. (2001). Variation in Finnish Phonology and Morphology. Dissertation summary. *Glot International* 5(1), 1-3.

13. Anttila, Arto and Adams Bodomo. (2000) Tonal polarity in Dagaare. In Vicki Carstens and Frederick Parkinson (eds.), *Trends in African Linguistics 4: Advances in African Linguistics*, Africa World Press, Trenton, New Jersey. pp. 119-134.

14. Baart, J.L.G. 1997. *The Sounds and Tones of Kalam Kohistani.* Arlington, TX: Summer Institute of Linguistics and National Institute of Pakistan Studies.

15. Bao, Z. (1999). *The structure of tone.* Oxford: Oxford University Press.

16. Bhaskararao, Peri. (1998). Voiced Aspiration and Tonogenesis in Some South-Asian Languages. In Kaji., Shigeki (Edt.). *Proceedings of the Symposium on Cross-Linguistic Studies of Tonal Phenomena: Tonogenesis, Typology, and Relate Topics.* Tokyo: ILCAA. (337-51 pp.)

17. Bhatia, Tej K. (1975). The Evolution of Tones in Punjabi. *Studies in the Linguistic Sciences*, 5/2 (Fall 1975): 12-24.

18. Bhattacharya, Krishna. (1910). *Bengali Phonetic Reader*. CIIL, Mysore.

19. Bickley, C. (1982). Acoustic analysis and perception of breathy vowels, *Speech Communication Group Working Papers I Research Lab. of MIT* pp. 71-82.

20. Blankenship, Barbara. (2002). The Timing of Nonmodal Phonation in Vowels. *Journal of Phonetics* Vol.30, pp 163-191.

21. Blankenship, Barbara. (1997). *The time course of breathiness and laryngealization in Vowels*. PhD dissertation, University of California, Los Angeles.

22. 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, Burling, Robbins. (1959). Proto-Bodo. *Language* 35, 433-453.

23. Cahill, Michael. (2004). Tonal polarity in Kɔnni nouns. Studies in African Linguistics 33(1):133.

24. Cahill, Michael. (1999). Aspects of the phonology of labial-velar stops. Studies in African Languages 28(2):155184. Cahill, Michael. (1999). From endangered to less endangered: Case histories from Brazil and Papua New Guinea. SIL Electronic Working Papers 1996-001.

25. Chatterjee, Suniti Kumar. (1926). *The Origin and Development of the Bengali Language*. London: George Allen and Unwin.

26. Chuang, C.K., Hiki, S., Sone, T. and Nimura, T. (1972). The acoustic features and perceptual cues of the four tones of standard colloquial Chinese. Proceedings of the 7th International Congress of Acoustics (Vol. 3): 297-300. Budapest: Akademial Kiado.

27. Coupe, A. R. (2003). *A phonetic and phonological description of Ao: a Tibeto-Burman language of Nagaland. Canberra: Pacific Linguistics.*

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

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

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

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

32. Doherty, B. (1993). *The acoustic-phonetic correlates of Cayuga word-stress.* Harvard University Ph.D. dissertation.

33. Dunn, J. (1979). *A reference grammar for the Coast Tsimshian language.* Ottawa: National Museums of Canada.

34. Dutta Indranil. (2009). *Acoustics of stop consonants in Hindi: Voicing, fundamental frequency and spectral intensity.* Verlag Dr. Mller: Saarbrcken.

35. DiCanio, Christian. (2009). The Phonetics of Register in Takhian Thong Chong. *Journal of International Phonetic Association* Vol. 39(2), pp. 162-188.

36. Edmondson, J.A., J. Bateman and H. Miehle. (1992). Tone contours and tone clusters in Iau. In L. Buszard-Welcher, J. Evans, D. Peterson, L.Wee and W. Weigel (eds.), *Special Session on the Typology of Tone languages. Proceedings of the 18th Annual meeting of the Berkeley Linguistics Society. Berkley: Berkley Linguistics Society,* pp. 92-103.

37. England, N. (1983). A grammar of Mam, a Mayan language. Austin: University of Texas Press.

38. Esposito, C.M. (2012). An Acoustic and Electroglottographic Study of White Hmong Phonation. *Journal of Phonetics*. Vol. 40(3), pp. 466476.

39. Esposito, C.M. & S.D. Khan. (2012). Contrastive Breathiness Across Consonants and Vowels: A comparative study of Gujarati and White Hmong. *Journal of the International Phonetic Association*. Vol. 42(2), pp. 123-143.

40. Esposito, C.M. (2011). The Perception of Pathologically-disordered Phonation by Gujarati, English and Spanish Listeners. *Language and Speech*. Vol. 54(3), pp. 116.

41. Esposito, C.M. (2010a). The Effects of Linguistic Experience on the Perception of Phonation. *Journal of Phonetics*. Vol. 38(2), pp. 303316.

42. Esposito, C.M. (2010b). Variation in Contrastive Phonation in Santa Ana del Valle Zapotec. *Journal of the International Phonetic Association*. Vol. 40(02), pp. 181198.

43. Esposito, C. M., Khan, S.D. & A. Hurst. (2007). Breathy Nasals and /Nh/ Clusters in Bengali, Hindi, and Marathi. *Indian Linguistics*. Vol. 68(34), pp. 275299.

44. Fischer-Jorgensen, Eli. (1967). Phonetic Analysis of Breathy (murmured) Vowels. *Indian Linguistics*. Vol. 28, pp. 71-139.

45. Flemming, E. (1995). Auditory Representations in Phonology. Doctoral dissertation, UCLA, Los Angeles.

46. Flemming, E., Ladefoged, P., & Thomason, S. (1994). Phonetic structures of Montana Salish, UCLA *Working Papers in Phonetics* 87, 1-33.

47. Fok, C.Y.Y. (1974). A perceptual study of tones in Cantonese. Hong Kong: University of Hong Kong Press.

48. Fromkin, V.A. (ed.). (1978). *Tone: A Linguistic Survey*. New York: Academic Press.

49. Gandour, Jack. (1974). On the representation of tone in Siamese. UCLA *Working papers in Phonetics* 27: 118-146.

50. Gandour, Jack. (1977). On the interaction between tone and vowel length: evidence from Thai dialects. *Phonetica* 34: 54-65.

51. Gandour, Jack. (1978). The perception of Tone. In Fromkin, V.A. (ed.). *Tone: A Linguistic Survey.* New York: Academic Press. pp. 41-76.

52. Gandour, Jack. (1979) "Perceptual Dimensions of Tone: Thai." South-east Asian Linguistic Studies Vol. 3. Ed. Nguyen Dang Liem. 3: 277-300.

53. Gandour, Jack. (1979) "Perceptual Dimensions of Cantonese Tones: a Multidimensional Scaling Reanalysis of Fok's Tone Confusion Data." South-east Asian Linguistic Studies Vol. 4. Ed. Nguyen Dang Liem. 4: 415-429.

54. Gandour, Jack. (1979). "Perceptual Dimensions of Cantonese Tones: a Multidimensional Scaling Reanalysis of Fok's Tone Confusion Data." South-east Asian Linguistic Studies Vol. 4. Ed. Nguyen Dang Liem. 4: 415-429.

55. Goldsmith, John. (1976). Autosegmental Phonology. PhD Dissertation. MIT.

56. Goldsmith, John. (1984). Meeussen's Rule. In M. Aronoff and R. Oehrle (eds.) *Language Sound Structure.* Cambridge, MA: MIT Press, pp. 245-9.

57. Goldsmith, John. (1990). *Autosegmental Phonology and Metrical Phonology.* Cambridge: Basil Blackwell.

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

59. Gope, Amalesh & Mahanta, Shakuntala (2014). "Lexical tones in Sylheti", In TAL-2014, 10-14.

60. Gope, Amalesh & Mahanta, Shakuntala (2015). An Acoustic Analysis of Sylheti Phonemes. Proceedings of the 18th International Conference of Phonetic Sciences (ICPhS 2015).

61. Gope, Amalesh & Mahanta, Shakuntala (Forthcoming- accepted). "Correlation between Sylheti tone and Phonation". Proceedings of the Speech Prosody 2016.

62. Gope, Amalesh & Mahanta, Shakuntala. (2016b). "Correlation between Sylheti tone and Phonation". Proceedings of the Speech Prosody 2016, Boston, Lincolnshire, England.

63. Gope, Amalesh & Mahanta, Shakuntala. (Forthcoming- Accepted). "Perception of Lexical Tones in Sylheti". Proceedings of the TAL-2016, Buffalo, NY.

64. Gordon, Mathew. (1996). The Phonetic Structures of Hupa. UCLA *Working Papersin Phonetics* Vol. 93, pp. 164-187. Gordon, Mathew. (1998). The Phonetics and Phonology of Non-modal Vowels: a cross linguistic perspective. *Berkeley Linguistic Society* Vol. 24, pp. 93-105.

65. Gordon Mathew and Ladefoged, Peter. (2001). Phonation Types.: A Cross Linguistic Overview. *Journal of Phonetics* Vol. 29, pp. 383-406.

66. Gordon, Mathew. (2001). Linguistic aspects of voice quality with special reference to Athabaskan. *Proceedings of the 2001 Athabaskan Languages Conference*, pp. 163-178.

67. Gordon, Raymond G., Jr. (Ed). (2005). Ethnologue: Languages of the World, Fifteenth edition. Dallas, Tex: SIL International Version.

68. Grierson, George. A. (1903). *The Linguistic Survey of India.* Vol. V. Indo-Aryan Family. Eastern Group. Part I. Specimens of the Bengali and Assamese languages. Calcutta: Office of the Superintendent of Government Printing, India.

69. Gussenhoven, Carlos., and P. van der Vliet. (1999). Tone and Intonation in Venlo. *Journal of Linguistics* 35.1: 99-136.

70. Gussenhoven, Carlos. (2000). The boundary tones are coming: on the non-peripheral realization of boundary tones. In M. Broe and J. Pierrehumbert (eds.), *Acoustic and the Lexicon: Papers in Laboratory Phonology* V. Cambridge: *Cambridge University Press*, pp. 132-151.

71. Gussenhoven, Carlos. (2004). The Phonology of Tone and Intonation. Cambridge: Cambridge.

72. Hanson, Helen M. (1995). *Glottal characteristics of female speakers.* PhD dissertation, Harvard University.

73. Hanson, H. and Chuang, E. (1999). Glottal characteristics of male speakers: acoustic correlates and comparison with female data. *Journal of the Acoustical Society of America* 106, 1064-77.

74. Hanson, H. M., Stevens, K. N., Kuo, H.-K. J., Chen, M. Y. and Slifka, J.(2001). Towards models of phonation, *Journal of Phonetics* 29, 451-480.

75. Haudricourt, Andr-Georges. (1954). De l'origine des tons en Viêtnamien. JA 242:69-82.

76. Hayes, Bruce. (1995). *Metrical Stress Theory: Principles and Case Studies.* Chicago: University of Chicago Press.

77. Hayes, Bruce. and Aditi Lahiri. (1991). Bengali Intonational Phonology. *Natural Languages and Linguistic Theory.* 9: 47-96.

78. Hirose, H.(1977). Laryngeal adjustment in consonant production, *Phonetica*34, 289294.

79. Hirose, H. (1995). Investigating the physiology of laryngeal structures. In W. Hardcastle & J. Laver (eds.), *The handbook of phonetic sciences*, pp. 116-36. Cambridge, MA: Blackwells.

80. Holmberg, E., Hillman, R., Perkell, J., Guiod, P., & Goldman, S. (1995). Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice, *Journal of Speech and Hearing Research* 38, 1212-23.

81. Hombert, Jean-Marie. (1977). Consonant types, vowel height and tone in Yoruba. *Studies in African Linguistics*, Vol. 8, pp. 173-190.

82. Hombert, Jean-Marie. (1978). Consonant types, vowel height and tone in Yoruba. In Victoria A. Fromkin (ed.) *Tone: A linguistic survey.* New York: Academic Press. pp. 77-111.

83. Hombert, Jean-Marie. Marie., John. J. Ohala. and William G. Ewan. (1979). Phonetic Explanations for the Development of Tones. *Language* Vol. 55 (1), pp. 37-58.

84. Huffman, M. K. (1987). Measures of phonation type in Hmong. *Journal of the Acoustical Society of America* Vol. 81, pp. 495-504.

85. Hyslop, Gwendolyn. (2009). Kurtop Tone: A tonogenetic Case Study. *Lingua, Vol. 19, Issue 6.* Elsevier.

86. Inkelas, Sharon. (1998). The theoretical status of morphologically conditioned phonology: a case study from dominance. Yearbook of Morphology 1997:121-155.

87. 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.

88. Ishihara, Shunichi. (1999). An acoustic-phonetic descriptive analysis of pitch realisations in Kagoshima Japanese. *Proceedings of the 1999 Conference of the Australian Linguistic Society.*

89. James, D. (1981). An autosegmental analysis of Siane. MS Summer Institute of Linguistics, Oklahoma.

90. Javkin, H. & Maddieson, I. (1985). An inverse filtering study of creaky voice. UCLA *Working Papers in Phonetics* 57, 115-25.

91. Joseph, U. V. and Robbins Burling. (2001). Tone Correspondences among the Bodo Languages. *Linguistics of the Tibeto-Burman Area.* 24(2), 41-55.

92. Joseph, U. V. and Robbins Burling. (2007). *The Comparative Phonology of the Boro-Garo Languages.* Mysore: Central Institute of Indian Languages.

93. Jun, J. (1995). *Perceptual and articulatory factors in place assimilation: An Optimality-theoretic approach.* UCLA Ph.D dissertation.

94. Jun, J. (1996). Place assimilation as the result of conflicting perceptual and articulatory constraints, *West Coast Conference on Formal Linguistics* 14, 221-38.

95. Kenstowicz, Michael. (1988). Tone and accent in Kizigua–a Bantu language. In P.M. Bertinetto & M. Loporcaro eds. Certamen Phonologicum I, Proceedings of the Cortona Phonology Conference. Torino: Rosier. 177- 188.

96. Khouw, Edward and Ciocca, Valter. (2007). Perceptual correlates of Cantonese tones. *Journal of Phonetics*, Volume 35, Issue 1, January 2007, Pages 104-117.

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

98. Kingston, John. and David Solnit. (1989). The inadequacy of underspecification. *Papers from the Annual Meeting of the North East Linguistic Society,* Vol. 19, pp. 264-278.

99. Kingston, John. (2011). Tonogenesis. in Oostendorp, Marc van et.al (Eds) *The Blackwell Companion to Phonology* (Vol. IV). Blackwell.

100. Kiparsky, Paul. (1982a). '*How are the levels linked?*' ms, Standford University.

101. Kirk, Paul. L., Jenny Ladefoged. and Peter Ladegoged. (1993). Quantifuing acoustic properties of modal, breathy and creaky vowels in Jalapa Mazatec. In *American Indian linguistics and ethnography in honor of Laurece C. Thompson* (A. Mattina and T. Montler, eds), Missoula, MIT: University of Montana Press.

102. Klatt, Dennis H. (1973). Discrimination of Fundamental Frequency Contours in Synthetic Speech: Implications for Models of Pitch Perception. *Journal Acoustic Society of America*, 53.1:8-16.

103. Klatt, Dennis. and Klatt, Laura. (1990). Analysis, synthesis and perception of voice quality variations among female and male talkers. *Journal of Acoustical Society of America*, Vol. 87, pp. 820-857.

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

105. Ladefoged, Peter and Maddieson, Ian, and Jackson M. (1988). Investigating phonation types in different languages. In O. Fujimura (ed.), *Vocal fold physiology: voice production, mechanisms and functions*, pp. 297-317. New York: Raven Press.

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

107. Ladefoged, Peter. (1983). The linguistic use of different phonation types. In Diane Bless and James Abba (eds), *Vocal fold physiology: Contemporary research and clinical issues*, pp. 351-360. San Diego, CA: College-Hill Press.

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

109. Lalrindikii.(1989). *Some Aspects of Autosegmental Phonology of English and Mizo.* M Litt dissertation. Hyderabad: Central Institute of English and Foreign Languages.

110. Laver, J. (1980). *The phonetic description of voice quality.* Cambridge: Cambridge University Press.

111. Leben, William. (1973). *Suprasegmental Phonology.* Ph.D. dissertation, MIT.

112. Leben, William. (1978). The representation of tone. In Fromkin, V.A. (ed.). *Tone: A Linguistic Survey.* New York: Academic Press. 177219.

113. Leer, J. (1979), *Proto-Athabaskan verb stem variation, part one: phonology.* Fairbanks: Alaska [Native Language Center Research Papers 1].

114. Lehiste, I. (1965). Juncture. *Proceedings of the 5th International Congress of Phonetic Sciences*, pp. 172-200.

115. Lehiste, I. (1979). Sentence boundaries and paragraph boundariesperceptual evidence. *In The Elements: a Parasession on Linguistic Units and Levels*, Chicago Linguistics Society 15, pp. 99-109.

116. Maddieson, I. and Ladefoged, P. (1985). 'Tense' and 'lax' in four minority languages of China, *Journal of Phonetics 13*, 433-54.

117. Mahanta, Shakuntala, Das, Kalyan and Gope, Amalesh (2016a). On the Phonetics and Phonology of Focus marking in Boro. Proceedings of the Annual Meeting of Phonology 2015, Vancouver, Canada.

118. Matisoff, J. A. (1973). Tonogenesis in Southeast Asia. In L. M. Hyman (Ed.), *Consonant types and Tone* (Vol. 1, pp. 73-95). Los Angeles: University of Southern California.

119. Matisoff, J.A. (1991). Sino-Tibetan linguistics: present state and future prospects. *Annual Review of Anthropology.* 20, 469-504.

120. Matisoff, J.A. (1999). In defense of Kamarupan. *Linguistics of the Tibeto-Burman Area.* 22.2, 173-82.

121. Matisoff, J.A. (2000). On 'Sino-Bodic' and other symptoms of neosubgroupitis. *Bulletin of the School of Oriental and African Studies (London).* 63.3, 356-69.

122. Mazaudon, Mertin. (1985). Proto-Tibeto-Burman as a two-tone language? Some evidence from Proto-Tamang and Proto-Karen. J. A. M. G. Thurgood, D. Bradley (eds.), *Linguistics of the Sino Tibetan Area: The State of the Art.* Canberra: Australian National University.

123. Meyase, Savio Megolhuto (2014): "Four versus five: the number of tones in Tenyidie", In *TAL-2014*, 19-22.

124. Mikuteit, S. and Reetz, H.(n.d.). Caught in the act: The timing of aspiration and voicing in East Bengali.

125. Moore, C.B., and Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese tones. *Journal of the Acoustical Society of America,* 102, 1864-1877.

126. Morén, B. & E. Zsiga.(2006). *The lexical and post-lexical phonology of Thai tones.* NLLT 24: 113178.

127. Newman, S. (1944). *Yokuts language of California.* New York: Viking Fund Publications in Anthropology.

128. Newman, S. (1947). Bella Coola phonology. *International Journal of American Linguistics* 13, 129-34.

129. N Chasaide, A. & Gobl, C. (1995). Voice source variation. In W. Hardcastle and J. Laver (eds.), The handbook of phonetic sciences, pp. 427-61. Cambridge, MA: Blackwells.

130. Pandit, P. (1957) Nasalization, aspiration and murmur in Gujarati, *Indian Linguistics* 17, 165-72.

131. Pike, K. L. (1948). *Tone Languages.* Ann Arbor, Michigan: University of Michigan Press.

132. Odden, D. (1995). Tone: African languages. In J. Goldsmith (Ed.), *Handbook of Phonological Theory* (pp. 444-75). Oxford: Blackwell.

133. Ohala, J.J. (1978). Production of tone. In V.A. Fromkin (Ed.), *Tone: A Linguistic Survey* (pp. 5- 40). New York: Academic Press.

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

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

136. Remijsen, Bert. (2002). Lexically contrastive stress accent and lexical tone in Ma'ya. In CarlosGussenhoven & Natasha Warner (eds.), Laboratory *Phonology* 7, 585-614. Berlin: Mouton De Gruyter.

137. Remijsen, Bert. (2013). Tonal alignment is contrastive in falling contours in Dinka. *Language* 89(2), 297-327.

138. Remijsen, Bert & Vincent J. van Heuven.(2005). Stress, tone, and discourse prominence in the Curaao dialect of Papiamentu. *Phonology* 22.205-235.

139. Remijsen, Bert & Leoma Gilley. 2008. Why are three-level vowel length systems rare? Insights from Dinka (Luanyjang dialect). *Journal of Phonetics.* 36.318-344.

140. Remijsen5 Remijsen, Bert & D. Robert Ladd. (2008). The tone system of the Luanyjang dialect of Dinka. *Journal of African Languages and Linguistics.* 29.149-189.

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

142. Rose, P. J. (1991). How effective are long term mean and standard deviation as normalisation parameters for tonal fundamental frequency? *Speech Communication* 10. 229-247.

143. Sapir, Edward. (1925). Sound patterns in language. Language (1): 3751.

144. Sarmah, Priyankoo. (2004). *Some Aspects of the Tonal Phonology of Bodo.* M Phil dissertation, Hyderabad: Central Institute of English and Foreign Languages.

145. Sarmah Priyankoo. (2009). Tone Systems of Dimasa and Rabha: A Phonetic and Phonological Study. PhD dissertation, University of Florida.

146. Sarmah, Priyankoo and Caroline Wiltshire. (2010). An Acoustic Study of Dimasa Tones. in Stephen Morey and Mark Post (Eds) *NEILS vol 2.* Cambridge University Press India Pvt. Ltm.

147. Schacter, P & Victoria Fromkin.(1968). A phonology of Akan: Akuapem, Asante & Fante. Working papers in phonetics 9, University of California, Los Angeles.

148. Schiefer, L. (1986). F0 in the production and perception of breathy stops: Evidence fromHindi, *Phonetica* 43, 43-69.

149. Schiefer, L. (1989). Voiced aspirated or breathy voiced and the case for articulatory phonology, *Forschungsberichte des InstitutsfÄurPhonetik und SprachlicheKommunikation derUniversitÄatMÄunchenpp.* 257-278.

150. Schiefer, L. (1992). Trading relations in the perception of stops and their implications for a phonological theory, In Papers in Laboratory Phonology, Vol. II, pp. 296-313.

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

152. Shue, Yen-Liang, Patricia A. Keating and Chad Vicenik. (2009). VoiceSauce: A programme for voice analysis. *Journal of the Acoustical Society of America* Vol. 124(4).

153. Shue, Yen-Liang, Patricia A. Keating and Chad Vicenik. (2009). VoiceSauce: A programme for voice analysis. Retrieved on March, 2012 from http://www.seas.ucla.edu/spapl/voicesauce/ .

154. Silverman, D. (1995). *Phasing and recoverability.* University of California, Los Angeles Ph.D. dissertation. [Published by Garland 1997].

155. Silverman, D. (1997). Laryngeal complexity in Otomanguean vowels, *Phonology* 14, 235-62.

156. Silverman, D. (to appear). Pitch discrimination during breathy versus modal phonation, *Papers in Laboratory Phonology VI.* Cambridge: Cambridge University Press.

157. Silverman, D., Blankenship, B., Kirk, P., & Ladefoged, P. (1995). Phonetic structures in Jalapa Mazatec, *Anthropological Linguistics* 37, 70-88.

158. Stevens, K. (1977). Physics of laryngeal behavior and larynx modes, Phonetica 34, 264-79.

159. Stevens, K. N.,Keyser, S. J., & Kawasaki, H. (1986). Toward a phonetic and phonological theory of redundant features. In J. Perkell & D. Klatt (Eds.), *Invariance and variability in speech processes*, pp. 426-449. Hillsdale, NJ: Lawrence Erlbaum Associates .

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

161. Temsunungsang, T. and Paroma Sanyal. (2005).Truncation in Ao Verbal Morphology. Paper presented at ICOSAL-6, Hyderabad, India.

162. Thompson, L. & Thompson, M. T. (1992). *The Thompson language.* Missoula, MT: University of Montana.

163. Thongkum, T. (1988). Phonation types in Mon-Khmer languages. In O. Fujimura (Ed.), *Vocal fold physiology: voice production, mechanisms and functions*, pp. 319-334. New York: Raven Press.

164. Thurgood, Graham (2002). Vietnamese and Tonogenesis: revising the model and the analysis. Diachronica, 19, 333-363.

165. Traill, A. & Jackson, M. (1988). Speaker variation and phonation type in Tsonga nasals, *Journal of Phonetics* 16, 385-400.

166. Trubetzkoy, N. (1939). Grundzge der Phonologie. Travaux du cercle linguistique de Prague 7.

167. Vijayakrishnan, K. G. (2003). Stress and Tone in Punjabi. In Thirumalai, M.S. (Eds.) *Occasional Papaers in Linguistics,* Vol. 10.Language in India, India.

168. Watkins, J. (2000). Phonation type phenomena in Burmese Tone. Tone Symposium. Troms, Norway.

169. Wayland, Ratree, Scott Gargash and Allard Jongman. (1994). Acoustic and perceptual investigation of breathy voice. *Journal of the Acoustical Society of America* Vol. 97, pp. 33-64.

170. Wayland, Ratree, and Allard Jongman. (2002). Acoustic correlates of breathy and clear volews: The case of Khmer. *Journal of Phonetics.* Vol. 31(2), pp. 181-201.

171. Weidert, A. (1987). *Tibeto Burman Tonology.* John Benjamins.

172. Xu, C. X. and Xu, Y. (2003). Effects of Consonant Aspiration on Mandarin Tones. *Journal of the International Phonetic Association.* 33, 165-181.

173. Xu, Y. (2001). Sources of tonal variations in connected speech. *Journal of Chinese Linguistics.* 17, 1-31.

174. Yadav, R. (1984). Voicing and aspiration in Maithili: a fiberoptic and acoustic study. *Indian Linguistics* 45, 27-35.

175. Yip, M. (1989). Contour tones. *Phonology*, 6(1), 149-174.

176. Yip, M. (1991). *The Tonal Phonology of Chinese.* New York: Garland Publishing.

177. Yip, Moira.(1980). *The Tonal Phonology of Chinese.* PhD dissertation, MIT.

178. Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.