

Workshop on Tone and Intonation: Theory, Typology and Computation

ABSTRACT BOOK



January 26-27, 2012

Department of Humanities and Social Sciences
Indian Institute of Technology Guwahati

Contents

Don't Look at Me in that Tone of Voice! KALIKA BALI	2
Modelling of F_0 Contour for Bangla Readout Speech TULIKA BASU and ANAL HAQUE WARSI	2
Tone Changes, Meaning Differs: A Focus on the Noakhali Bangla Tones UTTAM CHAKRABORTY	3
Tone in the Languages of Tripura SHYAMAL DAS	4
Tone in Riang Dialect SWAPAN DEBNATH	5
An Acoustic Study of Dzongkha (Paro Dialect) Tones KINLEY DORJEE	6
Lexical Tones of Tenyidie INDRANIL DUTTA, MIMI KEVICHÜSA EUNG, SHAKUNTALA MA- HANTA, SAVIO MEGOLHUTO MEYASE	7
Tone-mapping in Mizo LALRINDIKI T FANAI	8
Sentential Tone in Sylheti Bangla ARPITA GOSWAMI	9
Epoch Based Prosody Modification for Tone and Intonation Modification S R MAHADEVA PRASANNA	10
Mizo Tone Synthesis Using Epoch Based Prosody Modification PRIYANKOO SARMAH	10
Tones in Tedim Chin T TEMSUNUNGSANG	12
Focus in the Nalbariya Variety of Assamese ASIMUL ISLAM TWAHA	12
Emergence of Tone in Loan Phonology: The Case of Mizo and Tenyidie K G VIJAYAKRISHNAN	13

**Don't Look at Me in that Tone of Voice!
or, why Text-to-Speech Synthesisers sound the way they do.**

KALIKA BALI^a

^aMicrosoft Research India
Bangalore, India
kalikab@microsoft.com

A text-to-speech synthesis system aims to generate appropriate prosody from and for a given text. While text in MOST languages of the world does represent the spoken language at some level, it usually pays minimal attention to the prosodic structure.

Prosody provides vital information at different linguistic and paralinguistic levels that is crucial for communicating in the spoken mode. However, given that we can and do communicate effectively in the written mode, why is prosody essential for TTS? In this talk, we will look at the fundamental hurdles in generating prosody from text and the reason why researchers and developers look at work-arounds when it comes to synthesising the prosodic component of speech. The talk will highlight the inherent dissonance between textual and spoken form, which is further accentuated when it comes to prosody. We will also consider the distance between theory and practice in prosody generation for TTS the current techniques in use.

Modelling of F_0 Contour for Bangla Readout Speech

TULIKA BASU and ANAL HAQUE WARSI^a

^aC-DAC
Kolkata, India
tulika.basu@cdackolkata.in, anal.warsi@cdac.in

In India, speech synthesis is considered to be of primary need not only to empower disabled people, but also the functionally illiterate population. It is well known that the F_0 contour plays an important role in bestowing both intelligibility and naturalness in synthesized speech but the process of synthesizing the F_0 contour from the underlying linguistic information has not yet been elucidated for Bangla. So an attempt was made to analyze the contours of Bangla readout speech using the command-response model by Fujisaki et al., and define the prosodic units of Bangla. It has been observed that in Bangla tone is not phonemically significant. However intonation pattern of an utterance plays an important role in Bangla. The meaning of a Bangla utterance is changed with the change of intonation pattern. Based on the observations a model for generating the F_0 contours of Bangla readout speech was implemented and integrated with the Bangla TTS system developed by C-DAC, Kolkata.

Tone Changes, Meaning Differs: A Focus on the Noakhali Bangla Tones

UTTAM CHAKRABORTY^a

^aTripura University
Suryamaninagar, India
dhanitattauc@rediffmail.com

Tone as an intrinsic property of a morpheme, word, or a grammatical construction has a lexical or grammatical significance in languages. Especially when the homonyms or homophonous words are concerned, there the usages of different tonal properties are undeniable to convey different respective meanings. Even a synonym also is capable of expressing different condition or concept of the content while sharing different tones. A tone language can identify two or more distinctive types of tones as 'high', 'mid', 'low' and 'level', 'falling', 'rising' etc. The present paper aims at finding out the uses of tones and their contribution in determining the meanings of words or expression in Noakhali Bangla, a dialect of Bangla spoken in the south Tripura district and in some regions of Bangladesh. In Noakhali Bangla, examples are many where the use of different tones can introduce different or non-connective meanings from the same input expression. Below, I am representing few examples in Table 1 and Table 2 supporting my preliminary findings. The representative examples make the fact clear that sometimes a single word is

Table 1: [zɔl] 'water'









Tone	Meaning
	"water"
	"only water and water"
	"Is water?"
	"to give water"

Table 2: [at]

Tone	Meaning
	"bazaar"
	"eight"
	"to walk (imperative)"
	"Is eight?"

also used to convey a comprehensive syntactic meaning in Noakhali Bangla. The full paper will include some more representation of such variable tones for different meanings from the identical words in the dialect. It is however, a humble initiative to study the language variety from the tonal perspective.

Tone in the Languages of Tripura

SHYAMAL DAS^a

^aTripura University
Suryamaninagar, India
das_shyam232@rediffmail.com

Languages in Tripura belong to three broad families Sino-Tibetan, Indo-Aryan and Austric. Of these the Sino-Tibetan or Tibeto-Burman is the most diverse represented by Kokborok, Manipuri, Lushai, Garo, Mog etc.etc. Kokborok, the official language of tribal people of the state has again many dialects: Tripuri, Reang, Jamatia, Noatia, Ucoi, Murashing, Rupini and Kalai. Indo-Aryan family has various eastern dialects of Bangla as its representatives in Tripura such as the Noakhali, Kumilla, Chattagrami and Sylhetti dialects. Mundari is the sole representative of the Austric family.

Use of sentential tone or intonation is common amongst all the languages and dialects. But use of lexical tone is noted only in few. Amongst the languages of the Tribal folks Reang, Ucoi, and Kokborok in general are found to be employing tone for lexical distinction in a systematic manner. There are mainly two tones high and low in these varieties. For instance, in Reang dialect *ktoi* with low tone (LT) means younger in age and *ktoi* with high tone (HT) means sweet; *kra* (LT) to make fun, *kra* (HT) strong/solid etc. Similarly in Ucoi *ri* with a falling tone means to give, and *ri* with a rising tone signifies to draw a line etc. Acharjee (2009) notes the presence of at least two tones level and high in Kokborok. For instance *ca* with level tone is to be appropriate and the same with a high tone means to eat; *su* with level tone means to measure/ to repay and with high tone denotes to wash/to prick etc.

In the various dialects of Bangla also we notice the use of two tones high and low/level to contradistinguish otherwise homophones. For example *sa* (LT) tea and *sa* (HT) ask for/see; *asar* (LT) pickle and *asar* (HT) stumbling in the Kumilla dialect. In the Noakhali variety we notice use of tone to distinguish between pairs like *haDa* (LT) grinding stone and *haDa* (HT) to rind/burst; *kara* (LT) to husk paddy etc. and *kara* (HT) to snatch etc. In the Chattagram variety we come across the use of similar contrastive use of low and high tones: *sini* (LT) I know and *sini* (HT) sugar etc. Again in Sylheti dialect we notice the contrastive use of high and low tones: e.g. *kOTa* (LT) squirrel vs. *kOTa* (HT) salty; *Xala* (LT) black vs. *Xala* (HT) deaf etc.

In the present paper no claim to any new discovery is made. On the contrary efforts have been made to put together as much information as available in various works including the authors own on the tonal phonology in the languages and dialects spoken in Tripura.

Tone in Riang Dialect

SWAPAN DEBNATH^a

^aRamakrishna Mission
Viveknagar, India
raj_dn29@yahoo.com

Tripura is small and hilly state in the north-east corner of India. Its population is heterogeneous. Variety of ethnic elements can be traced in it. The Bengalees and the Tripuris are the two major communities. The greater tripuri communities include a number of tribes such as Tipra, Riang, Jamatia, Noatia, Uchai, Halam. Kuki, Lusai, Garo, Mog, Chakma, Murasing etc. The Riangs whose language particularly tonology is the subject matter of this abstract call themselves *Bru*. The Riangs are considered to be the most primitive amongst the tribal communities in Tripura. Their language is called *Kau Bru* i.e. the language of the Bru. This Kau Bru does not have any written literature or alphabets. This dialect exists in spoken form only and in the verge of extinction. However, the empirical study finds out the fact that this dialect allows lexical tone i.e. the pitch of a word in this language can change the meaning of a word, thus making it a tone language. The pitch not only changes the nuances of the language but also changes the core meaning of the language. Two tones are observed in the initial observation: High Tone and Low Tone¹. Some examples obviate the point under study.

^L ktɔi	'younger in age'	ADJ	LT
^H ktɔi	'sweet'	ADJ	HT
^L ktɔi.t ^h ag	'youngest'	ADJ	LT
^H ktɔi.t ^h ag	'sweetest'	ADJ	HT
^L kra	'fun'	N	LT
^H kra	'durable'	ADJ	HT
^L klai	'easy'	ADJ	LT
^H klai	'slip down'	VERB	HT

The language under study is very rich from linguistic point of view. The study and analysis of the tonology of this dialect may lead to the tone-phenomena of the other tribal languages/dialects in the state and can save the dialect from being extinction.

¹LT: Low Tone and HT: High Tone

An Acoustic Study of Dzongkha (Paro Dialect) Tones

KINLEY DORJEE^a

^aParo College of Education, Royal University of Bhutan
Paro, Bhutan
kinley_dorjee@pce.edu.bt

This paper presents results of an acoustic study of tones of Paro dialect of Dzongkha (the national language of Bhutan). In earlier literature, Thimphu and Chapcha dialects (Mazaudon and Michailovsky 1988) and Pasakha dialect (Watters 2002) were analyzed as four tone (divided into two registers) languages, namely: high-level, high-falling, low-level, low-falling. However, tones in Paro dialect have never been studied.

Acoustic recordings of four speakers of the Paro dialect in their 20s were conducted. Using Praat software, three tokens of each recording (in isolation) were analyzed. The results suggest that all the four tones are dynamic tones; high register tones fall and low register tones rise. The pitch pattern within the Register can best be analyzed with pitch height differences rather than level falling contour types.

In Paro dialect of Dzongkha, the tonal range can be divided into two pitch Registers, correlating with initial phonation types, i.e., 'Tense initials' (correlating with High) and 'Lax initials' (correlating with Low) phonation types in obstruents, while voiced sonorant onsets allow contrastive tones, with a default Mid tone at the right edge. Dzongkha is then an incipient, lexically bitonal language, supported by the lexical tone specification on the syllables with voiced sonorant onsets.

Lexical Tones of Tenyidie: Fundamental Frequency, Voice Quality and Pitch Modeling

INDRANIL DUTTA^a, MIMI KEVICHÜSA EUNG^b,
SHAKUNTALA MAHANTA^c, SAVIO MEGOLHUTO MEYASE^a

^aThe English and Foreign Languages University
Hyderabad, India
indranil.dutta.id@gmail.com, saviomm@gmail.com

^bUniversity of Hyderabad
Hyderabad, India
ezungm@gmail.com

^cIndian Institute of Technology, Guwahati
Guwahati, India
smahanta@iitg.ernet.in

In this study², we report on the status of lexical tones in Tenyidie. We present data from six female speakers of Tenyidie from Kohima between the age of 22 and 24. The target words for this study included the following four words in two sentential contexts; a frame sentence and an example sentence where the prosodic and segmental context was not controlled: /pɛ/, /ba/, and /kɛla/.

While in the phonology, there may be grounds for positing five tones in Tenyidie [2, 3], we argue that Tenyidie exhibits a four-tone lexical contrast, namely, Extra-High (EH), High (H), Mid (M) and Low (L). This result is consistent with the findings of a previous study [1], where mean F_0 on the tone bearing syllable was found to be significantly different between the four tones. In contrast, we present results from time-normalized F_0 at ten intervals on the syllable which better models the acoustic manifestation of lexical tone in Tenyidie.

It has also been observed that the vowels carrying the L tone in Tenyidie exhibit a breathy voice quality [1], however, in [1] this property of the vowels carrying the L tone was not measured. In addition to the time-normalized F_0 , in this study we also report on the voice quality characteristics of the vowels carrying the L tone based on three measures of voice quality, namely, the difference in the amplitude of the first and second harmonic (H1-H2), first formant bandwidth (H1-A1) and overall spectral tilt (H1-A3).

We conclude by suggesting that a more comprehensive study of Tenyidie tones needs to be conducted where in addition to the lexical status of tones in Tenyidie, consonant-tone interaction or consonant- F_0 perturbation, tone-intonation interaction and tone sandhi needs to be examined. This study will draw parallels with studies such as [4] and also comment on the need to build predictive models of tone production and perception in Tenyidie.

²The names of the contributors appear in alphabetical order by last name. All the authors contributed equally towards this study.

References

1. Blankenship, B., Ladefoged, P., Bhaskararao, P. and Nichumeno, C. 1992. Phonetic structures of Khonoma Angami. *Linguistics of the Tibeto-Burman Area*, vol. 16, no. 2, pp. 69-88
2. Burling, R. 1960. *Angami Naga phonemics and word list*. *Indian Linguistics*, vol. 21, pp. 51-60
3. Ravindran, N, 1974. *Angami Phonetic Reader*. CIIL Phonetic Reader Series 10. Mysore: Central Institute of Indian Languages.
4. Sarmah, P. and Wiltshire, C. 2010. A Preliminary Acoustic Study of Mizo Vowels and Tones. *Journal of Acoustical Society of India*. Vol. 37, No. 3, pp. 121-129.

Tone-mapping in Mizo

LALRINDIKI T FANAI^a

^aMizoram University
Aizawl, India
drdikifanai@gmail.com

The paper deals with tone-mapping in Mizo. There is a considerable evidence for assuming that the parameters for tone mapping/ linking is from right-to-left in Mizo rather than the expected left-to-right linking for other tone languages (specifically African tone languages). This hypothesis for right-to-left tone mapping in Mizo is assumed to hold true if a tone language has the following features:





- i. contrastive vowel length
- ii. sonorants as tone-bearing units
- iii. the presence of contour tones
- iv. tone change in derived environment

Sentential Tone in Sylheti Bangla

ARPITA GOSWAMI^a

^aTripura University
 Suryamaninagar, India
goswamiarpita2010shb@gmail.com

Intonation is a means for organizing languages into patterns that fit the communicative needs. It plays a vital role as far communication is concerned. Intonational choices made by speakers carry linguistic information and perform a variety of functions while conveying particular meaning by particular utterances. Intonation however, is a largely unconscious mechanism and a complex aspect of speech production. Functionally, intonation facilitates the hearer to work out the particular meaning from the context, whether the particular utterance is a statement or a question, a command or a request etc. This paper demonstrates the tones in Sylheti bangla (henceforth SHB), spoken in the three states of India, as Tripura, Assam and Maghalaya. It is also well practiced in the Sylhet district of Bangladesh. SHB has a rich inventory of sentential tones, which differentiates this variety from the other language varieties of Bangla. In this paper, I intend to explore the significant intonational features of SHB with a special reference to Standard Colloquial Bangla (henceforth SCB). This paper is based on isolated speech of the SHB speakers while comparing it with that the SCB sentential tone. Below, I am representing few comparative data in support of my issue.

	Speech	Tone	Remark	Gloss
(1)				
SCB	ami baꠞite jabo		mid-rise	"I will go home"
SHB	ami baꠞiṭ̣̣ zaimu		mid-fall-rise	"I will go home"
(2)				
SCB	ṭ̣̣ui koꠞ ^h ai jacc ^h ij		low-rise-fall	"Where will you go?"
SHB	ṭ̣̣ui xoi zaray		low-rise-fall-rise	"Where will you go?"

However, the examples display differences as well as similarities between SCB and SHB sentential tones. The tone contours may vary depending on the focus on the particular word of the speech. They may vary depending on the particular expression, whether it is a statement or it is a question. The full paper in length will include some of such important issues. Additionally, the paper will also include the impact of SHB intonation on English, while English is spoken by the native speakers of SHB as their target language (L2). Last but not least, it is an initiative to study the dialect from the intonational stand points. It is however, a little part of the ongoing research undertaken by me. More informative data and explanation will be with the full paper.

Epoch Based Prosody Modification for Tone and Intonation Modification

S R MAHADEVA PRASANNA^a

^aIndian Institute of Technology Guwahati
Guwahati, India
prasanna@iitg.ernet.in

The objective of this talk is to present the epoch prosody modification approach and to demonstrate its usefulness for tone and intonation modification. Prosody modification refers to the process of modifying the non-textual information present in speech. Humans can perform prosody modification in an effortless manner. However, to mimic the same on a machine seems difficult. The main limitation of prosody modification methods is the distortion that they introduce while performing modification. One of the major reasons attributed for this distortion is the poor localization of pitch markers. The epoch based methods are based on accurate pitch markers using epochs. Hence they tend to reduce the distortion significantly. Also the epochs provide convenient way for prosody modification. The talk will present the recent developments in epoch extraction and epoch based prosody modification methods. The significance of the epoch based prosody modification is illustrated in tasks like pitch, duration, intensity, tone and stress modification.

Mizo Tone Synthesis Using Epoch Based Prosody Modification

PRIYANKOO SARMAH^a

^aIndian Institute of Technology Guwahati
Guwahati, India
priyankoo@iitg.ernet.in

Accurate synthesis of lexical tones is crucial for the success of text-to-speech (TTS) systems for tone languages. At the same time, correct identification of lexical tones is also important for Automatic Speech Recognition (ASR) systems. For example, Huang and Zhu (2007) have reported that in case of Mandarin Chinese Automatic Speech Recognition (ASR) systems, 48.3% of the errors can be attributed to the misrecognition of the tones. Hence, for both TTS and ASR systems for tone languages, identification of cues associated with lexical tones is crucial for their effectiveness. In this work, we identify the acoustic cues associated with Mizo tones and synthesize its tones from those acoustic cues using the epoch based prosody modification method. Subsequently, we present the synthesized tones to the native speakers of Mizo to assess the goodness of the synthesized tones.

Mizo, is a tone language spoken by about 674,756 speakers in the province of Mizoram in North-East India. Mizo is reported to have four contrastive tones, namely high, rising, falling and low (Fanai, 1992). We identified that the primary acoustic cue associated with Mizo tones is the slope of the fundamental frequency (F_0) contour and came up with several temporal and acoustic cues associated with them.

Tone	Average F_0	F_d (F_0 offset - F_0 onset)	Duration
High	251.71 Hz	0.72	227 ms
Rising	214.54 Hz	11.57	223 ms
Falling	236.80 Hz	-22.51	176 ms
Low	210.81 Hz	-9.03	183 ms

With the help of the extracted values, we synthesized some Mizo tones on the [pa:] syllable recorded from a Mizo speaker pronounced in a low tone. While synthesizing the tones, an epoch based prosody modification method was used where; the epochs or the points of glottal closure are automatically estimated. This method modifies the epoch with the F_0 values of the target tone and finally reconstructs the waveform to generate a prosodically modified speech. Apart from modifying the F_0 , we also modified the duration of the synthesized tones.

After we synthesized the new Mizo tones, we asked native speakers of Mizo to identify and rate the goodness of the synthesized tones. Synthesized rising tones that had F_0 , F_d and duration modified, had a correct identification rate of 95%. At the same time native speakers rated their goodness as 4.65 out of 5.

The findings of the current work reveal that (a) in order to build an efficient TTS system for tone languages several acoustic and temporal cues need to be incorporated in the synthesized speech and (b) the epoch based prosody modification method is effective in synthesizing speech data for tone languages.

References

1. Huang, Hao and Jie Zhu. 2007. Discriminative tonal feature extraction method in mandarin speech recognition. *Journal of China University of Posts and Telecommunication*. 14 (4).
2. Fanai, Lalrindikii. 1992. *Some Aspects of the Lexical Phonology of Mizo and English: An Autosegmental Approach*, PhD dissertation, CIEFL, Hyderabad, India.

Tones in Tedim Chin

T TEMSUNUNGSANG^a

^aThe English and Foreign Languages University
Shillong, India
ttjamir@gmail.com

In this paper, we present a preliminary analysis of tones in Tedim Chin, a northern Kuki-Chin language belonging to the Tibeto-Burman family. It is spoken in the North-Eastern states of Manipur and Mizoram; and mostly in Chin state of Myanmar.

Henderson (1965) identifies three contrastive tones in Tedim Chin: Rising, Falling and Level in CV:C, CVV and CVC_{cont} syllable shapes. In addition, CVC_{stop} syllables always have a low level or low falling pitch. An exception is noted for *thak* 'to be new', which has a high tone. Hence, Tedim Chin attests upto five tones, of which the low and high tones are allotones of the falling and rising tones respectively. Thang's (2001) analysis is consistent with that of Henderson's.

We re-examine the status of the tones in Tedim Chin by looking at monosyllabic words, postulating three contrastive tones: High, Mid and Low. Consequently, we hypothesize the mora to be the Tone Bearing Unit. We will show that the contour tones (postulated in earlier works) are a result of phonetic implementation rather than phonological.

Focus in the Nalbariya Variety of Assamese

ASIMUL ISLAM TWAHA^a

^aIndian Institute of Technology Guwahati
Guwahati, India
asim4all@gmail.com

In this paper, we investigate the phonological and phonetic markers of focus in the Nalbariya variety of Assamese. We have found that focus is marked by phrasing, type of pitch accent and boundary tone. In order to investigate the relationship between word order and prosody, we conducted an experiment on native speakers of Assamese speaking the Nalbariya variety of Assamese. For the purpose of analysis we took 10 declarative sentences controlled for the number of syllables. Each sentence was produced in six different types: SOV (wide focus), SOV (subject focus), SOV (object focus) OSV (wide focus), OSV (object focus), OSV (subject focus). All these sentences were produced in response to a scripted question and answer pattern written in the Assamese orthography. Ten speakers were recorded for this experiment. We found that as far as the focus is concerned, in

SOV construction it is marked exclusively by pitch range; when the first constituent (i.e. S) is focused, we see a wider pitch span on the Subject. However, the second constituent does not show statistically significant difference from the wide focus baseline. Again in OSV construction, when the first constituent is focused, the pitch span is not wide enough to be greater than the wide focus base line; the reason here is the higher F_0 Minimum value of the first constituent. It reduces the pitch span of the focused item; otherwise the F_0 peak on the first constituent in OSV pattern is higher than that in SOV pattern. Similar to the SOV order, the focused second constituent does not exhibit statistically significant difference in pitch range from the wide focus pitch range on the second constituent. The focus on the second constituent in both SOV and OSV order is not statistically different from the wide focus baseline; the reason behind this may be the fact that the immediately preverbal position is the default focus position.

In both the word order (SOV and OSV), there is a global compression of the pitch contour of focused utterances. The only exception is in SOV word order, where there is no post focal compression, otherwise focal, pre-focal and post focal compression mark the Nalbaria variety of Assamese. Again whether focused or not, the initial constituent shows higher F_0 -Maximum peak in OSV order than in SOV order; while the second constituent in OSV undergoes greater compression than it is in SOV.

References

1. Khan, Sameer. 2007. Phrasing and Focus in Bengali. Poster presented at the satellite meeting of the 16th ICPhS, Saarbruecken, Germany, August 5, 2007.
2. Patil et. al. 2008. Focus, Word Order and Intonation in Hindi. *JASA*, 1(1):55-72

Emergence of Tone in Loan Phonology: The Case of Mizo and Tenyidie

K G VIJAYAKRISHNAN^a

^aThe English and Foreign Languages University
Hyderabad, India
vijayakrishnankg.1952@gmail.com

Mizo and Tenyidie are both tone languages but the nature of the realization of tone in the two languages is fairly distinct. Perceptually, Mizo tones are very salient but tones in Tenyidie are more subtle. Mizo has only two tones, namely High and Low, but it allows the two tones to link multiply to a syllable giving rise to compound tones (Rising and Falling) in the language. This phenomenon is not attested in Tenyidie which has four distinct tones with severe restrictions

on tone combinations in polysyllabic sequences. The working hypothesis is that due to greater tonal salience along with morphological alternation in Mizo, loan words may attest unpredictable tones, becoming totally phonologized. But even then compound tones may not be attested normally in loan words allowing us to generalize that only the less marked tonal linking is carried over from the language into the phonology of loans. In Tenyidie, on the other hand, tonal specification stays closer to the phonetic end, being more predictably associated with different segment types in the loan words.