International Journal of
Emerging Research in Management &Technology
ISSN: 2278-9359 (Volume-4, Issue-6)

Research Article

June
2015

# A Study on the Acoustics Properties of Speech with Respect to Assamese Language

**Mridula Medhi*, Parismita Sarma**
Department of Information Technology, Gauhati University
Assam, India

***Abstract*—**

***As devices become more complex, interaction between humans and computers also becomes more demanding and sets new requirements for the user interfaces specially those with speech synthesis. However, depending on the tone or in general intonational notation the same sentence can be storing different meaning depending on what pitch, energy etc has been used. This paper shows a study on this important factor of speech synthesis.***

***Keywords*— Intonation, pitch, linguistics, acoustics, TTS.**

## I. INTRODUCTION

In speech the variation in the pitch in terms of words spoken, intonation is used for a number of functions like indicating the attitudes and emotions of the speaker, signaling the difference between statements and questions and between different types of questions, focusing attention on important elements of the spoken part. It contrasts with tone, all languages in speech use pitch pragmatically in intonation — such as while making a question or a surprise statement, while raising anger or just a modest request etc. Hence in various tonal languages like (for example Assamese) for distinguishing different words in addition to using pitch, intonation is used. It serves different functions in terms of linguistics and para linguistics, which starts from marking of sentence modality and concludes till the expression of emotional and attitudinal nuances[1]. Intonation analysis involves checking to distinguish whether there is stress or accent, and, if it is an accent, what kind of pitch accent it is. It also includes condition checking to see if a boundary is present or not, and if it is so which pitch movement or level is used to mark it. There are also many gradient aspects like variation in the height of pitch or in the exact shape of the contour.

## II. INTONATIONAL SYSTEM

In this paper the main aim is to observe the variation in the functional and emotional intonation with reference to the assamese language. In speech recognition as well as in TTS systems it plays an important role for the correctness and naturalness of the speaking model. In the earlier days TTS (Text to Speech) systems typically had just a single "voice". But in recent time much attention has been given to the notion of having large numbers of voices in synthesis systems. A logistic requirement of this is that the speech on which these voices are modelled should be acquired quickly which implies automatic transcription techniques for all components including intonation. Hence we need some way to automatically analyse and parameterise data so that the intonational characteristics of a speaker can be captured. The main motive is providing a system of intonational description which on the other hand is also meaningful in terms of linguistics so that the representations can be shown based on the parts of an utterance's acoustics. It means the representations should contain information which plays a vital role in the linguistic information of an utterance's intonation [2]. An intonation system should be able to express the difference in utterances and represent these distinctions.

In this paper the main aim is to give a very basic introduction to the research for developing an intonational model in assamese language.

## III. DESIGN AND CONFIGURATION ISSUES

In this paper the main aim is to observe the variation in the functional and emotional intonation with reference to the assamese language.

### *A. Terminologies that needs to be incorporated*

➢ *Coarticulation*

Coupling effect, when two sounds are produced together

Production of /k/ and /a/ in isolation is different from producing "ক " /ka/

➢ *Energy*

Suitable energy contour

➢ *Pitch*

Pitch and its contour

(variation across the phones)

- *Duration*

How long each phone should be [4].

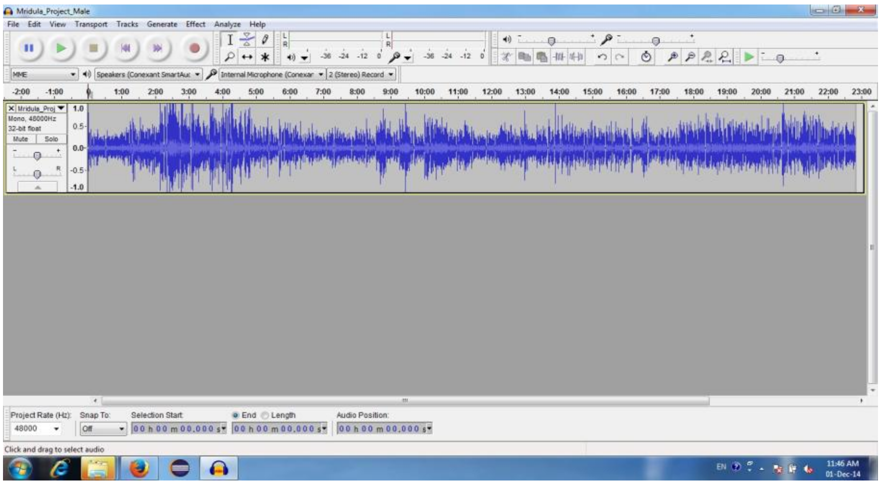
### *B. Data Collections*

The input is *text* which is based on Raw text, Formatted text (MS Word, PDF/PS, MS PPT), and Encoded text. Conversion from different formats (pdf/ps/doc) to a generic tagged format or raw text. Data is obtained from various sources like novels, newspapers etc. based on Assamese language and from the native speakers of the respective language and was syllabified by both male and female speakers. Approximately 1500 Assamese lines were re-written in Unicode of about two to four combination of words forming different sentences of varied emotions. In this model we are considering three types of sentences based on varied emotions i.e. normal, surprise and angry moods. The entire recording of the collected text was done by one male speaker and one female speaker respectively with their mother tongue Assamese. The recording was done in a quiet room with a noise cancelling microphone using the recording facilities of a typical multimedia computer system. The voice recorded was processed on Audacity. The recording was done on 1500(approx.) distinct lines once for the male speaker and next for the female speaker. The recorded speech is stored in as a (.wav) file separately for male and female respectively.

### *C. Methodology*

According to the different emotions the extracted files are numbered serially with distinctive markings for both the voices i.e. male and female along with the markings of the emotion used for that sentence. An instance for this kind of file is shown below-

- textm_a_1.wav(male speaker with angry emotion)
- textf_n_1.wav(female speaker with normal emotion)

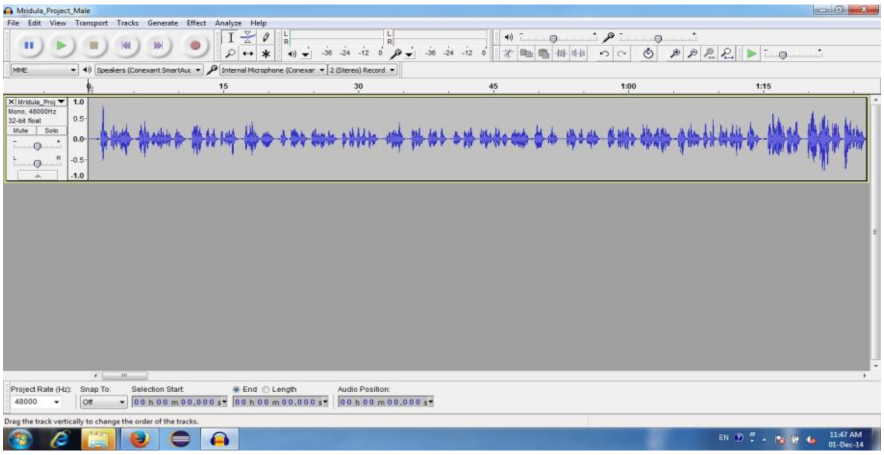


Mridula_Project_Male

Project Rate (Hz): 48000
Snap To: Off
Selection Start: 00 h 00 m 00.000 s
End: 00 h 00 m 00.000 s
Audio Position: 00 h 00 m 00.000 s

Click and drag to select audio

Fig. 1  wav file for the recorded voice of the female speaker



Mridula_Project_Male

Project Rate (Hz): 48000
Snap To: Off
Selection Start: 00 h 00 m 00.000 s
End: 00 h 00 m 00.000 s
Audio Position: 00 h 00 m 00.000 s

Drag the track vertically to change the order of the tracks.

Fig. 2 wav file for the recorded voice of the male speaker

The model is represented in three parts using histogram plots with the first showing the energy consumed while speaking the Assamese lines with their different emotions. The plot with the colour black represents normal tone, while the colors in green and yellow represents anger and surprise respectively. The second part in the histogram represents pitch in the different emotion in the sentences where the colour blue represents surprise in its tone while the colour magenta and black represents normal and anger in its tone respectively.

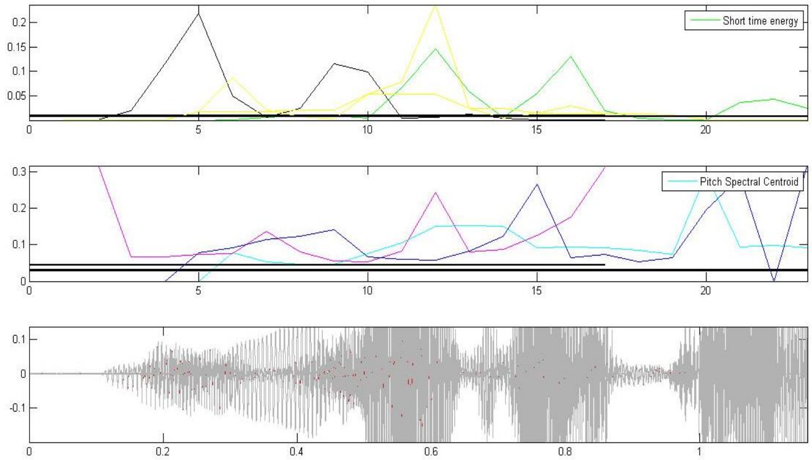


Fig. 3 histogram representation of energy, pitch and the recorded text of a male speaker with anger, normal and surprise tone.

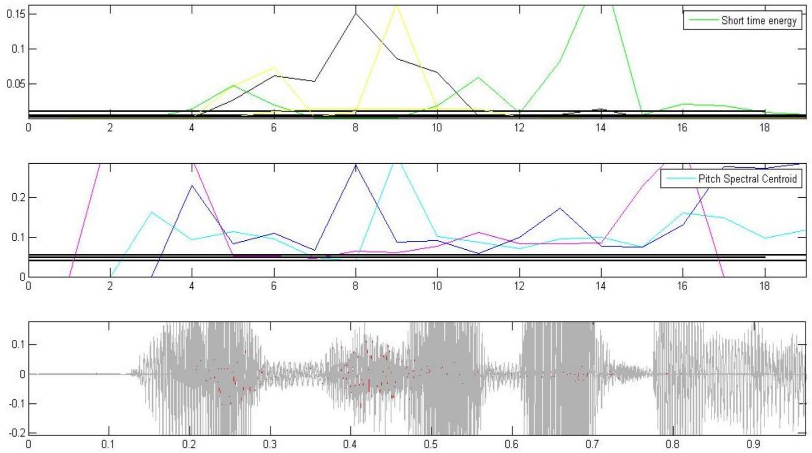


Fig. 4 histogram representation of energy, pitch and the recorded text of a female speaker with anger, normal and surprise tone.

The third part in the model figure is the histogram representation of the recorded speech.

## IV. EXPERIMENTAL ANALYSIS

In this paper we have analysed is how the same sentences can be expressed in different tones to convey different meanings. We have considered normal tone, anger and surprise tones e.g. the sentence "কত যাবা ?" Kot jaaba? can be expressed in all of the above three tones and even if it is the same sentence it will still convey different meanings. If conveyed with a higher pitch and varied energy it will be sounded as a surprise or anger in the emotion compared to a normal voice. And according to the research performed we observed that the emotional difference in the voice is measured by fundamental frequency, duration, rhythm and different aspects of voice quality. While the research was performed we divided the male and female wav files into distinctively 3000 wav files. The same text was given to both the speakers but the emotional tone e.g. The tone in Anger, Normal and Surprise etc. were comparatively different in both the cases. In terms of energy between the male and female voice a huge difference can be observed especially in case of anger . Also in respect to its own sound such as the male voice shows a significant change in its energy while speaking the same sentence in all the three modes – anger, surprise, normal. Considering the pitch we find a huge difference in anger and surprise tone in comparison to a normal tone.

## V. CONCLUSION

In general we know there are various sentences in Assamese which can be expressed in different tones with different expressions. The acoustic parameters of speech plays a vital role in the differences between the same sentences uttered in a different tone where energy, pitch, rate are significant. In this paper we have considered energy and pitch. As assumed we have found differences in the male and female recorded wav files. The different pitch values of the speakers of different sentences are also observed which is helpful for the study performed.

## ACKNOWLEDGMENT

The authors would like to acknowledge the Gauhati University Instrumentation Department for allowing us to do the recording and also the GUIST-Information Technology Department for giving us the opportunity to carry with our work in the NLP lab.

## REFERENCES

[1] Martine Grice1 & Stefan Baumann, "*An Introduction To Intonation – Functions And Models*", IfL – Phonetik, Universität Köln.

[2] Paul Taylor, "*Analysis and Synthesis of Intonation using the Tilt Model*", Centre for Speech Technology Research, University of Edinburgh.

[3] Peri bhaskararao, "*Salient phonetic features of Indian languages in speech technology*", Tokyo University of Foreign Studies, Tokyo, Japan.

[4] Bimal Kumar Kalita, Laba kr. Thakuria, Barnali Kalita, Purnendu Acharjee, P.H.Talukdar , "*A Model & Design Of A Database Of Functional And Emotional Intonation With Reference To Assamese Lanaguage*", Department of Instrumentation and USIC, Gauhati University, India.

[5] Dmitry Sityaev, Tina Burrows, Peter Jackson, Katherine Knill, "*Analysis and Modelling of Question Intonation in American Englis*h", Dmitry Sityaev, Tina Burrows, Peter Jackson, Katherine Knill Speech Technology Group, Toshiba Research Europe Ltd. Cambridge Research Laboratory, 1 Guildhall Street, Cambridge CB2 3NH, UK.

[6] SUN-AH JUN AND CÉCILE FOUGERON, "*A Phonological Model of French Intonation*".

[7] Bora, Mahendra (1981*). The Evolution of Assamese Script*. Jorhat, Assam: Assam Sahitya Sabha, Neog, Maheshwar (1980). *Early History of the Vaishnava Faith and Movement in Assam*. Delhi: Motilal Banarasidass. *"Assamese literature – An overview and historical perspective Linking into broader Indian canvas*". Retrieved 2012-01-04, "*Assamese writing System*". Archived from the original on 11 December 2007. Retrieved 2007-12-17, "*Antiques reveal script link – Inscriptions on 3 copper plates open new line of research*". The Telegraph (Kolkata). 25 January 2006. Retrieved 2007-12-17.

[8] Dmitry Sityaev, Tina Burrows, Peter Jackson, Katherine Knill, *Analysis and Modelling of Question Intonation in American English.*

[9] Beskow J. (1996). Talking Heads - Communication, Articulation and animation.Proceedings of Fonetik-96: 53-56.

[10] Dutoit T, "*High-quality text-to-speech synthesis: an overview. JInal of Electrical & Electronics Engineering*," Australia: Special Issue on Speech Recognition and Synthesis, vol. 17, pp 25-37 .

**[**11**]** Dmitry Sityaev, Tina Burrows, Peter Jackson, Katherine Knill.," *Annotating Speech Corpus for Prosody Modeling in Indian Language Text to Speech Systems".*

[12] Chandan Sarma, Prof. P.H Talukdar*," Dialect variation in Boro Language and Grapheme-to-Phoneme conversion rules to handle lexical lookup fails in Boro TTS System".*

[13] Time- and Text-Aligned Annotations: the SpLaSH Data Model

[14] Ivana Kruijff-Korbayov´a and Geert-Jan M. Kruijff, " *Discourse-Level Annotation for Investigating Information Structure".*

[15] Aljoscha Burchadt, Katrin Erk, Anette Frank, Andrea Kowalski, and Sebastian Pado,"*A Versatile Multi-Level Annotation Tool".*

[16] Flanagan J. (1972). *"Speech Analysis, Synthesis, and Perception "* .Springer-Verlag,Berlin-Heidelberg-New York.

[17] Klatt D, "*Review of text-to-speech conversion for English*", JInal of the Acoustical Society of America, vol. 82, pp 737-93 .