

Zero Crossing Rate Analysis of Assamese Vowel Phonemes

Bhargab Medhi*

Department of Instrumentation & USIC,
Gauhati University, Assam, India

Prof. P. H. Talukdar

Department of Instrumentation & USIC,
Gauhati University, Assam, India

Abstract— Speaker recognition is the identification of the person who is speaking by the characteristics of their voices. Assamese is a Indo-Aryan family of languages, mainly spoken in the North-Eastern of India. In this paper text dependent speaker modelling technique is used. The system contains training phase, the testing phase and the recognition phase. The database consists of utterance of 10 speakers with equal number of male and female speaker. Each phoneme is repeated 10 times by each speaker. The feature Zero Crossing Rate (ZCR) is used for the acoustic measure which can be helpful to design an Assamese speaker recognition system.

Keywords—Speech recognition, Feature Extraction, Zero Crossing Rate, Frame, Vowel Phonemes.

I. INTRODUCTION

Speech is natural mode of communication for people in our lives. It is very difficult to design Speaker recognition System in 100% accuracy. Speaker recognition can be classified into speaker identification and speaker verification. Speaker identification is the process of determining which registered speaker provides a given utterance [4]. Speaker recognition methods can be divided into two categories: text independent and text dependent. In a text independent system, speaker models capture characteristics of somebody's speech which show up irrespective of what one is saying. In a text dependent system, the recognition of the speaker's identity is based on the speaker's speaking one or more specific phrases or words [6].

Speaker recognition system contains two main modules: feature extraction and feature matching. In speaker recognition system the goal is to identify the speaker irrespective of what is being said, in speech recognition the goal is to recognize what is being said irrespective of who is speaking [7]. Automatic speaker recognition technology is becoming increasingly widespread in many applications. Speaker recognition is an example of biometric personal identification. The area can be security applications such as physical access control, computer data access control and so on.

The Assamese (IPA: ɔxɔmija) is a major language in the north-eastern India whose origin root is Indo-European family of languages. There are thirty two essential phonemes

in Assamese language out of which eight are vowel phonemes and twenty four are consonant phonemes [8]. Assamese scripts, derived from the Devanagari scripts consists of thirty nine consonant and eleven vowel symbols which are arranged in a well structured scientific manner based on phonetic principles [1]. Vowels are classified as front, mid, or back, corresponding to the position of the tongue hump, while consonants are basically classified depending on the touch point of the tongue inside the mouth as velar, palatal, retroflex, dental and labial.

The written symbols in Assamese vowel scripts and their corresponding vowel phonemes are presented in TABLE I. It is obvious from these tables that single phoneme may corresponds to more than two or three graphemes.

TABLE I
CLASSIFICATION OF ASSAMESE VOWELS AND THEIR
IPA REPRESENTATIONS

Tongue position →		Front		Central		Back	
Shape of Lips →		Unrounded		Neutral		Rounded	
Height of The Tongue ↓	Space in the Oral Cavity ↓	IPA	Assamese Vowel Phoneme	IPA	Assamese Vowel Phoneme	IPA	Assamese Vowel Phoneme
High	Close	i	ই			u	উ
High-Mid	Half Close	e	এ			o	ও
Low-Mid	Half Open	ɛ	এ			ɔ	অ
Low	Open			a	আ	ɒ	অ

II. ZERO CROSSING RATE

Zero Crossing Rate(ZCR) is proposed for sex identification and result of about 97% for gender classification is obtained. Zero Crossing rate is proposed for musical instrument identification and result reflects more effectively the difference in musical instrument [9]. Zero crossing rate is a measure of the number of times in a given time interval that

the amplitude of the speech signals passes through a value of zero. Zero crossing rate is an important parameter for voiced/unvoiced classification and for end point detection. Zero crossing rate for unvoiced speech is greater than that of voiced speech for its random nature. Detecting when a speech utterance begins and ends is a basic problem in speech processing which is referred to as end point detection [4]. End point detection is difficult if the speech is uttered in noisy environment. For silence zero crossing rate should be zero.

The notation of zero crossing is defined to be “The number of times in a sound sample that the amplitude of the sign wave changes sign.” But it is very difficult to get a noise free utterance. This means that there is some level of background noise, that interferes with the speech, meaning that the silent region actually have quite a high zero crossing rate as the signal changes from just one side of zero amplitude to the other and back again [10]. For this reason a tolerance threshold is included in the function that calculates zero crossing to try and alleviate this problem.

In our study we have used a threshold value of 0.001. This result states that any zero crossings that start and end in the range of ‘x’, where x lies $-0.001 < x < 0.001$, are not included in the total number of zero crossing for that window.

III. EXPERIMENT AND RESULT

The target sample was manually segmented using Audacity Software and stored with .wav extension.

The zero-crossing rate (ZCR) indicates the frequency of signal amplitude sign changes. It can be expressed as follows:

$$\frac{\sum_{N=1}^N \text{SGN } X(n) - \text{SGN } X(n-1)}{2N}$$

Where $\text{SGN}[]$ is a signum function and $X(n)$ is discrete audio signal.

Because of the slowly varying nature of speech signal, it is common to process speech into blocks (**Frames**) over which the properties of the speech waveform can be assumed to remain relatively constant [4]. We record the input

vowel signal wave at sampling frequency $f_s = 16$ KHz. We take Hamming window with the following specifications:

window size=256 samples

window overlap=100 samples

frameTime=((0:frameNum-1)*(frameSize-overlap)+0.5*frameSize)/fs

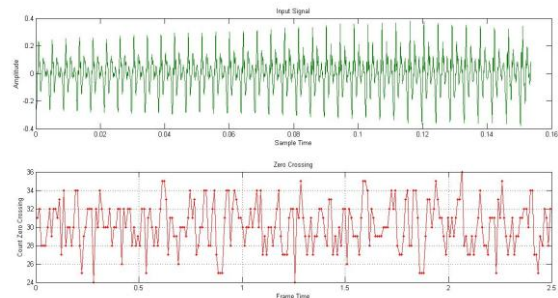
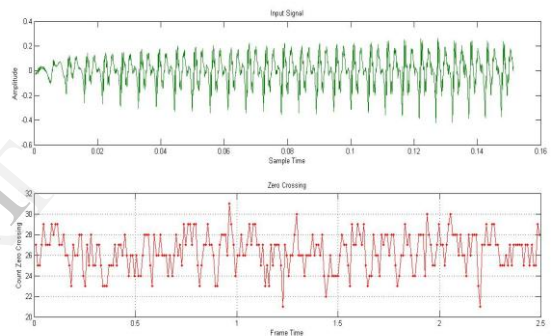


FIG 1: ZCR OF ASSAMESE VOWEL 'e' (IPA /e/) UTTERED BY FEMALE1



IG 1: ZCR OF ASSAMESE VOWEL 'e' (IPA /e/) UTTERED BY MALE1

F

The TABLE II and TABLE III depict the Zero Crossing Rate behaviour of Assamese vowel phonemes frame by frame (**in second**) where the results are taken by processing and executing the different vowel phonemes uttered by 10 male and female Assamese speakers equal each, step by step method, through Matlab.

TABLE II
ZCR OF DIFFERENT EIGHT ASSAMESE VOWELS OF MALE SPEAKERS

V O W E L ↓	F R A M E →	Frames																								
		Male1					Male2					Male3					Male4					Male5				
		0	.5	1	1.5	2	0	.5	1	1.5	2	0	.5	1	1.5	2	0	.5	1	1.5	2	0	.5	1	1.5	2
ই(ɪ)	50	51	50	49	49	49	48	51	51	50	51	49	48	49	51	50	51	51	49	49	50	53	51	49	51	
এ(ɛ)	46	51	48	48	49	48	52	49	49	47	51	48	48	49	48	49	49	47	51	48	48	48	49	48	52	
ঐ(ɛ i)	50	51	50	49	50	49	51	50	49	50	49	49	50	49	51	50	50	49	50	49	49	50	49	51	50	
আ(ɪ a)	48	51	46	49	49	49	49	49	47	51	48	48	49	46	51	48	48	49	48	52	49	46	51	51	49	
অ'(ɪ a)	51	51	49	49	50	53	51	51	50	49	50	49	49	51	48	49	48	52	49	49	48	49	47	51	48	
অ(ɪ n)	49	47	51	48	48	48	49	49	51	50	51	44	50	51	49	47	51	48	48	49	48	49	49	47	51	
ও(ɔ)	45	47	50	51	53	49	50	49	51	50	49	49	50	50	51	53	49	50	49	51	50	50	47	51	48	
উ(ɪ u)	49	48	52	49	49	51	44	50	51	49	53	49	50	49	51	50	53	51	51	50	50	48	52	49	49	

TABLE III
ZCR OF DIFFERENT EIGHT ASSAMESE VOWELS OF MALE SPEAKERS

V O W E L ↓	F R A M E →	Frames																								
		Female1					Female2					Female3					Female4					Female5				
		0	.5	1	1.5	2	0	.5	1	1.5	2	0	.5	1	1.5	2	0	.5	1	1.5	2	0	.5	1	1.5	2
ই(ɪ)		47	48	44	47	48	49	47	51	48	49	51	47	46	48	51	46	51	53	49	50	45	53	50	46	45
এ(ɛ)		48	49	47	51	48	49	51	47	49	47	51	49	49	49	49	47	51	48	48	49	46	51	48	48	49
ঐ(ɛ i)		48	49	49	51	50	51	44	50	51	49	47	51	50	51	48	48	49	46	51	48	48	49	48	52	49
আ(ɪ a)		48	49	51	47	46	48	51	46	51	53	48	48	47	49	47	51	49	49	49	49	51	50	51	44	51
অ(ɪ a)		51	50	51	44	50	51	49	47	51	49	47	51	48	49	51	47	49	47	51	49	49	48	52	49	49
অ(ɪ n)		49	44	47	48	49	47	49	51	50	53	51	51	50	50	48	52	47	48	48	49	48	49	49	47	51
ও(ɔ o)		45	47	50	51	53	49	50	49	51	50	49	49	50	50	51	53	50	51	53	49	50	49	51	50	49
উ(ɪ u)		48	48	49	48	49	49	47	51	48	48	49	48	50	51	50	51	44	50	51	49	47	51	51	50	51

IV. RESULT AND DISCUSSION

In this paper, we represent a method for isolated Assamese speech word recognition based on only Zero Crossing Rate features which would be beneficial for design an Automatic Assamese speaker Recognition System. Speech Recognition is a very difficult task. A speaker recognition works based on the premise that a person's speech exhibits characteristics that are unique to the speaker. We try to show how the Zero Crossing Rate changes in Eight Assamese vowel phonemes frame by frame with time. Result shows that Zero Crossing Rate reflects more effectively in the difference in case of speakers as well as gender classification.

REFERENCES

- [1] Banikanta Kakati, "Assamese, its Formation and Development", 5th edition, Guwahati, India, LBS Publications, 2007
- [2] J.L. Flanagan, "Speech Analysis, Synthesis, and Perception", 2nd edition, New York, 1972, Springer-Verlag.
- [3] J.R. Deller, J. H.L. Hansen, and J.G. Proakis, "Discrete-Time Processing of Speech Signal", New York, 2000, IEEE Press.
- [4] L.R. Rabiner and B.H. Juang, "Fundamentals of Speech Recognition", Englewood Cliffs, New Jersey, 1993, Prentice-Hall.
- [5] L.R. Rabiner and R. Schafer, "Digital Processing of Speech Signals", Englewood Cliffs, NJ, 1979, Prentice-Hall.
- [6] Costas Panagiotakis and G. Tziritas, "A speech/music discriminator based on RMS and Zero -Crossing", IEEE transactions on multimedia. Vol.7, no 1, February 2005.
- [7] Y.K. Lau and Chok K. Chan, "Speech recognition based on Zero-crossing rate", IEEE transactions on acoustic, speech and signal processing, vol, ASSP-33, No 1
- [8] T.K. Das and P.H. Talukder, "Cepstral Analysis of Assamese Vowel Phonemes", IJACST, Volume 2, No 9, August 2013.
- [9] S.K. Banchhor, A. Khan, "Musical Instrument Recognition using Zero Crossing Rate and Short-Time Energy", IJAIS, vol 1, No 3, Feb 2012.
- [10] A. U Khan, L. P. Bhaiya, S.K. Banchhor, "Hindi Speaking Person Identification Using Zero Crossing Rate", IJSCE, vol 2, Issue 3, July 2012.