Zero Crossing Rate Of The Voice And Unvoiced Speech Signal Of Assamese Words

Bitopi Sharma, Prof. P.H. Talukdar

Abstract — Voice activity detecting schemes try to distinguish between speech and noise from an input signal by using some operation such as heuristics methods. Zero-crossing rate, periodicity measures, total energy, low frequency energy, energy in different bands are some of the heuristic methods applied in voice activity detection schemes. In this paper, we used the zero-crossing rate Voice Activity Detecting scheme to identify the voiced and unvoiced sounds for the input speech signal in Assamese Language.

Index Terms—: Assamese scripts, voice and unvoiced sounds, Voice activity detecting schemes, Zero-crossing rate.

1 Introduction

Signal has the information of many types. Noise is the extraneous information in a signal, caused due to electronics, spurious response, and random events. Signal to noise ratio is generally constant and independent of the signal. The impact of noise is greatest on the lowest signal. The ratio of signal to noise is useful in evaluating data. Value of signal to noise varies from signal to signal.

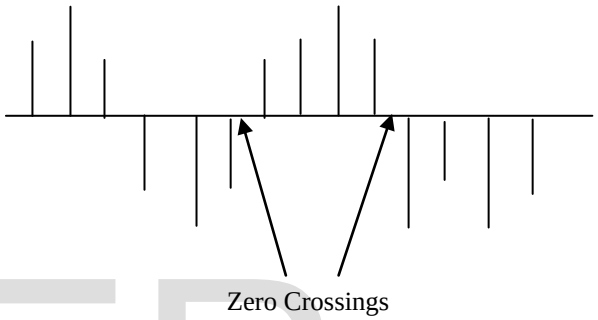
The zero-crossing rate is the rate of sign change along a signal to determine the voiced and unvoiced sounds of an input speech signal. The zero-crossing finds the rate at which the signal changes from positive to negative and vice-versa. This feature of Voice Activity Detection has been used for speech recognition and music information retrieval. Zero –Crossing Rate is defined formally as –

$$ZCR = \frac{1}{T-1}\sum_{t-1}^{T-1} II\{S_t S_{t-1} < 0\} \quad (1)$$

Where S is a signal of length T and the indicator function II{A} is 1 if its argument A is true and zero(0) otherwise.

- *Bitopi Sharma is currently a research scholar in the Department of Instrumentation and USIC, Gauhati University, Guwahati, India PH-+91-8822914024. E-mail: bitopisharma@gmail.com*
- *Prof. P.H. Talukdar is currently working as Professor and Dean, School of Computing Sciences,Kaziranga University, Jorhat-6, Assam, India,PH-+91-9864273197. E-mail: phtassam@gmail.com*

Fig. 1: Definition of zero-crossings rate



Let y(k) = x(k) +v(k) represent a signal x(k) that is corrupted with noise v(k). If Px is the average power of the signal, and Pv is the average power of the noise, then the SNR(Signal to Noise Ratio) is defined as

$$SNR(y) = 10\log10(Px/Pv)dB. \quad (2)$$

The average power Pv is the mean or expected value of the square of the signal

$$Px = E[x2(k)] \quad (3)$$

A signal that is all noise has a SNR of -∞ while a noise free signal has a SNR of +∞. When the SNR is equal to 0dB, the power of the signal is equal to the power of the noise.

The voiced sound in speech is important and two-third of speech signal is voiced. It consists of more or less constant frequency tones of some duration. In the articulatory process when the vocal cords vibrate, voiced sound is produced. The difference between the pair of sounds associated with the alphabets in a language is for the voiced sound. For example – In Assamese language the difference between "জ"/dʒ /, and "য"/dʒ / is understood due to the

IJSER

voiced sound. On the other hand unvoiced speech is non-periodic, random-like sounds, caused by air passing through a narrow constriction of the vocal tract. These sounds are resulted as the noise-like turbulence produced by forcing the air by the lungs with high velocities through a constriction in the vocal tract with glottis open. During whisperings, all sounds are unvoiced sounds.

2 ASSAMESE LANGUAGE

Assam located in the foothills in Himalayan range, popularly known as the land of red rivers and blue hills. Assamese is the main communicating language in the entire North Eastern region of India. It is the easternmost member of the Indo-
European family.Assamese is an old language developing since ancient times having its roots from Sanskrit, Orriya and Magadhi Prakrit. Today's modern Assamese language has a lot of influence from local dialects.

Assamese Language uses Assamese script. The script is developed from Brahmi through Devanagiri. The Assamese phonemic inventory consists of eight oral vowel phonemes, three nasalized vowel phonemes, 15 diphthongs and 21 consonant phonemes. The sound system of a language may be referred as the phonological structure. The Assamese phoneme inventory consists of eight vowels and twenty one consonants, fifteen diphthongs are attested. In a word three syllables may appear in succession comprising of five vowels.

In any language utterance is associated with pitch and juncture features and in Assamese it is also true. Assamese language is based on the pronunciation of the phonemes (letters) and the graphemes are classified- kanthabarna (কণ্ঠ্যৱৰ্ণ),talbarna (তালব্যবৰ্ণ),dantabarna (দন্ত্যৱৰ্ণ),ushabarna (ওষ্ঠ্যৱৰ্ণ), anunaxik barna (অনুনাসিক বৰ্ণ৩)

3 METHODOLOGY

A speech signal is continuous in nature but it can be treated independently, subdividing it into continuous frames. Our objective is to determine Zero Crossing Rate(ZCR) of Assamese speech signals. Depending upon the ZCR, we determine whether the sound is voiced or unvoiced. Data is collected from the Assamese dictionary "Hemkosh". Total 12 speakers – six females and six males have completed the recording. About 1,000 words were experimented. As per requirement, recording was completed in both noiseless and noise environment. The frequency for the kinds of recording was set on 48,000 Hz with 16bit PCM.

ZCR is an important parameter for voiced, unvoiced signal classification. It is often used as a part of the front-end processing in Automatic Speech Recognition system. Speech signals are broadband signals, thus its results are less precise when interpreted on an average. Mathematically it can be defined as-

$$Zn=\sum_{m=-\infty}^{\infty}|sgn[x(m)]-sgn[x(m-1)]|w(n-m) \quad (4)$$

Where

$sgn[x(n)]=1 \quad x(n)\geq 0$

$= -1 \quad x(n)<0$

And

$w(n) = 1/2N \quad 0\leq n \leq N-1$

$=0$ otherwise.

The equation can be diagrammatically represented as-

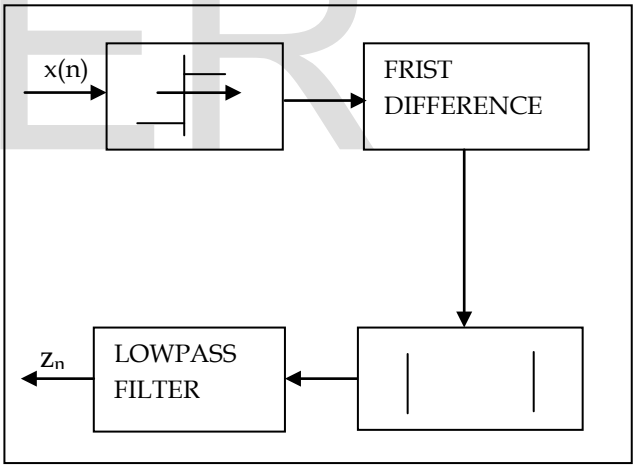


Fig 2: Block diagram representations of short-time average zero crossing.

The Short-Time Energy (STE) of speech signals reflects the amplitude variation. It is a processing technique that gives time domain features (Qn). It is mathematically represented as

$$Qn=\sum_{m=-\infty}^{\infty} T[x(m)]w(n-m) \quad (5)$$

Where

T[] is the transformation matrix. The transformation matrix could be either linear or nonlinear.

X(m) represents the data sequence and

W(n-m) represents a limited time window sequence.

The energy of the discrete time signal is defined as

$$E = \sum_{m=-\infty}^{\infty} X2(m) \quad (6)$$

The STE of the speech signal that reflects the amplitude variation may be represented as

$$En = \sum_{m=-\infty}^{\infty} [x(m)W(n-m)]2 \quad (7)$$

A simple algorithm to identify voiced and unvoiced sounds of an input speech signal is given below:

Step1: Accept input signal

Step2: Compute signal processing frame by frame for every signal.

Step3: Compute ZCR and Short Time Energy (STE)

Step3: Check whether is value is greater than STE.

Step 4: If the ZCR value is greater than STE then the input signal is voiced, else unvoiced.

4 RESULT

After taking the Assamese words as our input signal every word was passed through the frame-by-frame processing stage. After that the speech signal were segmented into a non-overlapping frame of samples. The step has been repeated until the entire speech signal is covered to get the ZCR values. The table below shows the ZCR values for a few Assamese words.

Word	Structure	ZCR (male)	ZCR (Female)
/jam/ 'যাম' 'will go'	CVC	883	1299
/asa/ আছা 'are you there'	VCV	1615	1632
/ma/ 'মা' 'mother'	CV	380	401
/ek/ 'এক' 'one'	VC	179	240

Table1: ZCR values for male and female.

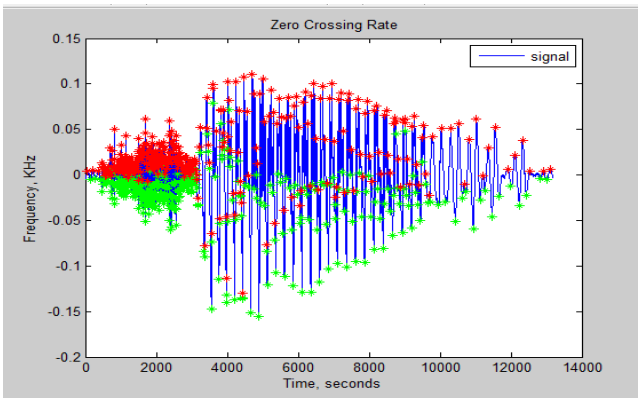


Fig 3: ZCR for the Assamese word /jam/ 'যাম' (female)

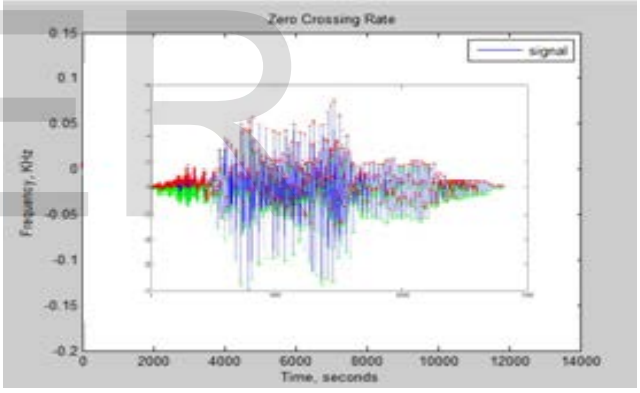


Fig 4: ZCR values for the Assamese word /jam/ 'যাম' (male)

5 CONCLUSION

In this paper we present an approach to identify Voice activity detecting scheme using ZCR to identify voiced and unvoiced signals for Assamese word. Result shows that the estimation of ZCR reflects effectively in voiced and unvoiced sounds. We found that separating the voiced and unvoiced sound of speech is a simple and efficient way.

REFERENCES

[1] D.S.Shete 1, Prof. S.B. Patil 2,Prof. S.B. Patil, Zero crossing rate and Energy of the Speech Signal of Devanagari Script.

IJSER

[2] Bachu R.G., Kopparthi S., Adapa B., Barkana B.D., Separation of Voiced and Unvoiced using Zero crossing rate and Energy of the Speech Signal.
[3] Arif Ullah Khan, L.P.Bhaiya Sumit Kumar Banchhor , Hindi Speaking Person Identification Using Zero Crossing Rate.
[4] Dr. Golukchandra Goswami,Structure of Assamese.
[5] http://www.iitg.ernet.in/rcilts/pdf/assamese.pdf .P7.
[6] Dr. Pran Hari Talukdar, Speech Production, Analysis and Coding Introduction to speech processing
[7] L.R.Rabiner, R.W. Schafer, Digital Processing of Speech Signal.

IJSER