IOSR Journal of VLSI and Signal Processing (IOSR-JVSP)
Volume 4, Issue 1, Ver. I (Jan. 2014), PP 01-05
e-ISSN: 2319 – 4200, p-ISSN No. : 2319 – 4197
www.iosrjournals.org

Zero crossing rate and Energy of the Speech Signal of Devanagari Script

D.S.Shete [1], Prof. S.B. Patil [2],Prof. S.B. Patil [3]

1(Department Of Electronics Engineering , J.J.Magdum College of Engg. ,Jaysingpur, Shivaji University Kolhapur,India)
2(Department Of Electronics Engineering , J.J.Magdum College of Engg. ,Jaysingpur, Shivaji University Kolhapur,India)
3(Department Of Electronics Engineering , D.Y.Patil College of Engg. & Tech., Kolhapur, Shivaji University Kolhapur,India)

Abstract: *In speech analysis, the voiced-unvoiced decision is usually performed in extracting the information from the speech signals. In this paper, we performed two methods to separate the voiced- unvoiced parts of speech from a speech signal. These are zero crossing rate (ZCR) and energy. In here, we evaluated the results by dividing the speech sample into some segments and used the zero crossing rate and energy calculations to separate the voiced and unvoiced parts of speech. The results suggest that zero crossing rates are low for voiced part and high for unvoiced part where as the energy is high for voiced part and low for unvoiced part. Therefore, these methods are proved more effective in separation of voiced and unvoiced speech*

Keywords*: Devnagari script, zero crossing rate,energy of speech signal*

I. Introduction

Speech can be divided into numerous voiced and unvoiced regions. The classification of speech signal into voiced, unvoiced provides a preliminary acoustic segmentation for speech processing applications, such as speech synthesis, speech enhancement, and speech recognition.“Voiced speech consists of more or less constant frequency tones of some duration, made when vowels are spoken. It is produced when periodic pulses of air generated by the vibrating glottis resonate through the vocal tract, at frequencies dependent on the vocal tract shape. About two-thirds of speech is voiced and this type of speech is also what is most important for intelligibility. Unvoiced speech is non-periodic, random-like sounds, caused by air passing through a narrow constriction of the vocal tract as when consonants are spoken. Voiced speech, because of its periodic nature, can be identified, and extracted .In recent years considerable efforts has been spent by researchers in solving the problem of classifying speech into voiced/unvoiced parts . A pattern recognition approach and statistical and non statistical techniques has been applied for deciding whether the given segment of a speech signal should be classified as voiced speech or unvoiced speech. Qi and Hunt classified voiced and unvoiced speech using non-parametric methods based on multi-layer feed forward network . Acoustical features and pattern recognition techniques were used to separate the speech segments into voiced/unvoiced .

The method we used in this work is a simple and fast approach and may overcome the problem of classifying the speech into voiced/unvoiced using zero-crossing rate and energy of a speech signal. Here we are using phonemes of devnagari script a speech sample. The methods that are used in this study are presented in the second part. The results are given in the third part.

II. The Devnagari Script

Devnagari script is a script of phonemes arranged in a well structured scientific manner showing unambiguous classification and grouping of phonemes according to the organs used in producing that sound. The letter order of Devnagari is based on phonetic principles which consider both the manner and place of articulation of the consonants and vowels they represent. Accordingly these letters (Akshar) are grouped into different classes called Varnas ("TulyasyaPrayatnam Savarnam") . Every letter and its pronunciation is unique and can't be represented or pronounced by using any other letter(s). This gives us a unique representation for every word uttered by human irrespective of human and context of speech. This feature is absent in languages like English in which one representation and pronunciation of a word or letter can be done in more than one way, e.g. bye, buy both are pronounced similarly.

The first 25 consonants of Devnagari script, arranged in a 5X5 matrix, form five different groups of phonemes as in Table 1 Each row of five consonants is generated in totally different way. First four rows are classified depending on the touch point of tongue inside the mouth as Kanthhawya (Velar), Talawya (Palatal), Murdhanya (Retroflex) and Dantawya (Dental). The fifth group is called Aushthawya (Labial) because it is

generated using lips only. The elements in a single row are generated using the same organs but varying the time period of touch and pressure at the same or near the touch point of group.Different phonemes in these varnas are Shown in following Table 1.

Table 1 Phonemes of Devnagari script

Phone class	Class variant				
	Non-voiced		voiced		Nasal
Kanthwya	ka	kha	ga	gha	nga
Talwaya	cha	chha	ja	jha	nja
Murdhanya	ta	tha	da	Dha^	na^
Dantawya	ta	tha	da	dha*	na*
Aushthawya	pa	pha	ba	ma	ma

III. Present work

The objective of work is to determine zero crossing rate and energy calculation of Devnagari speech sample. Depending upon the ZCR and energy ,we determine whether the segment is voiced or unvoiced.The steps used in the present work system are discussed below.

3.1 Input Acquisation

After capturing the speech by using microphone the speech data is saved in .wav files. For that purpose window XP sound recorder is used. We are using complete list of Devnagari alphabets uttered by 10 different persons (5 females + 5 males) in normal daily use rooms at 8 kHz with 8 bits per sample.The work being focused mostly on calculating zero crossing rate and energy calculation of speech sample.

3.2 The Speech Utterance (Data Collection)

The source of data is a database consisting of 25 characters taken from 5 phone classes and spoken 10 times by 10 speakers; those are 5 males and 5 females of various ages.The data, which is speaker dependent, will be used for further processing.. These characters are recorded by Windows XP sound recorder with sampling rate 8 kHz, 8-bit and mono is used to record the utterance.

IV. Method

In our design, we combined zero crossings rate and energy calculation. Zero-crossing rate is an important parameter for voiced/unvoiced classification. It is also often used as a part of the front-end processing in automatic speech recognition system. The zero crossing count is an indicator of the frequency at which the energy is concentrated in the signal spectrum. Voiced speech is produced because of excitation of vocal tract by the periodic flow of air at the glottis and usually shows a low zero-crossing count], whereas the unvoiced speech is produced by the constriction of the vocal tract narrow enough to cause turbulent airflow which results in noise and shows high zero-crossing count. Energy of a speech is another parameter for classifying the oiced/unvoiced parts The voiced part of the speech has high energy because of its periodicity and the unvoiced part of speech has low energy.

4.1. Zero-Crossings Rate

In the context of discrete-time signals, a zero crossing is said to occur if successive samples have different algebraic signs. The rate at which zero crossings occur is a simple measure of the frequency content of a signal. Zero-crossing rate is a measure of number of times in a given time interval/frame that the amplitude of the speech signals passes through a value of zero, Fig 1. Speech signals are broadband signals and interpretation of average zero-crossing rate is therefore much less precise However, rough estimates of spectral properties can be obtained using a representation based on the shorttime average zero-crossing rate .

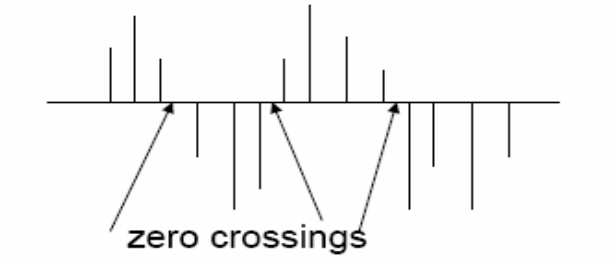


Fig. 1: Definition of zero-crossings rate

A definition for zero-crossings rate is:

$$Z_n = \sum_{m=-\infty}^{\infty} \left| \text{sgn}[x(m)] - \text{sgn}[x(m-1)] \right| w(n-m) \qquad (1)$$

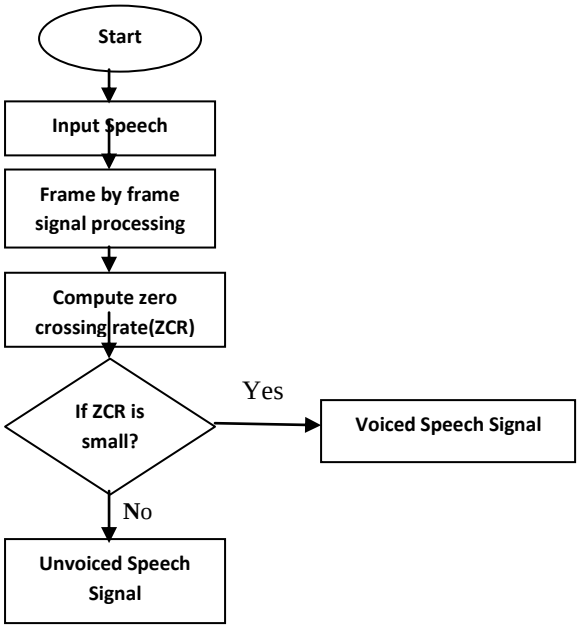
where

$$\text{sgn}[x(n)] = \begin{cases} 1 & x(n) \geq 0 \\ -1 & x(n) < 0 \end{cases}$$

And w(n) is the windowing function with a window size of N samples

$$W = \begin{cases} 1/2N & 0 \leq n \leq N-1 \\ 0 & \text{otherwise} \end{cases}$$

The analysis for classifying the voiced/unvoiced parts of speech has been illustrated in the flow chart in Fig.2



4.2 Energy of the Discrete Speech Signal

The amplitude of unvoiced segments is noticeably lower than that of the voiced segments. The short-time energy of speech signals reflects the amplitude variation. In a typical speech signal we can see that its certain properties considerably changes with time.For example, we can observe a significant variation in the peak amplitude of the signal and a considerable variation of fundamental frequency within voiced regions in a speech signal. These facts suggest that simple time domain processing techniques should be capable of providing useful information of signal features, such as intensity, excitation mode, pitch, and possibly even vocal tract parameters, such as formant frequencies. Most of the short time processing techniques that give time domain features (Qn), can be mathematically represented as

$$Q_n = \sum_{m=-\infty}^{\infty} T[x(m)]w(n-m) \qquad (2)$$

where T[] is the transformation matrix which may 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

$$E = \sum_{m=-\infty}^{\infty} X^2(m) \qquad (3)$$

Such a quantity has little meaning or utility for speech since it gives little information about time dependent properties of speech signal. We have observed that the amplitude of the speech signal varies appreciably with time. In particular, the amplitude of the unvoiced segment is generally much lower than amplitude of the voiced segment. The short time energy of the speech signal provides a convenient representation that reflects the amplitude variation and can be defined as

$$En= \sum_{m=-\infty}^{\infty} [x(m)W(n-m)]^2 \quad (4)$$

The major significance of *En* is that it provides a basis for distinguishing voiced speech segment from unvoiced speech segment. It can be seen that the value of *En* for the unvoiced segments are significantly smaller than voiced segments. The energy function can also be used to locate approximately the time at which voiced speech become unvoiced speech and vice versa, and for high quality speech (high signal to noise ratio) the energy can be used to distinguish speech from silence. In voiced speech the short-time energy values are much higher than in unvoiced speech , which has a higher zero crossing rate. The above discussion cites the importance of energy function (*En*) for speech analysis purpose.

V. Results

MATLAB 7..3 is used for our calculations. We chose MATLAB as our programming environment as it offers many advantages. It contains a variety of signal processing and statistical tools, which help users in generating a variety of signals and plotting them. MATLAB excels at numerical computations, especially when dealing with vectors or matrices of data.One of the speech signal used in this study is given with Fig.6. Proposed voiced/unvoiced classification algorithm uses short-time zero-crossings rate and energy of the speech signal. The results of voiced/unvoiced decision using our model are presented in Table2.

In the frame-by-frame processing stage, the speech signal is segmented into a non-overlapping frame of samples. It is processed into frame by frame until the entire speech signal is covered. Table 2 includes the voiced/unvoiced decisions for phoneme "ka." It has 3600 samples with 8000Hz sampling rate. At the beginning, we set the frame size as 100 samples. For every 100 samples we calculate zero crossing rate and energy of speech signal.

Figure 3 original speech signal for phoneme ka

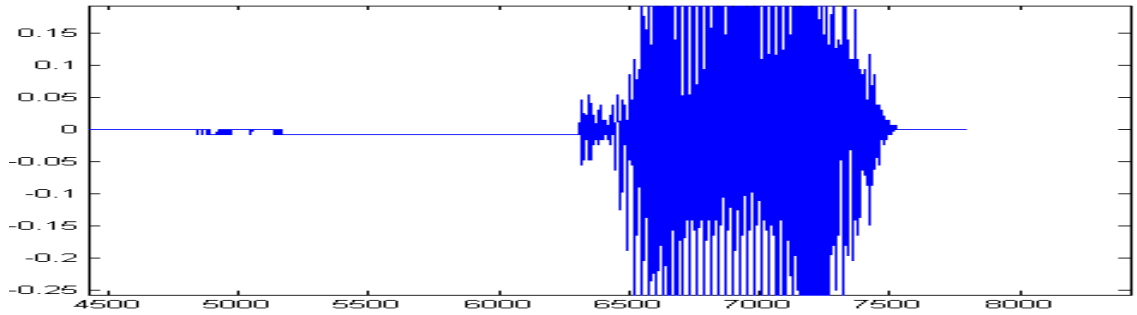


Figure 4 Frame-by–frame processing of speech signal.

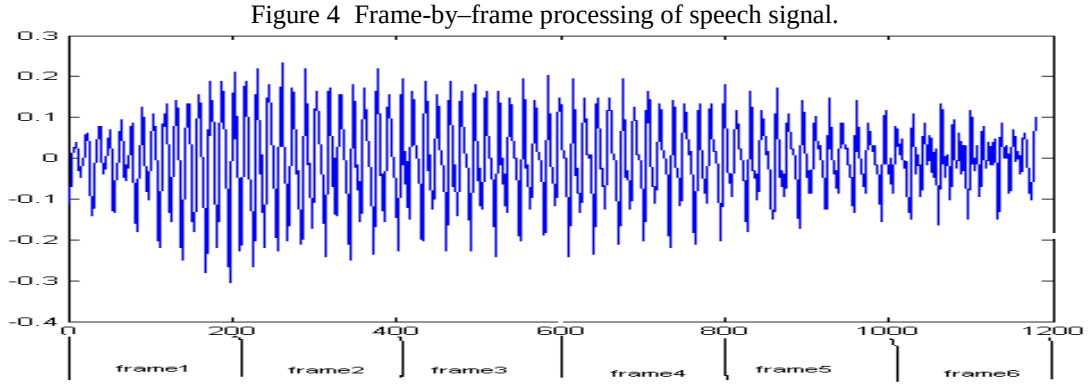


Table 2: Voiced/unvoiced decisions for the phoneme "ka"

Frames Phoneme ka Sampling frequency=8khz	ZCR	Energy	Decision
Frame-1(100 Samples)	0.16	0.2188	voiced
Frame-2(100 samples)	0.14	0.435	voiced
Frame-3(100 samples)	0.20	0.0902	unvoiced
Frame-4(100 samples)	0.16	0.1382	voiced
Frame-5(100 samples)	0.19	0.0804	unvoiced
Frame-6(100 samples)	0.14	0.0617	unvoiced
Frame-7(100 samples)	0.15	0.0373	unvoiced
Frame-8(100 samples)	0.15	0.0748	unvoiced

VI. Conclusion

We have presented an approach for separating the voiced /unvoiced part of speech in a simple and efficient way. The algorithm shows good results in classifying the speech as we segmented speech into many frames. For unvoiced speech, most of the energy is found at higher frequencies. Since high frequencies imply high zero crossing rates, and low frequencies imply low zero-crossing rates, there is a strong correlation between zero-crossing rate and energy distribution with frequency. A reasonable generalization is that if the zero-crossing rate is high, the speech signal is unvoiced, while if the zero-crossing rate is low, the speech signal is voiced.

References

Jouneral Papers:

[1] 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

[2] Jong Kwan Lee, Chang D. Yoo, "Wavelet speech enhancement based on voiced/unvoiced decision",Korea Advanced Institute of Science and Technology The 32nd International Congress and Exposition on Noise Control Engineering, Jeju International Convention Center, Seogwipo, Korea,August 25-28, 2003.

[3] B. Atal, and L. Rabiner, "A Pattern Recognition Approach to Voiced-Unvoiced-Silence Classification with Applications to Speech Recognition," IEEE Trans. On ASSP, vol. ASSP-24, pp. 201-212, 1976. [3] S. Ahmadi, and A.S. Spanias, "Cepstrum-Based Pitch Detection using a New Statistical V/UV Classification Algorithm," IEEE Trans. Speech Audio Processing, vol. 7 No. 3, pp. 333-338, 1999.

[4] Y. Qi, and B.R. Hunt, "Voiced-Unvoiced-Silence Classifications of Speech using Hybrid Features and a Network Classifier," IEEE Trans. Speech Audio Processing, vol. 1 No. 2, pp. 250-255, 1993.

[5] L. Siegel, "A Procedure for using Pattern Classification Techniques to obtain a Voiced/Unvoiced Classifier", IEEE Trans. on ASSP, vol. ASSP-27, pp. 83- 88, 1979.

[6] T.L. Burrows, "Speech Processing with Linear and Neural Network Models", Ph.D. thesis, Cambridge University Engineering Department, U.K., 1996.

[7] D.G. Childers, M. Hahn, and J.N. Larar, "Silent and Voiced/Unvoiced/Mixed Excitation (Four-Way) Classification of Speech," IEEE Trans. on ASSP, vol. 37 No. 11, pp. 1771-1774, 1989.

[8] Jashmin K. Shah, Ananth N. Iyer, Brett Y. Smolenski, and Robert E. Yantorno "Robust voiced/unvoiced classification using novel features and Gaussian Mixture model", Speech Processing Lab., ECE Dept., Temple University, 1947 N 12th St., Philadelphia, PA 19122-6077, USA.

[9] Jaber Marvan, "Voice Activity detection Method and Apparatus for voiced/unvoiced decision and Pitch Estimation in a Noisy speech feature extraction", 08/23/2007, United States Patent 20070198251.

[10] Thomas F. Quatieri, Discrete-Time Speech Signal Processing: Principles and Practice, MIT Lincoln Laboratory, Lexington, Massachusetts, Prentice Hall, ISBN-13:9780132429429.

[11] Rabiner, L. R., and Schafer, R. W., Digital Processing of Speech Signals, Englewood Cliffs, New Jersey, Prentice Hall, 512-ISBN-13:9780132136037, 1978. Short Biographies of the Authors: Rajesh G. Bachu is Graduate Assistant in Electrical Engineering at the University of Bridgeport, Bridgeport, CT