Journal of Science and Technology

ISSN: 2456-5660 Volume 6, Issue 1, Jan-February 2021

www.jst.org.in **DOI: *https://doi.org/10.46243/jst.2021.v6.i1.pp167-171***

Devanagari Script using Energy of Speech Signal

J.Sneha Latha 1, G.Usha Rani 2

1(Assistant Professor, Department Of Electronics and Communication Engineering , MGIT, Hyderabad, India)

2(Assistant Professor, Department Of Electronics and Communication Engineering , MGIT, Hyderabad, India)

1Corresponding Author: sneha.jaladi@gmail.com

To Cite this Article

J.Sneha Latha, G.Usha Rani "**Devanagari Script using Energy of Speech Signal**", *Journal of Science and Technology, Vol. 06, Issue 01,Jan-Feb 2021, pp.:167-171.*

Article Info

Received: 08-02-2021 *Revised: 10-02-2021* *Accepted: 13-02-2021* *Published: 15-02-2021*

ABSTRACT:

The voiced-unvoiced decision is often used in speech analysis to extract information from speech signals. Voiced and unvoiced components of speech were separated using two approaches in this study. ZCR and energy are the terms used to describe these concepts. Using the zero crossing rate and energy estimates, we were able to differentiate the voiced and unvoiced sections of the speech in this study. As a consequence of these findings, it seems that zero crossing rates for voiced and unvoiced parts are different, with the former having a higher energy level and the latter having a lower one. Because of this, they have been shown to be more adept in separating spoken and unspoken language.

Devnagari script, zero crossing rate, and voice signal energy are among the key terms.

Introduction

Numerous voiced and unvoiced areas are available in speech. Preliminary acoustic segmentation for speech processing applications such as speech synthesis, speech augmentation, and speech recognition is provided by voiced and unvoiced categorization of the audio signal "Voiced speech is made up of tones of varying frequencies and durations, which are produced when vowels are pronounced. It is caused by the vibrating glottis resonating through the vocal tract at a frequency that is dependent on the geometry of the vocal tract. Two-thirds of communication is vocal; this is the most crucial kind of speech for comprehension. Non-periodic, random-sounding sounds are produced by air flowing through a tight constriction of the vocal tract, as is the case when consonants are utilised in speech. Voiced speech may be recognised and extracted due of its periodic character. Researchers have spent a lot of time recently trying to find a solution to the challenge of categorising speech into voiced and unvoiced sections. For determining whether a particular segment of a speech signal should be classed as voiced or unvoiced, statistical and non-statistical approaches have been used. Qi and Hunt used non-parametric approaches to classify voiced and unvoiced speech, based on a multi-layer feed forward network. The speech samples were divided into voiced and unvoiced using acoustic characteristics and pattern recognition algorithms.

Using zero-crossing rate and energy of a speech signal, we may be able to solve the challenge of determining whether a speech signal is voiced or not. Phonemes from the devnagari script are used in this voice sample. Methods for this investigation are detailed in the next section. The findings are presented in the last section.

The Devnagari Script

For the Devnagari script, which is an alphabet of phonemes, the categorization and grouping of phonemes according to the organs involved in making that sound is well defined. When it comes to Devnagari, the alphabet is organised according to phonetic principles, which take into account both how and where each letter is spoken. Varnas (TulyasyaPrayatnam Savarnam) are arranged according to these letters (Akshar). In order to symbolise or pronounce a letter, you cannot use another letter (s). Every human-uttered word now has its own unique representation, independent of the speaker or the context in which it was spoken. In languages like English, where a word or letter may be represented and spoken in a variety of ways, this trait is missing. For example, bye and buy are both pronounced the same.

As seen in Table 1, the first 25 consonants of the Devnagari script create five distinct groups of phonemes. A unique set of five consonants is created for each row of the alphabet. To classify these first four rows of teeth into one of the following: Kanthhawya (Velar), Talawya (Palatal), Murdhanya (Retroflex), and Dantawawya (Dental) (Dental). Aushthawya (Labial) is the name given to the fifth group, which is formed only by the lips. All the components in a group have the same organs but differ in how long and how hard they are pressed against each other. Table 1 shows the several phonemes that fall within each of these varnas.

Journal of Science and Technology

ISSN: 2456-5660 Volume 6, Issue 1, Jan-February 2021

www.jst.org.in **DOI:** *https://doi.org/10.46243/jst.2021.v6.i1.pp167-171*

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

Present work

Devnagari speech sample zero crossover rate and energy calculation are the goals of this project. Whether a segment is voiced or not is determined by the ZCR and energy. The current system of work is described below.

This section focuses on the acquisition of input.

WAV files are created when the audio has been recorded using a microphone. Window XP's sound recorder is used for this purpose. Devnagari alphabets are being pronounced by 10 distinct people (5 ladies and 5 men) in regular everyday usage rooms at a sampling rate of 8 kHz and an average bit depth of 8 bits. Working on zero crossing rate and energy computation of voice samples is the primary emphasis.

3.2 The Utterance of Speech (Data Collection)

A database of 25 characters drawn from five phonetic classes and uttered ten times by ten speakers, five men and five women of varied ages, served as the data source.

Further processing will be done using the speaker-dependent data. Windows XP sound recorder with sampling rate 8 kHz, 8-bit, and mono is utilised to record the utterances of these characters.

Method

The zero crossings rate and the energy calculation were included into our design. For voiced/unvoiced categorization, zero-crossing rate is critical. Front-end processing is also often utilised in automated voice recognition systems. Counting the number of zero crossings in a signal spectrum indicates the frequency at which the energy is most concentrated. Vocal tract excitation is caused by the periodic flow of air at the glottis, which results in a low zero-crossing count, whereas constriction of the vocal tract narrow enough to cause turbulent airflow, which results in noise and a high zero-crossing count, is responsible for producing unvoiced speech. Another factor in determining whether a speech is oiced or unvoiced is its level of intensity. Because of its regularity, the voiced portion of the speech is more energetic than the unvoiced portion.

Rate of No Crossings

Zero crossing occurs when the algebraic signs of consecutive samples are different, in the setting of discrete-time signals. The frequency content of a signal may be measured by the frequency of zero crossings. zero crossing rate is an indicator of how many times speech signal amplitudes pass through a value of zero in any particular time period or frame. The average zero-crossing rate of a speech signal is substantially less exact since it is a broadband signal. However, the short-term average zero-crossing rate may be used to assess the spectral features of a signal.

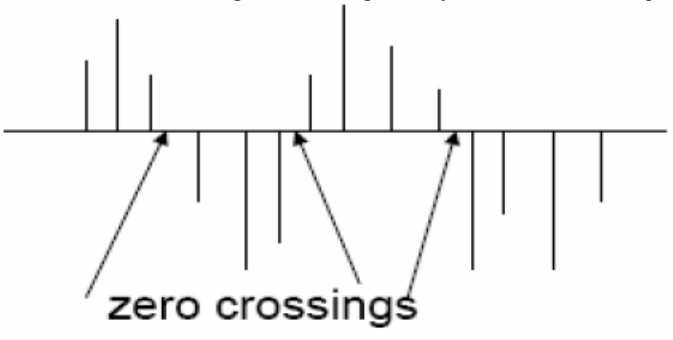


Fig. 1: Definition of zero-crossings rate

A definition for zero-crossings rate is:

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

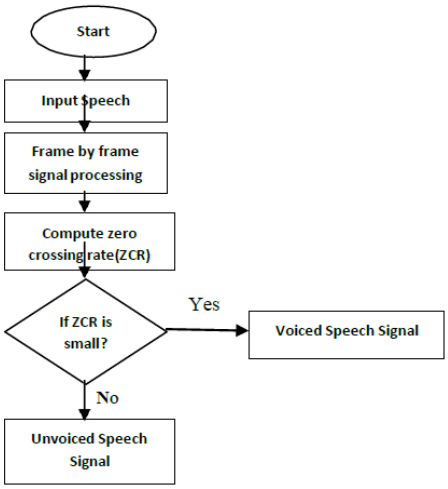
where

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

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



Energy of the Discrete Speech Signal

Segments that are not voiced have much lower amplitudes. The amplitude volatility of speech transmissions is reflected in the speech signal's short-term energy. Time-varying features are evident in speech signals, as we can observe in a typical speech signal. In a speech signal, for example, we may see large variations in peak amplitude and fundamental frequency among the voiced areas. In light of these facts, basic time domain processing approaches should be able to provide valuable information on signal properties such as strength, excitation mode and pitch, and maybe even vocal tract parameters such as formant frequencies. Short-term processing strategies (Qn) may be quantitatively described as a function of time.

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

W(n-m) depicts a restricted time window sequence where T[] is the transformation matrix that may be either linear or nonlinear. The discrete time signal's energy is denoted by

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

For speech, such a number has no significance or usefulness since it offers no insight into the time-dependent features of the voice signal. When listening to a voice signal, we've noticed that its loudness changes significantly over time. The unvoiced segment's amplitude is often substantially lower than the voiced segment's. To easily express the amplitude fluctuation, the voice signal's short time energy may be defined as

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

As a starting point for discriminating between voiced and non-voiced speech segments, En is critical. Unvoiced segments have a lower En value than voiced segments, as can be seen in the graph. It is possible to use the energy function to find roughly the period at which voiced speech

becomes unvoiced speech and vice versa, and for high-quality speech (high signal to noise ratio), the energy may be utilised to discriminate speech from silence. Compared to unvoiced speech, which has a lower zero crossing rate, the short-time energy values in voiced speech are substantially greater. Speech analysis relies heavily on the energy function (En), which has been discussed in the preceding section in some detail.

The results

For our computations, we utilise MATLAB 7..3 and 7.4. Our programming environment of choice is MATLAB, which has a number of benefits. Users may generate and visualise a variety of signals with the aid of a wide range of signal processing and statistical tools. When it comes to numerical calculations, MATLAB is the best. Fig.6 depicts one of the speech signals employed in this investigation. Zero crossing rate and energy of the speech signal are used in a proposed algorithm for determining whether a person is speaking or not. Table 2 shows the outcomes of our model's voiced/unvoiced choice.

A non-overlapping frame of samples is created at the frame-by-frame processing step. Frame by frame, the complete voice stream is decomposed and processed. Voiced versus unvoiced judgments for the phoneme "ka" may be found in Table 2. It features a sampling rate of 8000 Hz and 3600 samples. The frame size was initially set at 100 samples. Zero crossing rate and voice signal energy are calculated for every 100 samples.

Figure 3 original speech signal for phoneme ka

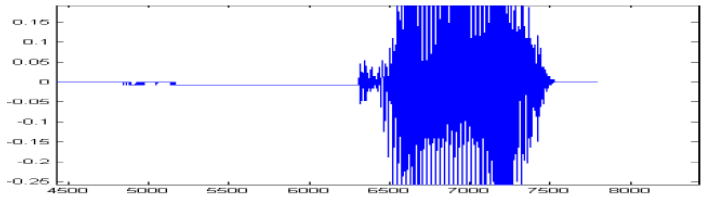


Figure 4 Frame-by–frame processing of speech signal.

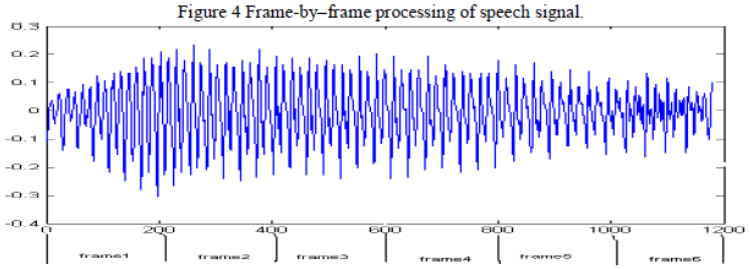


Table 2: Voiced/unvoiced decisions for the phoneme "ka"

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

Conclusion

We have shown a simple and quick method for differentiating the voiced and unvoiced parts of speech. We were able to get decent results with the algorithm since we divided the voice into a number of frames. Unvoiced speech's energy is concentrated in the higher frequencies. There is a significant association between zero-crossing rate and energy distribution with frequency since high frequencies predict high zero crossing rates and low frequencies imply low zero crossing rates. According to the zero-crossing frequency, the voiced and unvoiced signals may be distinguished by looking at the zero-crossing frequency.

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