International Journal of Advanced Information Science and Technology (IJAIST) ISSN: 2319:2682
Vol.18, No.18, October 2013

Voiced/Unvoiced Classification of Speech Signal Using Average Zero Crossing Index Difference Function

A.Milton
Asst. Professor, Department of ECE,
St. Xavier's Catholic College of Engineering,
Nagercoil, India.
milton@sxcce.edu.in

S. Ashitha Dayana
PG Student, Applied Electronics,
St. Xavier's Catholic College of Engineering,
Nagercoil, India.
ashithadayana@gmail.com

S.Tamil Selvi
Professor,
National Engineering College
Kovilpatti, India.
tamilgopal2004@yahoo.co.in

***Abstract*— In this paper, a new method is proposed for the classification of voiced/unvoiced speech signals. The vocal cord vibration is present in the voiced region and it is absent in the unvoiced region. The voiced, unvoiced classification is required for many applications, including analysis, synthesis, and recognition applications. We propose Average zero crossing index difference function (AZIDF) to classify voiced and unvoiced region and the performance are analyzed. In this method, the average of the difference between adjacent zero crossing index values are compared with a threshold to classify voiced and unvoiced region. The performance of AZIDF classification is compared with the zero crossing rate (ZCR) classification. The classification rate is high in AZIDF classification compared with the ZCR classification.**

***Index Terms*- Average zero crossing index difference function, zero crossings,ZCR,Voiced and Unvoiced classification.**

I. INTRODUCTION

The voiced, unvoiced and silence (absence of speech) classification is usually performed for the extraction of information from the speech signals. Voiced sounds are produced when the vocal cords vibrate during the pronunciation of a phoneme, thus interrupting the flow of air from the lungs to the vocal tract and producing quasi-periodic pulses of air as the excitation. In unvoiced sounds, the vocal chords are not vibrated, so there is no vibration in the throat. Unvoiced sounds results when the excitation is a noise like turbulence produced by forcing air at high velocities through a constriction in the vocal tract while the glottis is held open. The periodicity of this vibration makes the voiced segments periodic and so distinguishable from the noisy-like unvoiced segments. Since the speech signals are quasi-periodic, making the decision gets hard. The pitch of the particular speech signal is present only in the voiced part of the signal. Numerous approaches have been proposed by the researchers for the classification of voiced and unvoiced sounds. This classification is required for many applications, including modeling for analysis/synthesis, detection of model changes for segmentation purposes and signal characterization for indexing and recognition applications [16,17].

Most commonly used method is the extraction of features from speech segments and makes the voiced, unvoiced decision according to whether the value of the feature is above or below a pre-determined threshold. The extracted features are cepstral peaks [11], zero-crossing rate [4,13,14], energy [4,13,14], auto-correlation function peak [4,11], or harmonic to noise ratio in the sinusoidal model of speech signal[11]. Different methods have been used in the field of multi- feature voicing decision. The authors Alexandru Caruntu, and Alina Nica [3] investigates a few methods of automatic classification of speech in silence/voiced/unvoiced (SUV) regions, using both time and frequency domain parameters. The features include zero-crossings, root mean square energy (RMSE), but also a modified version of Teager energy. Results proved that RMSE and Teager energy in conjunction with zero crossings have an accuracy of around 66%.

Yingyong Qi and Bobby R. Hunt [15] done the classification using Multilayer feed forward network and the feature vectors used for this classification is the combination of both cepstral coefficients and waveform features. The authors Jashmin K. Shah et.al., [9], proposed two approaches with Gaussian Mixture Model classifier based on Mel frequency cepstral coefficient and reduced dimensional Linear Predictive Coding (LPC) residual for voiced and unvoiced classification. A multilayer perceptron classifier[5] also have been used for the classification and the features are trained using, Steepest Descent Minimization algorithm[5],Genetic algorithms[8].Ji-Hyun Song and Joon-Hyuk Chang [10] used Gaussian Mixture Model(GMM) to classify voiced/ unvoiced sounds for Selectable Mode Vocoder(SMV). The feature vectors which are applied to the GMM are selected from relevant parameters of SMV.

A.E. Mahdi and E. Jafer [1] proposed a new wavelet-based algorithm for voice/unvoiced classification of speech segments based on the statistical analysis of the energy-frequency distribution of the speech signal using wavelet transform. The authors Dhany Arifianto and Takao Kobayashi [6,7] proposed voiced/ unvoiced algorithm which uses instantaneous frequency amplitude spectrum in adverse environment using harmonicity measure acquired after evaluating instantaneous frequency is derived from short time Fourier transform of a signal as a function of time and frequency .The authors Arthur P. Lobo and Philipos C. Loizou [2] proposed an algorithm for voiced/unvoiced speech discrimination in noise that is based on the Gabor atomic decomposition of the speech waveform. N. Dhananjaya and B. Yegnanarayana [12] proposed a new method for voiced/nonvoiced detection based on epoch extraction. Zero-frequency filtered speech signal is used to extract the instants of significant excitation. The robustness of

the method is to extract epochs in the voiced regions, even with small amount of additive white noise.

In this paper, we propose a new method for the classification of voiced and unvoiced sounds using Average zero crossing index difference function which is a time domain technique. This paper is organized as follows. The description of the AZIDF is given in Section II. Voiced/Unvoiced classification using the new method is presented in Section III. In Section IV, the results and the discussion are presented. Finally, the conclusion is given in Section V.

II. Average Zero Crossing Index Difference Function

Average zero crossing index difference is a time domain function, capable of providing time domain information. The basic feature used in this function is zero crossings.

Consider, the speech signal x(n).The total number of samples present in the signal is N i.e. n=1,2…N. There may be M number of zero crossings. Zero crossing is the algebraic sign changes along the signal. Zero crossing index (ZCI) is the time index at which the signal crosses zero. The time value of each zero crossings present in the frame are considered as the zero crossing indices. The zero crossing indices are stored in the vector z.

$$z(m) = \underset{n}{arg}\left\{zero\left(sgn\left(x(n)\right) + sgn\left(x(n+1)\right)\right)\right\} \quad (1)$$

where,

n=1, 2…..N

m=1, 2….M

The difference between adjacent zero crossing index value is taken as the Zero crossing index difference (ZCID).

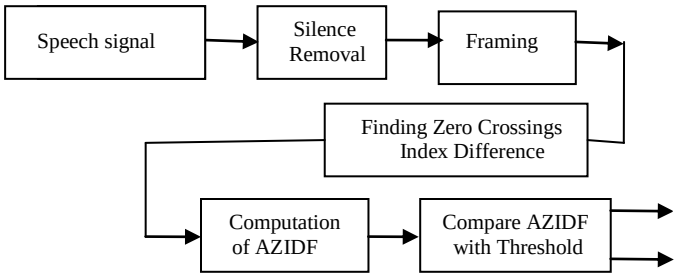

Fig. 1 Block diagram of proposed work

$$ZCID = z(m+1) - z(m) \quad (2)$$

AZIDF is measured by taking the averages of ZCID i.e. sum of zero crossing index difference to the length of zero crossing index difference

$$AZIDF = \frac{1}{M-1}\sum_{m=1}^{M-1}\left(z(m+1) - z(m)\right) \quad (3)$$

III. Voiced/Unvoiced Classification

The database used in this work is Berlin database. Totally there are 535 wave files. It is a German emotional database (Emo-DB). A block diagram representation of the algorithm is shown in Fig.1. First the silence region is removed from the speech signal.The parameters used to remove the silence region from the speech signal are Zero crossing rate and Energy.

A. Zero-Crossings Rate

ZCR is a measure of number of times in a given time interval, the amplitude of the speech signal passes through a value of zero.

$$ZCR = \frac{1}{N-1}\sum_{n=1}^{N-1}|sgn[x(n)] - sgn[x(n-1)]| \quad (4)$$

where,

$$sgn[x(n)] = \begin{cases} 1, x(n) \geq \\ -1, x(n) \end{cases} \quad (5)$$

N denotes number of samples in a frame

The number of zero crossings in the voiced part of the speech is high and the number of zero crossings in the unvoiced part speech is low. The number of zero crossings present in the silence region is very low.

B. Energy

Energy is defined as the strength of the signal. The amplitude of the signal varies with time. The energy of the speech signal provides a representation that reflects these amplitude variations.

$$En = \sum_{n=1}^{N-1}[x(n)w(n-m)]^2 \quad (6)$$

where,

N - Number of samples in a frame

w - Window used for analysis

The voiced part of the speech has high energy because of its city and the unvoiced part of speech has low energy [3]. The energy present in the silence region is very low. The very low zero crossings and the very low energy are used to detect and remove the silence region. The number of zero crossings present is almost equal to zero, and the energy of the signal will be close to zero.

The silence removed speech signal is divided into frames by 20 ms rectangular window without overlapping. In each frame the AZIDF is calculated and compared with a threshold in order to classify the voiced and unvoiced region. The threshold is set empirically by manual examination of AZIDF

in the voiced and unvoiced regions. The voiced/unvoiced classification performance of AZIDF is compared with the popular time domain parameter ZCR.

IV Results And Discussion

For analysis, 120 signals are taken from the Berlin database and Average Zero Crossing Index difference values for voiced and unvoiced region are manually calculated to set the threshold. When the threshold is greater than or equal to 5, the frame is considered as voiced frame otherwise an unvoiced frame. To compare the results with ZCR classification, the threshold is set manually for ZCR classification. When ZCR is less than 0.2 the frame is considered as voiced region otherwise an unvoiced frame.

TABLE I Performance Comparison of AZIDF and ZCR Classification

Actual classification	Number of frames	Experimental classification by		Correct classification by
		AZIDF	*ZCR*	
UV frames	156	UV	V	AZIDF
V frames	15	UV	V	ZCR
UV superimposed on V	59	UV	V	ZCR
V+UN,V<UN	5	UV	V	AZIDF
V+UN,V>UN	5	UV	V	ZCR

To calculate the error, comparison has been done between AZIDF and ZCR classification. Error calculation has been done manually. Total number of signals taken for analysis is 50. In that signals, contradiction occurred in 240 frames between AZIDF classification and ZCR classification. All the other frames are classified correctly and both the AZIDF and ZCR classification is same. Table I shows the performance comparison of AZIDF and ZCR classification in contradiction frames. Out of 240 frames, AZIDF classified 161 frames correctly and the ZCR classified the remaining frames correctly.

V. Conclusion

A new method for voiced/unvoiced classification has been proposed based Average zero crossing index difference function. The major advantage of this method is the source information is directly used for the process. The performance of AZIDF classification is compared with the ZCR classification. The comparison analysis between AZIDF and ZCR is done with manual examination. The results proved that the AZIDF performs better classification than ZCR in classifying unvoiced regions.

References

[1] A.E. Mahdi and E. Jafer (2008) "Two Feature Voiced/Unvoiced Classifier Using Wavelet Transform", The Open Electrical and Electronic Engineering Journal, pp, 8-13.

[2] A. P. Lobo and P. C. Loizou (2003) "Voiced/unvoiced speech discrimination in noise using Gabor atomic decomposition," in Proc. Int. Conf. Acoustics Speech and Signal Processing, Hong Kong, pp. I-820 – I-823.

[3] Alexandru Caruntu Gavril Toderean Alina Nica (2005) "Automatic Silence/Unvoiced/Voiced Classification of Speech Using a Modified Teager Energy Feature", WSEAS Int. Conf. on Dynamical Systems and Control, Venice, Italy, November 2-4.

[4] B. Atal and L. Rabiner (2003) "A Pattern Recognition Approach to Voiced Unvoiced-Silence Classification with Applications to Speech Recognition," IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. 24, No. 3, pp. 201-212.

[5] Darren K Emge and Tulay Adali (1999) "Least Relative Entropy for Voiced/Unvoiced Speech Classification**",** IEEE International Conference Vol.5, pp. 2976-2978.

[6] D. Arifianto, (2007) "Dual parameters for voiced-unvoiced speech signal determination," in Proc. Int. Conf. Acoustics Speech and Signal Processing, Honolulu, HI, pp. IV-749–IV-752.

[7] D. Arifianto, T. Kobayashi, (2005) "Voiced/unvoiced determination of speech signal in noisy environment using harmonicity measure based on instantaneous frequency", Proc. ICASSP, Vol. 1, pp.877-880, Philadelphia, USA.

[8] F. Beritelli, S. Casale, and S. Serrano (2007) "Adaptive V/UV speechdetection based on acoustic noise estimation and classification", Electronics Letters, vol. 43, no. 4, pp. 249–251.

[9] J.K.Shah, A. N. Iyer, B. Y. Smolenski and R. E. Yantorno (2004) "Robust Voiced/Unvoiced Classification Using Novel Features and Gaussian Mixture Model," IEEE International Conference on Acoustics, Speech, and Signal Processing, Philadelphia, pp. 17-21.

[10] Ji-Hyun Song (2009) "Efficient Implementation of Voiced/Unvoiced Sounds Classification Based on GMM for SMV Codec", IEICE Trans. Fundamentals Vol. E92-A, No.8 .pp.2121-2123.

[11] Mojtaba Radmard, Mahdi Hadavi, Mohammad Mahdi Nayebi (2011)" A New Method of Voiced/Unvoiced Classification Based on Clustering" Journal of Signal and Information Processing, pp. 336-347.

[12] N. Dhananjaya and B. Yegnanarayana (2010*)* "Voiced/Non-voiced Detection Based on Robustness of Voiced Epochs" IEEE Signal Processing Letters, Vol. 17, No. 3.

[13] R. G. Bachu, S. Kopparthi, B. Adapa and B. D. Barkana (2008) "Separation of Voiced and Unvoiced Using Zero Crossing Rate and Energy of the Speech Signal," American Society for Engineering Education (ASEE) Zone Conference Proceedings, pp. 1-7.

[14] S. Ahmadi and A. S. Spanias (2002) "Cepstrum Based Pitch Detection Using a New Statistical V/UV Classification Algorithm," IEEE Transactions on Speech and Audio Processing, Vol. 7, No. 3, pp. 333-338.

[15] Y.Qi and B.R.Hunt (2002) "Voiced-Unvoiced-Silence Classifications of Speech Using Hybrid Features and a Network Classifier," IEEE Transactions on Speech and Audio Processing, Vol. 1, No. 2, pp. 250-255.

[16] John G. Proakis and Dimitris G. Manolakis, Digital Signal Processing, 3rd edition, Prentice Hall, Inc., Englewood Cliffs, New Jersey, ISBN 0-13-373762-4, 1996.

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

Authors Profile

A.Milton received the **B.E.** degree in Electronics and Communication Engineering from the Govt. College of Engineering, Tirunelveli, Madurai Kamaraj University, Madurai, India, in 1993, the **M.Tech** degree in Microwave and Television Engineering from the College of Engineering Thiruvananthapuram, University of Kerala, India, in 2003. Currently working as assistant professor in St. Xavier's Catholic College of

Engineering, Nagercoil, India. His research interest includes digital speech signal processing and statistical pattern classification.

S.Ashitha Dayana received the **B.E.** degree in electronics and communication engineering from the Vins Christian College of Engineering, Nagercoil, Anna University, Chennai, India, in 2011.Currently doing **M.E.** in electronics and communication engineering (Applied electronics) in St.Xavier's Catholic College of Engineering,, Nagercoil, Anna University, Chennai,India. Her research interest includes digital speech signal processing and wireless communication.

S.Tamil Selvi received the **B.E.** degree in Electronics and Communication Engineering from Madurai Kamaraj University, Madurai, India, in 1988, the **M.E** degree from College of Engineering, Guindy, Anna University, Chennai, India, in 1997, the **Ph.D** degree from Manonmoniam Sundaranar University, Tirunelveli, India, in 2009. Currently working as professor in National Engineering College, Kovilpatti, India. Her research interest includes digital speech signal processing, image processing, wireless and optical communication. She has published 10 papers in international journals.