Voice and Unvoiced Classification Using Fuzzy Logic

Mohammed Algabri[1], **Mansour Alsulaiman**[2], **Ghulam Muhammad**[2], **Mohammed Zakariah**[2], **Mohamed Bencherif**[2],**Zulfiqar Ali**[2]
[1] Computer Science Department , King Saud University, Riyadh, Saudi Arabia
[2] Computer Engineering Department , King Saud University, Riyadh, Saudi Arabia
{malgabri, msuliman, ghulam, mzakariah, mbencherif, zuali}@ksu.edu.sa

Abstract - *In this paper, we proposed a system for automatic classification of speech. A speech signal contains three different regions voiced, unvoiced and silence. In the proposed system, Zero-Crossing rate and short term energy are used in a fuzzy logic control this classification. Arabic digits of the KSU database is used to test our proposed method. The proposed method achieves 2.5 % error between human classification and automatic classification using this method.*

Keywords: *Voice detection, fuzzy logic controller, zero-crossing, and short term energy.*

1 Introduction

A speech signal contains three different regions –voiced, unvoiced and silence. It is to determine in speech recognition system it is very helpful to discriminate between speech and background noise. This well give better result in a shorter time. The discrimination between voiced and unvoiced speech is also very helpful in designing high performance speech recognition system. Voiced phonemes are produced due to the vibration of the vocal folds, while unvoiced phonemes are made without vocal cord vibration [1]. Researchers have done significant efforts during the recent years for classification of speech into voice and unvoiced detection [2-12]. Based on the statistical and non-statistical techniques and pattern recognition approach a decision is made on the given segment of speech signal to classify it as a voice or unvoiced speech [6,7,8, and 9].The speech segment is classified into voice and unvoiced based on the acoustic features and pattern recognition techniques[12]. Speech is said to be intelligible if two third of it is voiced. Non periodic speech is called unvoiced, sounds like random, when consonants and spoken air is passed through a narrow constraints of the vocal tract causing unvoiced speech. Identification and extraction of voice speech is done because of its periodic nature. The Energy and Zero-Crossing rate are two important parameter are used to classify voice and unvoiced speech. They are used as front-end processing in automatic speech recognition system. The energy present in the signal spectrum and its frequency is indicated by the zero crossing counts. The low zero crossing count shows that the air is flowing in periodic form due to the extraction of vocal cords producing voiced speech [13]. Voice activity detector (VAD) is an algorithm implemented to detect the presence and absence of the speech. Numerous techniques are applied to the art of VAD. The very common features used in the detection process in the early VAD algorithm stage were short-time energy, zero-crossing rate and linear prediction coefficients [14]. Cepstral coefficients [15], spectral entropy [16], a least-square periodicity measure [17], wavelet transform coefficients [18] are examples of recently proposed VAD features. Because of the variety and varying nature of human speech and also because of the background noise none of the above technique proves to be perfect for all the applications. The decision to classify a segment into voiced or unvoiced is based on the values of the energy and zero-crossing. Since these values are not precise it will be helpful to use fuzzy logic. So in this paper, we present a method that classify the speech based on fuzzy logic of short-time energy and zero-crossing. This paper will be structured as follows: Section 2 presents the literature review. Section 3 gives the details of our proposed method. Sections 4 give the experimental results. Section 5 concludes the paper and suggests some future works.

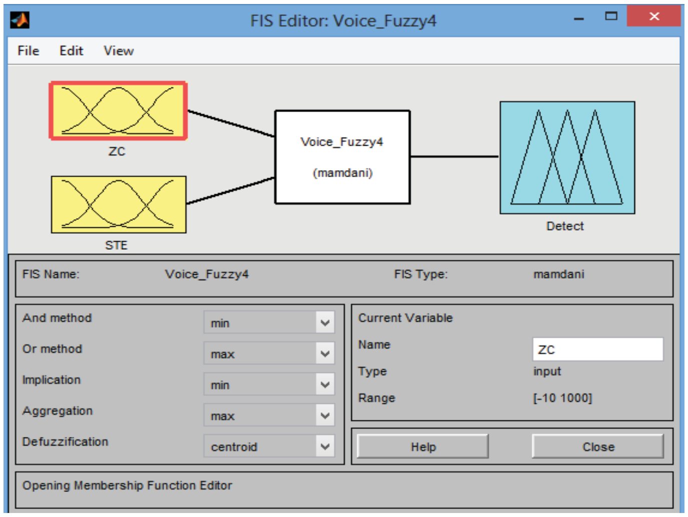
2 Literature Review

In [19] a speech recognition system based on zero-crossing rate and energy was presented. It used a vocabulary of ten Cantonese digits, and achieved a recognition accuracy of 97.2 percent has been achieved. Speech recognition using zero-crossing feature is presented in [20]. The zero-crossing features are extracted while speaking are done in the training phase, then stored in the database. Using the same technique the features for testing data are extracted and compared with the template in the database during the recognition phase. A VAD algorithm is presented in [21] for speech signal with very low signal to noise ration (SNR). The short-term energy of the speech signal is viewed as positive frequency of the magnitude spectrum of a minimum phase signal then the group delay function is computed for this signal. The speech regions of the signal are identified by well-defined peaks, while the non-speech regions are identified by well-defined valleys. Speech/ silence classification algorithm based on energy is proposed in [22]. The algorithm is able to track non-stationary signals and calculate the value of threshold using adaptive scaling parameter. Computed threshold can be obtained using maximum and minimum values of short-term energy. Voiced /Unvoiced classification based on clustering is developed in [23]. They used cepstral peak, zero crossing rate, and autocorrelation function peak of short time segments of speech by using some clustering methods. They achieved good results for classification of voice and unvoiced segments of speech. Zero crossing and short-term energy function are used for VAD algorithm for speech recognition applications in [24]. The method is labeling the speech samples based on if they are silence, voiced or unvoiced speech. Zero crossing rate and short term energy of speech are extensively used to detect the endpoints of an utterance.

3 Proposed method

Energy for unvoiced speech is significantly smaller than for voiced speech hence the short time energy can be used to distinguish voice and unvoiced speech. The short time energy can also be used to distinguish speech from silence. But the use of short time energy is not sufficient alone hence it is used coupled with the zero crossing in the classification of speech. Hence in [25] the authors present an algorithm for a heuristically developed method that cooperate zero crossing and energy. The authors states that "voice speech should characterized by relatively high energy and relatively low zero crossing rate, while unvoiced speech will have relatively high zero crossing rate and relatively low energy". They also state "we have not said what we mean by high and low values of short time zero crossing rate, and it is really not possible to be precise". Hence we see that this is a problem that fit to be solved by fuzzy logic.

In this paper, we propose a method to classify the speech into silence voiced and unvoiced detection using short time energy and zero-crossing in a fuzzy logic system. Figure 1 presents the fuzzy logic system, where the zero-crossing (ZC) and short term energy (STE) are the inputs of fuzzy logic control and the (Detect) is the output. The signal was segmented into frames with duration 10 ms. Then, hamming window was applied to prevent discontinuity. The mean of zero-crossing and short term energy was computed for each frame and set as inputs to fuzzy logic control. Voice, unvoiced and silence detection is an output of fuzzy logic control.



FIS Name: Voice_Fuzzy4

FIS Type: mamdani

And method: min

Or method: max

Implication: min

Aggregation: max

Defuzzification: centroid

Current Variable

Name: ZC

Type: input

Range: [-10 1000]

Opening Membership Function Editor

Figure 1: Fuzzy Logic System.

To define the membership function of each linguistic variables we used three membership functions for inputs (ZC and STE). The notation for Zero-crossing is Low, Mid and High as shown in Figure 2. The notation of short term energy (STE) id Low, Mid and High as shown in Figure 3. The notation of fuzzy output (Detect) is Sil, Unvoice and Voice as shown in Figure 4. The membership functions were tuned after many experiment manually to achieve good results.

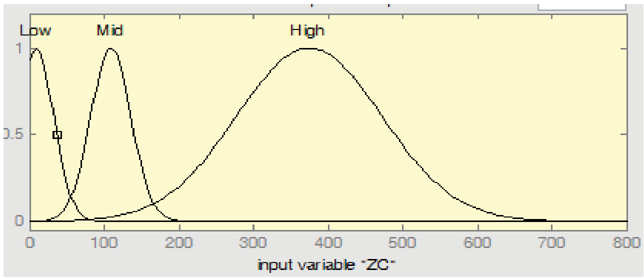


Figure 2: Membership Function of ZC.

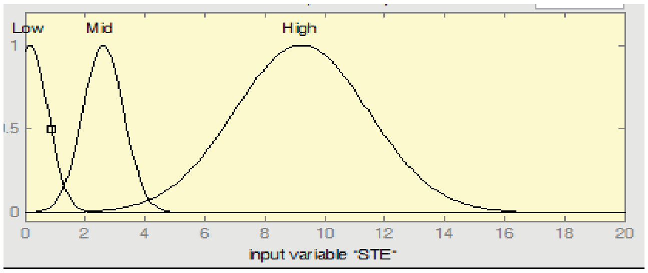


Figure 3: Membership Function of STE.

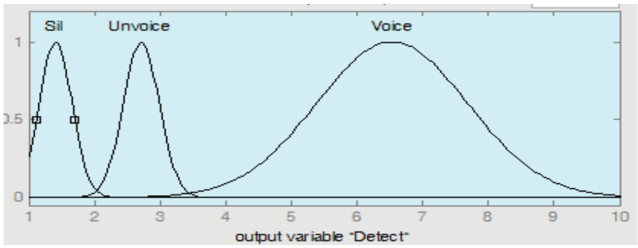
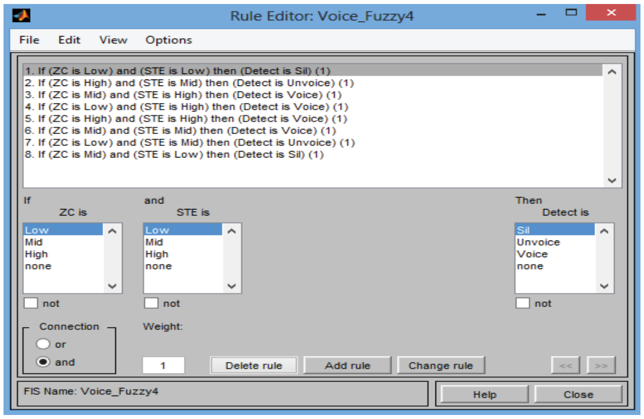


Figure 4: Membership Function of Detect.

The fuzzy rules of the control are defined in Figure 5. They were selected based on the study in [26].



1. If (ZC is Low) and (STE is Low) then (Detect is Sil) (1)
2. If (ZC is High) and (STE is Mid) then (Detect is Unvoice) (1)
3. If (ZC is Mid) and (STE is High) then (Detect is Voice) (1)
4. If (ZC is Low) and (STE is High) then (Detect is Voice) (1)
5. If (ZC is High) and (STE is High) then (Detect is Voice) (1)
6. If (ZC is Mid) and (STE is Mid) then (Detect is Voice) (1)
7. If (ZC is Low) and (STE is Mid) then (Detect is Unvoice) (1)
8. If (ZC is Mid) and (STE is Low) then (Detect is Sil) (1)

FIS Name: Voice_Fuzzy4

Figure 5: Rule based of fuzzy logic.

4 Experimental Results

The experimental results are obtained using the speech of Arabic digit of a King Saud Speech Database [27]. The database has 327 speakers and is rich in different aspects and different nationalities. We performed many experiments to segment the Arabic digits. The classification result was excellent. As an example Figure 6 shows the original wav file that contains the utterances of Arabic digits from zero to nine. We applied the proposed method on this file to classify it into voiced, unvoiced and silence. The output of our proposed method is shown in Figure 7.

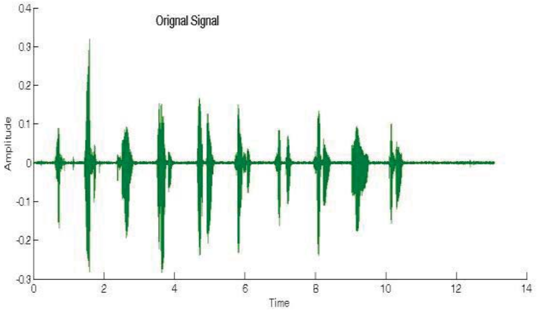


Figure 6: Original wav file.

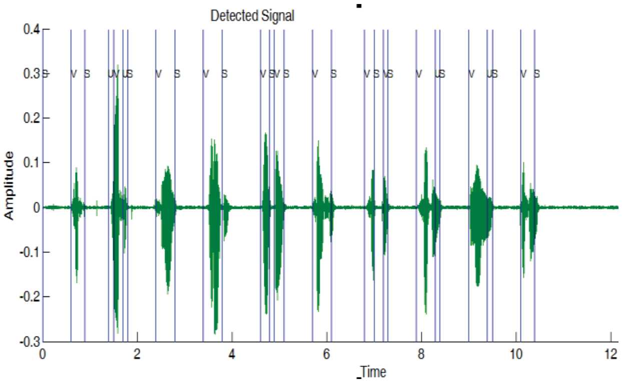


Figure 7: Segmented wav file [S-Silence-V-Voice, U-Unvoiced].

In order to evaluate the proposed method for voice and unvoiced detection using fuzzy logic, we used the error difference between voice detected by human and voice detected using this method. The error was calculated using equation (1).

$$Error = \frac{|TH - TC|}{TH} \times 100 \qquad (1)$$

Where TH is the manually voice frame length detected by human as shown in table 1, and TC is the length of voice detected using the proposed approach.

Table 1: Automatic and manual voice segmentation length of digit file.

	Voice frame length detected (ms)												total
TH	0.262	0.26	0.416	0.43	0.19	0.21	0.397	0.17	0.097	0.416	0.445	0.33	3.623
TC	0.3	0.2	0.4	0.4	0.2	0.2	0.4	0.2	0.1	0.4	0.4	0.3	3.5
TH-TC	-0.038	0.06	0.016	0.03	-0.01	0.01	-0.003	-0.03	-0.003	0.016	0.045	0.03	0.093
Error	14.50	23.08	3.85	6.98	5.26	4.76	0.76	17.65	3.09	3.85	10.11	9.09	**2.57**

So from the results in table 1, we can calculated the overall detection error using equation in (2).

$$OverallError = \frac{\sum_{i=1}^{n} |TH - TC|}{\sum_{i=1}^{n} TH} \times 100 \qquad (2)$$

Where n is a number of voice frames detected. So from the table 1 above the overall error segmentation is approximate **2.5%**.

As another example, Figure 8 shows the original wav file that contains the speech of ten phonetic distinctive words in Arabic language. We applied the proposed method on this file to classify it into voiced, unvoiced and silence. The output of our proposed method is shown in Figure 9.

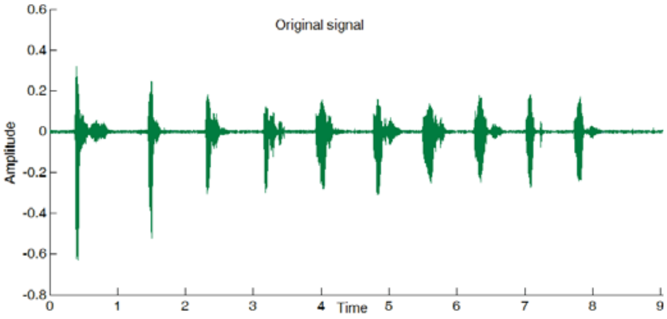


Figure 8: Original wav file.

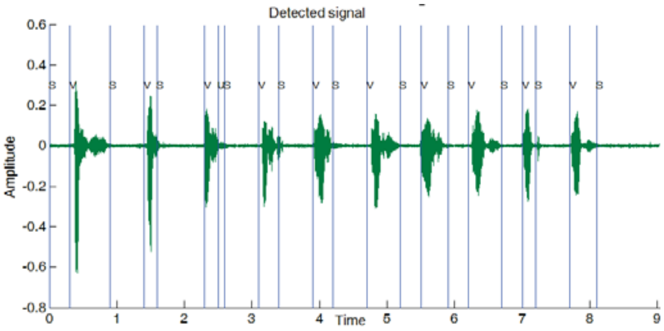
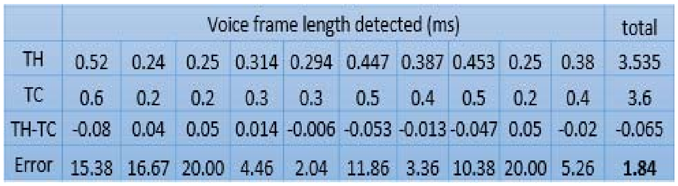


Figure 9: Segmented wav file [S-Silence-V-Voice, U-Unvoiced].

The calculated error based on equations (1, 2) between human segmented and our proposed method shown in Table 2. The overall error segmentation is approximate **1.84%.**

Table 2: Automatic and manual voice segmentation length of words file.



	Voice frame length detected (ms)										total
TH	0.52	0.24	0.25	0.314	0.294	0.447	0.387	0.453	0.25	0.38	3.535
TC	0.6	0.2	0.2	0.3	0.3	0.5	0.4	0.5	0.2	0.4	3.6
TH-TC	-0.08	0.04	0.05	0.014	-0.006	-0.053	-0.013	-0.047	0.05	-0.02	-0.065
Error	15.38	16.67	20.00	4.46	2.04	11.86	3.36	10.38	20.00	5.26	**1.84**

5 Conclusion

Silence/unvoiced/voiced classification using fuzzy logic of Arabic speech was proposed in this paper. Zero crossing and short term energy were used as features. Fuzzy logic controller used to distinguish three categorize of speech (Silence, Voice and unvoiced). The experiments were performed on Arabic KSU database in MATALB environment and their results showed that the proposed method successfully classified the speech.

Acknowledgment

This project was funded by the National Plan for Science, Technology and Innovation (MAARIFAH), King Abdulaziz City for Science and Technology, Kingdom of Saudi Arabia, Award Number (12-MED2474-02).

References

[1] D. Jurafsky, J. H. Martin "Speech and Language Processing", Publisher Pearson.

[2] E. Fisher, J. Tabrikian and S. Dubnov, "Generalized Likelihood Ratio Test for Voiced-Unvoiced Decision in Noisy Speech Using the Harmonic Model," IEEE Trans-actions on Audio, Speech, and Language Processing, Vol. 14, No. 2, 2006, pp. 502-510. doi:10.1109/TSA.2005.857806

[3] Y. Qi and B. R. Hunt, "Voiced-Unvoiced-Silence Classifications of Speech Using Hybrid Features and a Network Classifier," IEEE Transactions on Speech and Audio Processing, Vol. 1, No. 2, 2002, pp. 250-255. doi:10.1109/89.222883

[4] B. Atal and L. Rabiner, "A Pattern Recognition Approach to Voicedunvoiced-Silence Classification with Applica-tions to Speech Recognition," IEEE Transactions on Ac- oustics, Speech and Signal Processing, Vol. 24, No. 3, 2003, pp. 201-212. doi:10.1109/TASSP.1976.1162800

[5] F. Y. Qi and C. C. Bao, "A Method for Voiced/Unvoiced/Silence Classification of Speech with Noise Using SVM," Acta Electronica Sinica, Vol. 34, No. 4, 2006, pp. 605-611.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

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

[11] 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.

[12] 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.

[13] JaberMarvan, "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.

[14] B. S. Atal and L. R. Rabiner, "A pattern recognition approach to voiced-unvoiced- silence classi_cation with applications to speech recognition, " IEEE Trans. Acoustics, Speech, Signal Processing, vol. 24, pp. 201-212, June 1976.

[15] J. A. Haigh and J. S. Mason, "Robust voice activity detection using cepstralfea-tures, " in Proc. of IEEE Region 10 Annual Conf. Speech and Image Technologies for Computing and Telecommunications, (Beijing), pp. 321-324, Oct. 1993.

[16] S. A. McClellan and J. D. Gibson, "Spectral entropy: An alternative indicator for rate allocation, " in IEEE Int. Conf. on Acoustics, Speech, Signal Processing, (Adelaide, Australia), pp. 201-204, Apr. 1994.

[17] R. Tucker, "Voice activity detection using a periodicity measure, "IEE Proc.-I, vol. 139, pp. 377-380, Aug. 1992.

[18] J. Stegmann and G. Schroder, "Robust voice-activity detection based on the wavelet transform, " in Proc. IEEE Workshop on Speech Coding for Telecommunications, (Pocono Manor, PN), pp. 99-100, Sept. 1997.

[19] Lau, Yiu-Kei, and Chok-Ki Chan. "Speech recognition based on zero crossing rate and energy." IEEE transactions on acoustics, speech, and signal processing 33, no. 1 (1985): 320-323.

[20] Aye, Yin. "Speech Recognition Using Zero-Crossing Features." InElectronic Computer Technology, 2009 International Conference on, pp. 689-692. IEEE, 2009.

[21] Hari Krishnan P, S., R. Padmanabhan, and Hema A. Murthy. "Robust voice activity detection using group delay functions." In Industrial Technology, 2006. ICIT 2006. IEEE International Conference on, pp. 2603-2607. IEEE, 2006.

[22] Sakhnov, Kirill, Ekaterina Verteletskaya, and Boris Simak. "Approach for energy-based voice detector with adaptive scaling factor." IAENG International Journal of Computer Science 36, no. 4 (2009): 394.

[23] Radmard, Mojtaba, Mahdi Hadavi, and Mohammad Mahdi Nayebi. "A New Method of Voiced/Unvoiced Classification Based on Clustering." Journal of Signal and Information Processing 2, no. 04 (2011): 336.

[24] Lokhande, N. N., D. S. Nehe, and P. S. Vikhe. "Voice activity detection algorithm for speech recognition applications." In IJCA Proceedings on International Conference in Computational Intelligence (ICCIA2012), vol. iccia, no. 6, pp. 1-4. 2012.

[25] Rabiner, Lawrence R., and Ronald W. Schafer. "Theory and application of digital speech processing.", Person , 2011.

[26] Greenwood, M., & Kinghorn, A. (1999). Suving: Automatic silence/unvoiced/voiced classification of speech. Undergraduate Coursework, Department of Computer Science, the University of Sheffield, UK.

[27] Alsulaiman, M., Ali, Z., Muhammed, G., Bencherif, M., & Mahmood, A. (2013). KSU Speech Database: Text Selection, Recording and Verification. In Modelling Symposium (EMS), 2013 European (pp. 237-242). IEEE.