Jurnal
Teknologi

Full Paper

EFFECTIVE ARABIC SPEECH SEGMENTATION STRATEGY

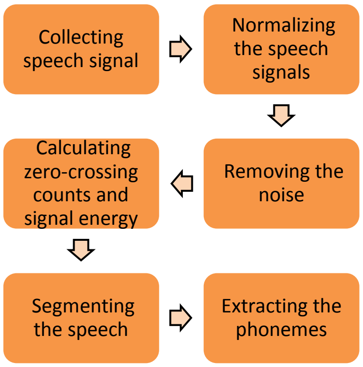
Article history
Received
25 November 2014
Received in revised form
7 April 2015 2015
Accepted
1 October 2015

Abduljalil Radman[a,c]*, Nasharuddin Zainal[a], Cila Umat[b], Badrulzaman Abdul Hamid[b]

[a]Department of Electrical, Electronic & Systems Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Malaysia
[b]School of Rehabilitation Sciences, Faculty of Health Sciences, Universiti Kebangsaan Malaysia, 50300 Kuala Lumpur, Malaysia
[c]Department of Communication and Computer Engineering, Faculty of Engineering and Information Technology, Taiz University, Yemen

*Corresponding author
abdu_rad@yahoo.com

Graphical abstract



Abstract

Speech segmentation is a process to segment speech utterances into small chunks, where each chunk represents a phoneme. The phoneme is an essential unit in any speech which is recognizable. In this paper, a segmentation speech approach was proposed to segment consonant and vowel phonemes from speech utterances of Arabic basic syllables, in order to analyze the consonant production of a group of Malay-speaking normal hearing children and adults. The approach is a combination of zero-crossing counts and signal energy. The zero-crossing counts were used to extract the noise signals, whilst the signal energy was utilized for identifying the speech signals. The spectrogram was used to determine the frequencies with the most intense energy.

Keywords: Component speech segmentation, Zero-crossing counts, Signal energy, consonants

Abstrak

Segmentasi pertuturan merupakan suatu proses untuk mengasingkan bunyi pertuturan kepada bahagian-bahagian yang lebih kecil di mana setiap bahagian tersebut mewakili satu fonem. Fonem merupakan unit asas dalam bunyi pertuturan yang boleh dikenal pasti. Dalam kajian ini, pendekatan segmentasi pertuturan diperkenalkan untuk memisahkan fonem konsonan dan vokal dari bunyi pertuturan suku kata asas Bahasa Arab bagi tujuan menganalisa penghasilan bunyi konsonan oleh kanak-kanak dan dewasa yang merupakan penutur natif Bahasa Melayu. Pendekatan ini merupakan kombinasi pengiraan silangan-sifar dan tenaga signal. Pengiraan silangan-sifar digunakan untuk mengasingkan kebisingan pada signal sementara tenaga signal digunakan untuk mengenal pasti bunyi pertuturan yang dikehendaki. Spektrogram digunakan untuk menentukan frekuensi dengan tenaga bunyi signal yang paling tinggi.

Kata kunci: Komponen segmentasi pertuturan, pengiraan silangan-sifar, tenaga signal, konsonan

1.0 INTRODUCTION

Arabic speech segmentation is a process to partition the original continuous audio waves into segmented waves that carry meaning and consist of consonants and vowels. There are rules that control the existence of consonants and vowels in Arabic speechs such as, there is no transition from vowel to vowel and there is always one of the six patterns of CV, CV:, CVC, CVCC, CV:C, CV:CC [1], where C denotes a consonant, V denotes a vowel and V: denotes a long vowel. The Arabic basic syllables consist of 29 types of consonants and vowels that have different voicing, place and manner of production. In the literature, several methods have been proposed to segment speech for further processing [2-3]. Almisreb *et al.* [4] also proposed a segmentation technique for Arabic letter spoken by Malay speakers. First, they applied a multiscale principal component to remove the noise. Then, the Zero-Crossing Rate function was used to segment the voice signals. In other approach [5], the authors considered the silence as the main factor for speech segmentation. Mousmita and Sarma [6] also proposed a soft computing framework for phoneme segmentation used for speech recognition.

Indeed, there are a lot of variations in the production of Arabic basic syllables by native Arabic speakers and non-native speakers. However, non-native speakers who read the holy Quran which is in Arabic have to master the pronunciation of these basic syllables before being able to read the Quran with proper recitation. Therefore in this study, we aimed to quantify the variability in the production of these basic Arabic syllables among Malay-speaking normal hearing children and adults by performing acoustic phonetic analyses. In order to be objective in segmenting the speech signals which contain consonants and vowels and purifying the samples by separating the noise from the speech energy, an effective Arabic speech segmentation method had been proposed.

The described method in this study focused on segmenting speech to define the start and end points of each phoneme in the speech signal. Thus, samples energy and zero-crossing counts were calculated to be used for distinguishing between speech and non-speech segments. The calculation of zero-crossing counts and signal energy has been widely used for speech segmentation such as that reported in [7-8]. However, the proposed method in this paper differs from the commonly used speech segmentation using zero-crossing counts and signal energy by normalizing and smoothing the speech signals before calculating the zero-crossing counts and signal energy. This modification made the speech segmentation more accurate, as well as more robust to noise.

The remaining of the paper is organized as follows. Section 2 discusses the speech segmentation process. Section 3 describes the proposed speech segmentation method. Section 4 introduces the data collection, and Section 5 discusses the experimental results. In Section 6, the conclusion is presented.

2.0 SPEECH SEGMENTATION

Speech segmentation is a process to segment the speech utterances into small chunks in which each chunk represents one phoneme. The phoneme is the essential unit in any speech; it is meaningful and recognizable. Several parameters have an effect on the speech signal such as the style of the speaking, the language, phoneme inventory, and the speaker's origin. Hence, all of these parameters have an impact on the speech segmentation process [9-11]. However, the idea of the speech segmentation process is to detect the start and end of each of the phoneme in the speech utterance of Consonant-Vowel (CV) format.

3.0 PROPOSED SPEECH SEGMENTATION METHOD

In this paper, a simple and efficient speech segmentation method to segment Arabic basic syllables was proposed. The speech samples used to evaluate the proposed method were obtained from eight (8) Malay-speaking, highly trained Quranic reciters with 3 to 30 years experiences in teaching Al-Quran at an institution of higher learning for Quranic recitation in Malaysia. Moreover, speech samples from 25 Malaysian children that have the ability to identify Al-Quran phonemes were obtained. The proposed speech segmentation method in this paper used the zero-crossing counts and signal energy to detect the start and end of each of the phoneme in the speech utterance.

First, each speech signal was divided into small frames (80 samples/frame). Then, the zero-crossing counts and signal energy for each frame were calculated in order to be used to isolate the phonemes form the speech signal. However, the speech signals for different subjects are known not to be within the same amplitude range, because the different loudness cues. Thus, a preprocessing process was applied to normalize the speech signals into the same amplitude range. After that, the ambient noise was removed by filtering the normalized speech signal with a 1-D Gaussian filter; this process was called smoothing. This operation allowed the value of each sample more in tune with the values of its neighbors. Additionally, we assumed that each phoneme must reach certain milliseconds to be counted as a speech, so as to avoid segmenting the noise as a speech signal. The flow chart in Figure 1 shows the major processes to segment the speech signals by the proposed method.

As described above, a combination of zero-crossing counts and signal energy were utilized to segment the speech signals. However, the zero-crossing counts were used to detect the noise signals, whilst the signal energy was utilized to identify the speech signals. Based on this method, the targeted phonemes were found to have a low number of zero-crossing counts but high total energy. In contrast, noise signals were found to have a high number of zero-crossing counts, but lower total energy as compared to the targeted phonemes.

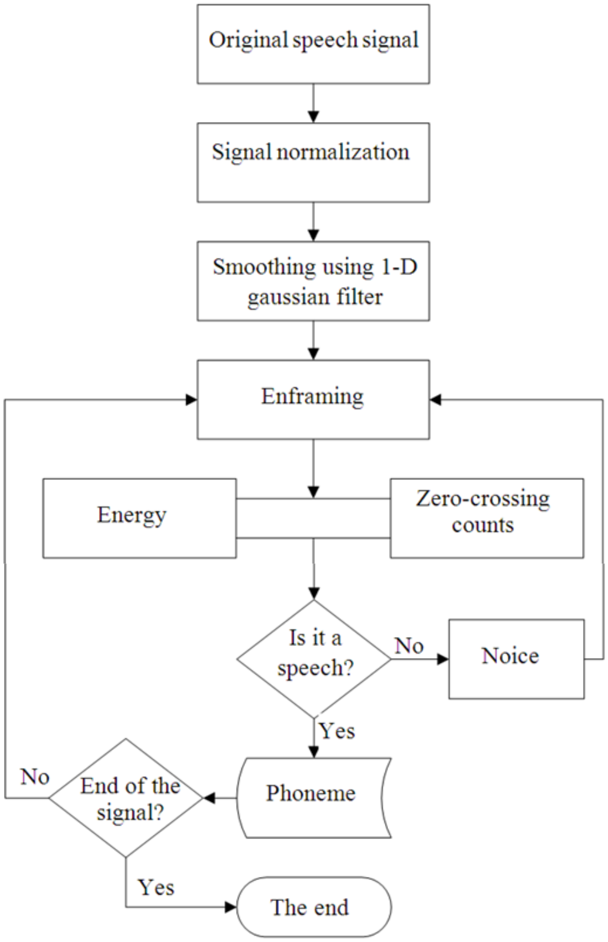


Figure 1 Flow chart of the proposed speech segmentation method

4.0 DATA COLLECTION

The data used to validate the accuracy of the proposed speech segmentation method in this paper were collected at a public institution of Al-Quran higher learning Darul Quran (DQ) in Kuala Kubu Baru-Selangor, and at the Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia (UKM), Bangi-Selangor. The first set of data was collected from eight adult subjects from Darul Quran, who were native Malay speakers aged between 25 to 40 years old with teaching experience ranged from 3 to 30 years. All of them had formal education in Al-Quran learning from various Middle East higher learning institutions. The second set of speech data was collected from 25 Malay-speaking, normal hearing children aged between 6 to 16 years old. All pediatric subjects had the ability to identify the basic Al-Quran syllables and able to recite the Quran. Hearing screening and oro-motor examinations were performed prior to the speech recording session, in order to ensure normal hearing and normal speech motor structures for accurate speech production. These procedures were performed by qualified audiologists and speech language pathologists for all subjects' adults and children.

Subjects' task was to produce sequentially the basic Al-Quran syllables (أ, ب, ت, ث, ج, ح, خ, د, ذ, ر, ز, س , ش , ص , ض, ط, ظ, ع, غ , ف , ق, ك, ل, م, ن, ه, و, ي, ء) in the CV-format. For all consonants, the vowel attached to them was /a/. Subjects were informed to have a pause in between each syllable to facilitate the analyses. Three repetitions were made to ensure data was reliable and average values could be used for the later analyses. At both study sites (DQ and UKM), speech samples had been recorded in fairly quiet rooms with the fixed microphone position of the mouth of about 20 cm.

5.0 RESULTS AND DISCUSSION

All collected speech signals in this study were segmented by the proposed speech segmentation method described in Section 3, so as to isolate each phoneme separately. Results revealed that the proposed method performs well even with speech signals that contain ambient noise. Figure 2 shows an example of correct speech segmentation obtained using the proposed method. As an example for phoneme isolation, Figure 3 presents the segmented waveform of phoneme 'ق' attained by the proposed method. However, in each speech signal, 87 phonemes were detected. In the segmented speech, there were three samples for each phoneme (29 arabic phonemes).

Due to the overlapping for some successive phonemes in some speech signals, especially those produced by children subjects, the proposed method failed to isolate each phoneme separately. Thus, phonetic analysis software, Praat [12] was used to allow a small period of time between the overlapped phonemes before we applied our method to segment the speech signals. On the contrary, the clear speech signals collected from adult subjects were effectively recognized from noise through calculating the zero-crossing counts and signal energy. The noise signals were found to have low energy and zero-crossing counts. As a results, the proposed method was able to successfully identify the start and end of different phonemes analyzed (Figures 2 and 3). Even though, automatic speech segmentation is a challenging task [13], the proposed method proved that it is sufficient for segmenting speech signals with a moderate background noise as those collected in this study.

After segmentation, all the segmented phonemes were analyzed to study the features of the different type of phonemes. However, we found that all passive phonemes such as ('ب','ت','د','ض','ط','ق','ك') could be recognized by the relatively high energy for a short duration at the beginning; whereas , fricative phonemes such as ('ف','ث','د','ص','ز') could be identified by high frequency spectrum with low energy.

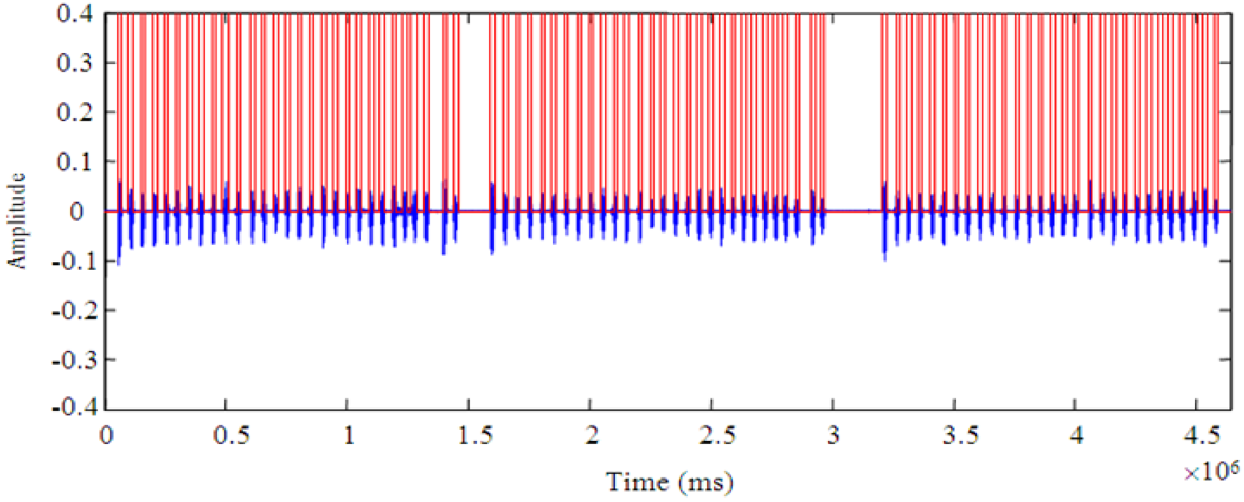


Figure 2 An example of correct speech segmentation by the proposed method

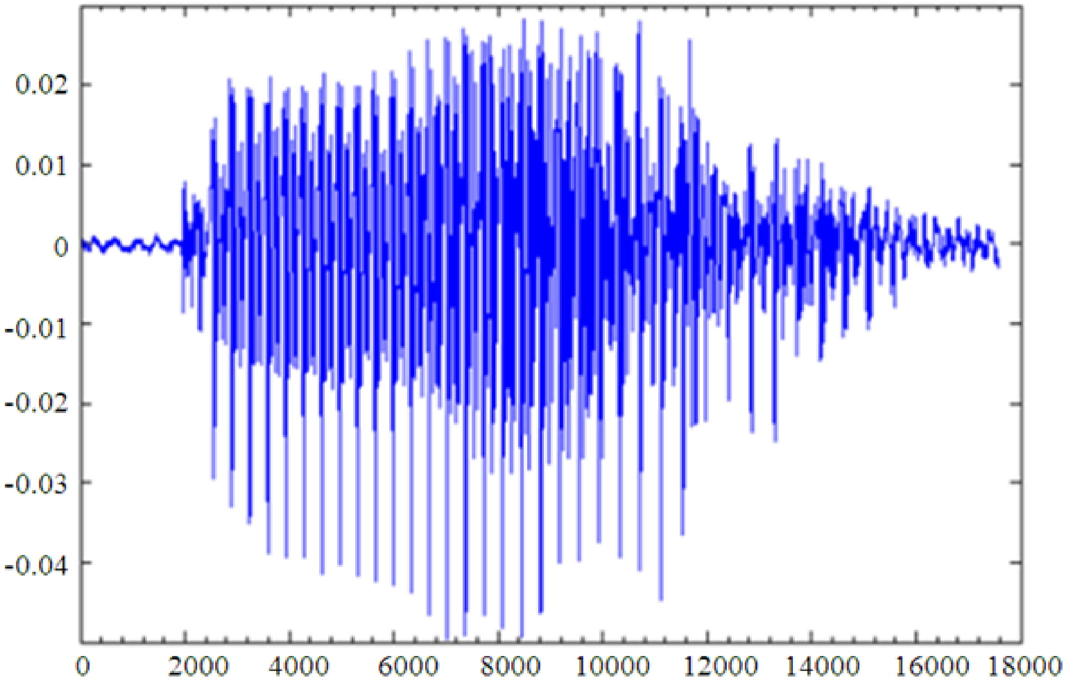


Figure 3 Waveform of phoneme 'ق' extracted from the segmented speech signal

The segmented phonemes were also analyzed to obtain the frequency spectrum of each phoneme; the frequency areas with highest energy for each phoneme were extracted from the frequency spectrum by means of the Fast Fourier Transform (FFT) analyses. Next, formant frequencies or frequencies with the highest or most intense energy were utilized to identify the speech of each phoneme.

6.0 CONCLUSION

Segmenting speech in order to analyze the phoneme production represents a challenging task, as it commonly involves subjective judgment of the naked eyes. In this paper, a speech segmentation method was proposed. The proposed method was based on the calculation of zero-crossing counts and signal energy. As a preprocessing step, the speech samples were normalized to quantify all speech signals into the same amplitude range. Then, the samples were smoothed by means of Gaussian filter in order to eliminate the noise signals. Experimental results showed that the proposed method automatically segment the speech signals into their basic phonemes. The start and end points for each of the Arabic phonemes in the speech signals were clearly identified. Hence, the clearly separated targeted

Arabic phonemes allow further analyses on the main frequencies and intensities of the phonemes as produced by Malay-speaking normal hearing listeners-children and adults.

Acknowledgement

The researcher would like to thank the audiologists and speech language pathologists who helped with the hearing screening and oro-motor examination on all subjects prior to the speech data collection. We also thank Darul Quran and the children and their parents for their participation in this study. The study was funded by the UKM Research Grant Scheme UKM-GUP-2011-115.

References

[1] God, A. M. 1999. Speech Processing Using Wavelet Based Algorithms. Ph.D. Thesis, Cairo University.

[2] Alginahi, Yasser M. 2013. A Survey on Arabic Character Segmentation. *International Journal on Document Analysis and Recognition (IJDAR)*. 16(2): 105-126.

[3] Al-Manie, Mohammed A., Mohammed I. Alkanhal, and Mansour M. Al-Ghamdi. 2010. Arabic Speech Segmentation: Automatic Verses Manual Method and Zero Crossing Measurements. Indian Journal of Science and Technology. 3(12): 1134-1138.

[4] Almisreb, Ali Abd, Ahmad Farid Abidin, and Nooritawati Md Tahir. 2013. Segmentation of Arabic Letters Signal using Multiscale Principal Component analysis and Zero-Crossing Rate based on Malay speakers. *IEEE International Conference in Control System, Computing and Engineering (ICCSCE)*. 483-486.

[5] Abushariah, Mohammad Abd-Alrahman Mahmoud, Raja Noor Ainon, Roziati Zainuddin, Moustafa Elshafei, and Othman Omran Khalifa. 2012. Arabic Speaker-Independent Continuous Automatic Speech Recognition Based on a Phonetically Rich and Balanced Speech Corpus. *Int. Arab J. Inf. Technol*. 9(1): 84-93.

[6] Sarma, Mousmita, and Kandarpa Kumar Sarma. 2014. Phoneme-based Speech Segmentation Using Hybrid Soft Computing Framework. *Computational Intelligence and Complexity*. New Delhi: Springer.

[7] AL-Haddad, S. A. R., Salina Abdul Samad, Aini Hussein, K. A. Ishak and A. A. Azid *et al.* 2006. Automatic Segmentation and Labeling for Malay Speech Recognition. *Proceedings of the 6th WSEAS International Conference on Signal Processing, Computational Geometry & Artificial Vision, (GAV' 06), Stevens Point, Wisconsin, USA*. 217-221.

[8] Salam, M. S., D. Mohamad and S. H. Salleh. 2010. Speech Segmentation Using Divergence Algorithm with Zero Crossing Property. *Proceedings of the 13th International Conference on 13th International Conference on Computer and Information Technology, Dec. 23-25, IEEE Xplore Press, Dhaka.* 488-493.

[9] Anwar, M. J., M. M. Awais, S. Masud and S. Shamail. 2006. Automatic Arabic Speech Segmentation System. *Int. J. Inform. Technol.* 12: 102-111.

[10] Abdul-Kadir, N. A., R. Sudirman and N. M. Safri. 2010. Modelling of the Arabic Plosive Consonants Characteristics Based on Spectrogram. *Proceedings of the 4th Asia International Conference on Mathematical/Analytical Modelling and Computer Simulation, May. 26-28, Kota Kinabalu, Malaysia*. 282-285. DOI: 10.1109/AMS.2010.63.

[11] Abdul-Kadir, N. A. and R. Sudirman. 2011. Difficulties of Standard Arabic Phonemes Spoken by Non-Arab Primary School Children based on Formant Frequencies. *J. Comp. Sci.* 7: 1003-1010. DOI: 10.3844/jcssp.2011.1003.1010.

[12] Paul, B. Praat. 2001. A System for Doing Phonetics by Computer. *Glot Int.* 5: 341-345. http://www.researchgate.net/publication/208032992_Praat_a_system_for_doing_phonetics_by_computer.

[13] Tolba, M. F., T. N. Azmy, A. A. Abdelhamid and M. E. Gadallah. 2005. A Novel Method for Arabic Consonant/Vowel Segmentation Using Wavelet Transform. *IJICIS*. 5: 353-364.