Semi-Automatic Segmentation System for Syllables Extraction from Continuous Arabic Audio Signal

Mohamed S. Abdo

Systems and Biomedical Engineering Department,
Faculty of Engineering, Minia University
Minia, Egypt

Ahmed H. Kandil

Systems and Biomedical Engineering Department,
Faculty of Engineering, Cairo University,
Cairo, Egypt

Abstract—**The paper describes a speaker independent segmentation system for breaking Arabic uttered sentences into its constituent syllables. The goal is to construct a database of acoustical Arabic syllables as a step towards a syllable-based Arabic speech verification/recognition system. The proposed technique segments the utterances based on maxima extraction from delta function of 1ˢᵗ MFC coefficient. This method locates syllables boundaries by applying the template matching technique with reference utterances. The system was applied over a data set of 276 utterances to segment them into their 2544 constituent syllables. A segmentation success rate of about 91.5% was reached.**

Keywords—Arabic speech syllables; automatic segmentation; boundaries detection; delta-MFCC features

I. INTRODUCTION

Speech and natural language processing (SNLP) is a vital topic in recent research. Computer Aided Language Learning (CALL) systems have received considerable attention in recent years. CALL system are used to improve learning and to evaluate pronunciation quality of speakers. Arabic is the spoken language in 60 countries around the world, so it is the second most spoken language in terms of the number of speakers [1]. Quran is the basic reference of Arabic language. One of the most important issues in the Arabic world is the learning of Quran recitations [2].

A robust language learning system should have a vocabulary database in order to recognize uttered speech, localize and identify pronunciation mistakes and provide meaningful feedback to help users to improve their performance. Building an acoustical Arabic database of syllables, phonemes, etc., according with the requirements of the used application is the base for these researches. In this paper, a new method for the automatic segmentation of Arabic audio signal into its syllables is introduced. Classical Arabic is the old form of the Arabic language used in the Quran. Modern Standard Arabic "MSA" is a formal language commonly used in all Arabic-speaking countries.

Our system seeks to perform accurate allocation of syllables boundaries from continuous speech as a step towards building an Arabic database that contributes in developing many applications, such as:

a) Diagnosis and treatment of speaking pathology.

b) Teaching the recitation rules of the Holy Quran.

c) Training systems for correct Arabic pronunciation for children and non-native speakers.

d) Facilitate the man-machine communication and help its progress.

This paper is organized as follows; section 2 presents segmentation. Section 3 defines the Arabic syllable segmentation. Section 4 introduces the implementation of the proposed system. Section 5 discusses the segmentation experimental results. In section 6, conclusion and future work are presented.

II. RELATED WORK

A. Selecting a Template

Due to the importance of the subject, intensive studies have been conducted on speech segmentation employing different features. A theoretical framework for MSA speech segmentation using dynamic level building was introduced byEl Arif et al. [3]. Gody presented a speech segmentation approach depending on wavelet transform and spectral analysis; the accuracy was 95% [4],. Tolba used wavelet transform achieving 81% accuracy [5]. Yacine Yekache et al. reported a step toward developing Quranic reader using sphinx4 framework [6]. Wang et al. [7], Fu et al. [8] introduced zero crossing rate "ZCR", pitch and energy profile as features for the segmentation of speech. In [9], a survey on Punjabi speech segmentation into syllables is presented using negative derivative of Fourier transformations. . In [10], a syllable based recognition system based on pseudo articulatory method is presented which contributes of more plausible style of speech recognition and new modeling of speech behavior. In [11], a group delay based approach is proposed which the short-term energy is processed for determining segment boundaries. An attempt is made by Sarada et al. [12] to automate the syllable transcription task for Indian languages. The method does not require any manual segmentation and a new feature extraction strategy is explored using multiple frame sizes and rates for both training and testing datasets. A technique based on short term energy was implemented in [13] for the automatic segmentation of speech signals in Punjabi speech into syllables. In [14], biologically inspired auditory attention cues are proposed for syllables segmentation from continuous speech. The method achieved 92.1 % accuracy of syllable boundary detection at frame level when tested on TIMIT. In [15], a time-frequency representation and the fusion of intensity and voicing

measures were introduced for the segmentation of speech into syllables. A practical method for blind segmentation of continuous speech is presented by Villing et al. [16] using amplitude onset velocity and coarse spectral makeup to identify syllable boundaries. Mijanur Rahman et al. [17] developed a system that automatically segments words from the continuously spoken Bangla sentences. Mijanur Rahman et al. [17] developed a system that automatically segments words from the continuously spoken Bangla sentences. Our prior works of [18], [19] presented an algorithm for segmenting a subset of emphatic and non-emphatic sounds automatically from continuous spoken Arabic, where achieved a segmentation accuracy of up to 90 %.

III. THE PROPOSED SEGMENTATION SYSTEM

A. Arabic syllables

Speech units can be phonemes, letters, syllables, words, etc. The segmentation problem may be viewed as an unlabeled splitting problem where the input sequence needs to be split into subsequences. This study focuses on the isolation of syllables; the syllable consists of nuclear vowel plus neighboring consonants. The vowel should be preceded by a consonant and followed by zero, one or two consonants. Thus Arabic language has five standard types of syllables: {CV, CV %, CVC, CV%C, CVCC} where "C": consonant, "V": vowel, "V%": long vowel.

B. System block diagrm

The block diagram of the proposed system is shown in Fig. 1.

It consists of five modules developed to allocate the syllables boundaries in Arabic utterance

1) Data collection, preprocessing, feature extraction and forming reference template
2) Test data entry and preprocessing.
3) Features Extraction
4) Automatic allocation of syllables boundaries through matching process.
5) Evaluation of the resulting isolated syllables.

Our system is based on the Holy Quran. The recordings of twelve speakers; each recited 23 continuous sentences (verses), constitute a dataset of 276 utterances. The texts of the collected data with its IPA mapping for according to [20] are reported in table 4 which includes (2544) syllables to be detected and segmented. The recordings from the reader "Mahmoud Khaleel El-Hosary" has been selected to form the template dataset which will be used as reference throughout the matching process, this choice is based on the well-known of his good realization for the rules of recitations of Quran verses.

Wide variability of speech may affect the accuracy of its analysis. So, good setup of pre-processing phase improves the performance of speech segmentation. The audio signal is divided into fixed length frames with overlapping to insure continuity [21]. MIR© software toolbox [22] is used for applying the pre-processing steps, which are: trim silence at begnning and end of the audio signal, normalize the recorded

data, form fixed length frames of 30 ms with 60% overlapping and smooth frame boundaries using Hamming window.

C. Extraction of features from template utterance

The selected tool for allocating syllables boundaries is the feature vector of local maxima picked from Delta function of the first Mel Frequency Cepstrum coefficient [23].
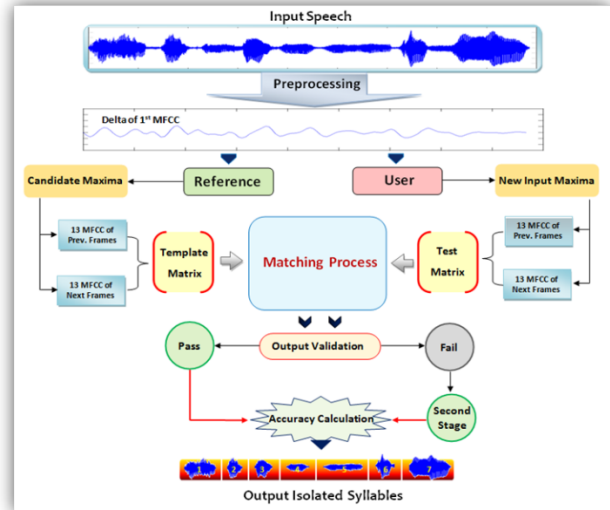

Fig. 1. Schematic diagram of the proposed system

Fig. 2 shows the procedure of obtain MFCC from the audio signal and Fig. 3 illustrates how local maxima are extracted from delta function of 1st MFCC along speech signal.
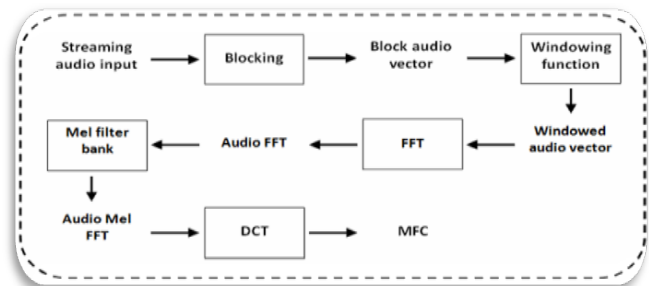

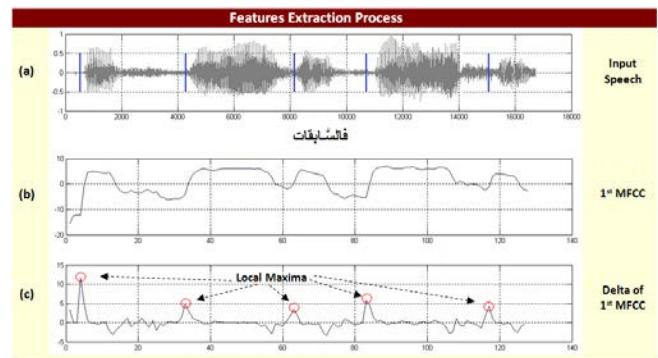Fig. 2. Obtaining the feature vector from speech stream

Fig. 3. Local maxima extraction from delta function of 1st MFCC. (a) Input speech signal. (b) 1st MFCC along whole signal. (c) Delta function of 1st MFCC with a selected local maxima

Fig. 4 clarifies the template matrix creation through selection of candidates. There are two factors affecting the accuracy of getting candidate that identify a syllable boundary at the template utterance, the frame length and/or the percentage of frames overlapping and the ML value "number of holding local maxima", as illustrated in Fig. 5.
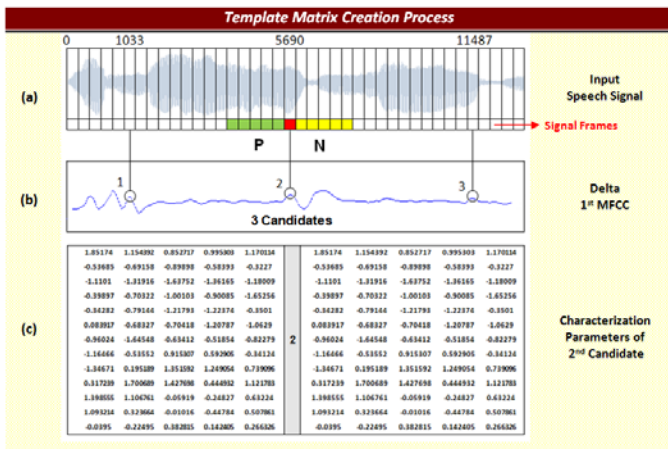

Fig. 4. Template matrix of candidates' parameters (a)Input signal. (b) Delta function of 1st MFCC with candidate maxima. (c) Characterization parameters of the 2nd candidate
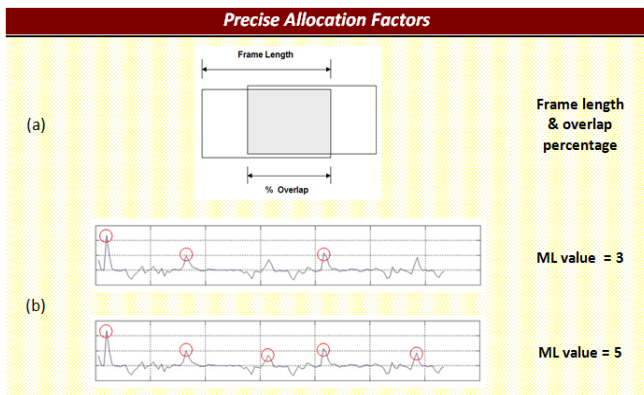

Fig. 5. Factors affect the precise allocation of syllables through hand tuning (a) Frame length and percentage of frames overlapping, (b) Delta function of 1st MFCC with "ML value=3" and "ML value=5"

D. Process the new input utterance

This module is responsible for processing the new input utterance that needs to be segmented. Each verse (utterance) has a fixed "ML value" to extract the local maxima from (PD-1st-MFC), depending on the syllabic structure of the utterance. The limiting value should be more than twice the number of existing syllables in the audio signal. Lower value of this limit results in less boundaries detection and more occurrences of connected syllables and vice versa. According to the pre-specified picked local maxima from user input, characterization parameters are obtained constituting the test matrix. The matching process is now ready to be applied between Template Matrix "output of the second module" and Test Matrix "output of the third module".
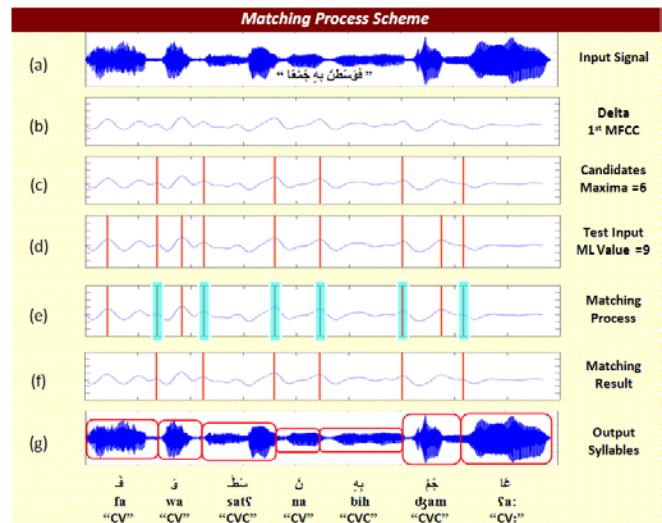

Fig. 6. Segmentation of speech syllables from input signal through template matching.(a) Input Signal. (b) Smoothed D elta function of 1^{st}MFCC along whole signal. (c) Candidate maxima selection using hand tuning. (d) Peaks configuration of the test input with ML value=9. (e) Matching process between (c), (d). (f) Resultant maxima from matching process. (g) Output segmented syllables with Arabic/IPA labels

E. Automatic identification of syllables boundaries through matching process

Identification of syllables boundaries of the user utterance is carried out by comparing its characterization parameters with the stored ones of the reference utterance using the matching technique schemed as shown in figure 6. Several methods can be used for the formulation of the rules in matching process based on distance measures techniques like Euclidean distance [24], Mahalanobis distance and Saito divergence [25]. The Euclidean distance is used for measuring closeness throughout the matching of this module. The final allocation of boundaries is obtained by passing the output of matching process through two stages of decomposition as will discussed in the next two subsections.

a) Primary Allocation of Syllables:

Table 1 shows the matching result as distances measure between user and template parameters. The closest local maxima are identified according to the minimum distance as displayed at the last row of table 1. These maxima represent locations of the target boundaries.

TABLE I. MATCHING RESULT BETWEEN TWO IDENTICAL UTTERANCES FROM THE SPEAKER (HSARY)

112 001	(Input utterance from "HSARY")										ML value = 13		
		1	2	3	4	5	6	7	8	9	10	11	12	13
(Template utterance) Candidate Maxima =6	1	37	0	30	29	37	43	42	37	34	46	50	38	42
	2	34	30	0	23	19	31	30	31	29	29	34	39	40
	3	38	37	19	28	0	29	29	32	32	30	31	37	38
	4	43	34	29	22	32	35	28	25	0	34	43	39	47
	5	44	50	34	40	31	36	37	40	43	26	0	37	34
	6	37	38	39	33	37	29	32	33	39	43	37	0	29
BP			2	3		5				9		11	12	

TABLE II.   MATCHING RESULT BETWEEN TWO UTTERANCES OF
DIFFERENT SPEAKERS (HSARY & AUOOB)

112 001	(Input utterance from "AUOOB")												ML value = 13
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	31	42	40	44	46	47	46	48	43	38	38	37	--
2	44	35	48	47	46	48	41	38	42	41	41	38	--
3	49	39	43	40	30	39	43	39	47	45	46	43	--
4	42	35	39	37	37	40	32	41	42	39	40	38	--
5	55	42	42	40	37	39	43	35	43	33	47	46	--
6	53	49	45	48	45	41	45	42	32	35	37	30	--
BP	1	2			5		7			10		12	

(Template utterance) Candidate Maxima =6

The first case in table 1 is the matching result between two identical utterances **"Qul huwa ʔallahu ʔahad , قُلْ هُوَ اللَّهُ أَحَد"** for same speaker (HSARY), therefore the closest distances are zeroes, this ensures the efficiency of the algorithm.

Each utterance from the test dataset is processed through the matching module with a template of the reference speaker. As reported in the last row of table 2, the boundaries of targeted syllables is represented by the local maxima of indexes (1st, 2nd, 5th, 7th, 10th & 12th) where these maxima had the best closeness among the picked thirteen maxima from (PD-1st-MFC) of the reader (AUOOB) with the six candidate maxima from the template utterance of the reader (HSARY).

b) Connected Syllables Breakup:

In some cases the output has one or more connected syllables, as shown in Fig. 7. This phenomenon appears as a result of one of the following interpretations:

a. Adjusted "ML value" is not enough to get the ideal number of local maxima, so that the maximum which represents the missing boundary was not taken it into account.

b. Framing duration is big rather while one or more of speech articulations disappeared inside one frame, since the frame is the smallest unit in the speech signal and should be selected precisely to avoid merged or spited syllables.

c. The recorded audio has a composite noise at this point, where the selected feature as a tool for segmentation is unable to detect the boundary between syllables in this area.

In this situation, detection of the missed boundary is performed in a semi-automatic manner. Local maxima are picked with a number equal to the missing boundaries.
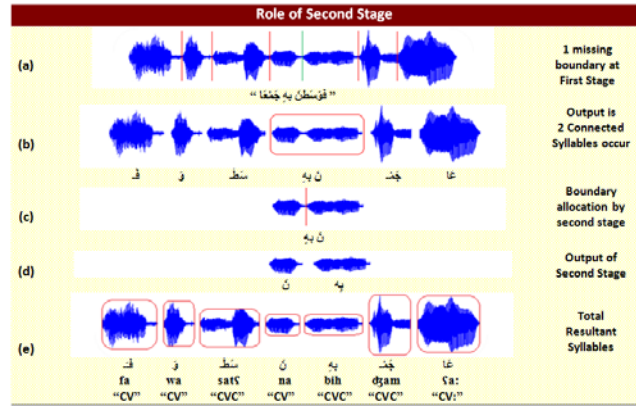

Fig. 7. The Role of second stage in the fourth module (a)One missing boundary at the first stage. (b) Two connected syllables occurrence. (c) Missing boundary allocation by the second stage. (d) Output isolated syllables by the second stage. (e) Total resultant syllables

IV. SEGMENTATION RESULTS

Since the system is speaker independent, utterances from different speakers were tested with an overall accuracy of 91.5 % as shown in table 3.

As an example, the results of applying the algorithm for segmenting the utterance **"Qul huwa ʔallahu ʔahad , قُلْ هُوَ اللَّهُ أَحَد"** from different speakers are tabulated in table 4.

TABLE III.   ACCURACY RESULTS OF THE TEST SAMPLES

No.	Verse Code	Verse Text Arabic/IPA	Syllables Count			Accuracy
			Reference	Correct	Error	
1	110001	إِذَا جَاءَ نَصْرُ اللهِ وَالْفَتْحُ ʔiða: dʒa:ʔa nasrullahi walfatħ	120	115	5	95.83 %
2	110002	وَرَأَيْتَ النَّاسَ يَدْخُلُونَ فِي دِينِ اللهِ أَفْوَاجًا waraʔajta ʔalnasa jadxulu:na fi: di:nillahi afwadʒa:	216	204	12	94.44 %
3	110003	فَسَبِّحْ بِحَمْدِ رَبِّكَ وَاسْتَغْفِرْهُ إِنَّهُ كَانَ تَوَّابًا fasabbiħ biħamdi rabbika wastaɣfirhu ʔinnaiu kana tawwaba:	252	201	51	79.76 %
4	111001	تَبَّتْ يَدَا أَبِي لَهَبٍ وَتَبَّ tabbat jada: ʔabi: lahabin watab	120	111	9	92.50 %
5	111002	مَا أَغْنَى عَنْهُ مَالُهُ وَمَا كَسَبَ ma: ʔaɣnā ʔanhu ma:luhu wama:kasab	144	140	4	97.22 %
6	111003	سَيَصْلَى نَارًا ذَاتَ لَهَبٍ sajasʕlā na:ran ða:ta lahab	108	107	1	99.07 %
7	111004	وَامْرَأَتُهُ حَمَّالَةَ الْحَطَبِ wamraʔatuhu hammalata ʔalhatab	108	104	4	96.30 %
8	111005	فِي جِيدِهَا حَبْلٌ مِنْ مَسَدٍ fi: dʒi:diha hablun min masad	108	106	2	98.15 %
9	112001	قُلْ هُوَ اللهُ أَحَدٌ qul huwa ʔallahu ʔahad	84	81	3	96.43 %

10	112002	اللّهُ الصَّمَدُ ʔallahu ʔalsˤamad	60	40	20	66.67 %
11	112003	لَمْ يَلِدْ وَلَمْ يُولَدْ lam jalid walam ju:lad	72	71	1	98.61 %
12	112004	وَلَمْ يَكُنْ لَهُ كُفُوًا أَحَدٌ walam jakun lahu kufuwan ʔahad	108	97	11	89.81 %
13	113001	قُلْ أَعُوذُ بِرَبِّ الْفَلَقِ qul ʔaʕu:ðu birabbi ʔilfalaq	108	101	7	93.52 %
14	113002	مِنْ شَرِّ مَا خَلَقَ min ʃarri ma: xalaq	72	65	7	90.28 %
15	113003	وَمِنْ شَرِّ غَاسِقٍ إِذَا وَقَبَ wa min ʃarri ɣa:siqin ʔiða: waqab	120	112	8	93.33 %
16	113004	وَمِنْ شَرِّ النَّفَّاثَاتِ فِي الْعُقَدِ wa min ʃarri ʔalnnaffa:θa:ti fi: ʔilʕuqad	120	113	7	94.17 %
17	113005	وَمِنْ شَرِّ حَاسِدٍ إِذَا حَسَدَ wa min ʃarri ha:sidin ʔiða hasad	120	113	7	94.17 %
18	114001	قُلْ أَعُوذُ بِرَبِّ النَّاسِ qul ʔaʕu:ðu birabbi ʔilna:s	96	91	5	94.79 %
19	114002	مَلِكِ النَّاسِ maliki ʔilna:s	48	35	13	72.92 %
20	114003	إِلَهِ النَّاسِ ʔilahi ʔilna:s	48	40	8	83.33 %
21	114004	مِنْ شَرِّ الْوَسْوَاسِ الْخَنَّاسِ min ʃarri ʔilwaswa:si ʔilxanna:s	96	88	8	91.67 %
22	114005	الَّذِي يُوَسْوِسُ فِي صُدُورِ النَّاسِ ʔllaði: juwaswisu fi: sˤuduri ʔilna:s	132	125	7	94.70 %
23	114006	مِنَ الْجِنَّةِ وَالنَّاسِ min ʔaldʒennati waʔalna:s	84	81	3	96.43 %
			2544	2341	203	91.48%

TABLE IV.   SPEAKER INDEPENDENCY TEST RESULTS

Test Utterance : قُلْ هُوَ اللَّهُ أَحَدٌ
Utterance Type : Quranic Verse   **ML value** : 13
Utterance Code : 112001   **Actual Syllables** : 7 (Bulk:0)
Reference Reader: HSARY   **Segmentation Stages:** 2

No.	Reader Code	Start/ End Trim	# O/P Syllables					Calculated Accuracy	Total Accuracy
			First Stage			Second Stage			
			Matched	Merged	Splitted	Matched	Un-matched		
1	HSARY	0/0	7	–	–	–	–	100	
2	AKHDR	1/0	2	3,2	–	7	–	100	
3	AUOOB	0/1	6	–	1	6	1	86	
4	BASIT	1/1	3	2,2	–	7	–	100	
5	GHMDY	0/1	5	2	–	7	–	100	
6	HZYFI	1/1	2	2,3	–	7	–	100	96.5 %
7	JBRIL	0/1	5	2	–	7	–	100	
8	MAQLY	1/1	6	–	1	6	1	86	
9	QASIM	1/1	5	2	–	7	–	100	
10	REFAY	0/1	2	5	–	6	1	86	
11	SHTRY	0/1	3	2,2	–	7	–	100	
12	SHRIM	0/2	5	2	–	7	–	100	

V.   CONCLUSION

The main purpose of this paper is to implement precise semi-automatic speaker independent system for building a database of *Syllables* banks from continuous Arabic uttered speech. The developed method employs the vector of local maxima picked from peaks of the delta function of first Mel Frequency Cepstrum Coefficient as cutting tools that predict possible locations of syllables boundaries inside the continuous speech. The final boundaries are allocated by taking into account the number of segments predicted and the closeness between the predicted and the reference segment boundaries. The results have shown that the system was able to break up a set of 276 Arabic utterances into its syllables with up to 91.5 % accuracy.

VI. LIST OF ABBREVIATIONS

MFC: Mel Frequency Cepstrum.

MFCC: Mel Frequency Cepstrum Coefficient.

SNLP: Speech and Natural Language Processing.

CALL: Computer Aided Language Learning.

MSA: Modern Standard Arabic.

ZCR: Zero Crossing Rate.

PD-1st -MFC: Peaks of Delta function of the first Mel Frequency Cepstrum Coefficient.

ML value: Maximum value of picked Local Maxima from the utterance for processing.

IPA: International Phonetic Alphabet.

BP: Boundary Position.

REFERENCES

[1] Kirchhoff K., Bilmes J., Das S., et al. "Novel approaches to Arabic speech recognition": Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 03), vol.1, pp.344–347, 2003.

[2] Hassan Dahan, Abdul Hussin, Zaidi Razak, Mourad Odelha "Automatic Arabic Pronunciation Scoring for Language Instruction". In Proceedings of EDULEARN11 Conference, pp. 145-150, Barcelona, Spain, 2011.

[3] T. I. El Arif, Z. T. Fayed, M. E. Gad Allah, and A. I. Eldesoky "Automatic phonetic segmentation of Arabic speech without linguistic knowledge". ICICIS conference, pp. 265-271, 2005.

[4] Amr M. Gody "Speech Processing Using Wavelet Based Algorithms", PhD Thesis, Cairo University, 1999.

[5] M. F. Tolba, T. Nazmy, A. A. Abdelhamid, and M. E. Gadallaha "A Novel Method for Arabic Consonant/Vowel Segmentation using Wavelet Transform". International Journal on Intelligent Cooperative Information Systems, 5 (1), pp. 353-364, 2005.

[6] Yacine Yekache, Yekhlef Mekelleche and Belkacem Kouninef "Towards Quranic reader controlled by speech". International Journal of Advanced Computer Science and Applications, Vol. 2, No. 11, 2011.

[7] J. F. Wang, C. H. Wu, S. H. Chang, and J. Y Lee "A Hierarchical Neural Network Model Based on a C/V Segmentation Algorithm for Isolated Mandarin Speech Recognition". IEEE Trans. On Signal Processing, 39 (9), pp.2131-2136, 1991.

[8] Stephen W. K. Fu, C. H. Lee, O. L. Clubb "A Robust C/V Segmentation Algorithm for Cantonese". IEEE TENCON, pp. 42-45, 1996.

[9] Anupriya Sharma, Amanpreet Kaur "A Survey on Punjabi Speech Segmentation into Syllable-Like Units Using Group Delay". International Journal of Advanced Research in Computer Science and Software Engineering, 3, Issue 6, 2013.

[10] Li Zhang "A Syllable-based Pseudo-articulatory Approach to Speech Recognition", PhD thesis, School of Computer Science, Faculty of Science, University of Birmingham, 2004.

[11] V. Kamakshi Prasad, T. Nagarajan, Hema A. Murthy "Automatic segmentation of continuous speech using minimum phase group delay functions". International Journal of Speech Communication, 42, pp. 429–446, 2004.

[12] G. Sarada, A. Lakshmi, Hema A. Murthy and T. Nagarajan "Automatic transcription of continuous speech into syllable-like units for Indian languages". International Journal of Sadhana, 34, pp. 221–233., 2009.

[13] Er. Amanpreet Kaur and Er. Tarandeep Singh "Segmentation of Continuous Punjabi Speech Signal into Syllables". Proceedings of the World Congress on Engineering and Computer Science, WCECS, Vol I, pp.20-22, San Francisco, USA, 2010.

[14] Ozlem Kalinli "Syllable Segmentation of Continuous Speech Using Auditory Attention Cues". Interspeech, pp. 425–428, 2011.

[15] Nicolas Obin, François Lamare, Axel Roebel "SYLL-O-MATIC: An Adaptive Time-Frequency Representation for the Automatic Segmentation of Speech into Syllables". In Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 26-31, 2013.

[16] Rudi Villing, Joseph Timoney, Tomas Ward and John Costello "Automatic Blind Syllable Segmentation for Continuous Speech". In Irish Signals and Systems Conference, Queens University, Belfast, 2004.

[17] Mijanur Rahman, Fatema Khatun, and Al-Amin Bhuiyan "Blocking Black Area Method for Speech Segmentation". International Journal of Advanced Computer Science and Applications, Vol. 4, No.2, 2015.

[18] Mohamed S. Abdo, A. H. Kandil, A. M. El-Bialy, S. A. Fawzy "Automatic Detection for Some Common Pronunciation Mistakes Applied to Chosen Quran Sounds". In Proceedings of IEEE Cairo International Biomedical Engineering Conference, pp. 219-222, Cairo, Egypt, 2010.

[19] Mohamed S. Abdo, Ahmed H. Kandil, Sahar A. Fawzy "MFC Peak Based Segmentation for Continuous Arabic Audio Signal". In Proceedings of IEEE 2nd Middle East Conference on Biomedical Engineering, pp. 17-20, Doha, Qatar, 2014.

[20] Brierly, C., Sawalha M., Heselwood B. And Atwell E. S. "A verified Arabic-IPA mapping for Arabic transcription technology, informed by Quranic recitation, traditional Arabic linguistics and modern phonetics". Journal of Semitic Studies, 2014.

[21] Simon Lindholm "A speech recognition system for Swedish running on Android". MSc. Thesis, Lunds universitet, LTH, 2010.

[22] Olivier Lartillot, Petri Toiviainen, Tuomas Eerola "A Matlab Toolbox for Music Information Retrieval". In C. Preisach, H. Burkhardt, L. Schmidt-Thieme, R. Decker (Eds.), Data Analysis, Machine Learning and Applications, Studies in Classification, Data Analysis, and Knowledge Organization, Springer-Verlag, 2008.

[23] Trabelsi I. and Ayed D. Ben "On the Use of Different Feature Extraction Methods for Linear and Non Linear kernels". IEEE, Sciences of Electronic, Technologies of Information and Telecommunications, 2012.

[24] Gorriz J. M., Ramirez J., Segura J. C. & Puntonet C. G. "An effective cluster-based model for robust speech detection and speech recognition in noisy environments". Journal of the Acoustical Society of America, 120, pp. 470, 2006.

[25] Ramirez J., Segura J. C., Benitez C., Torre A. & Rubio A. "A New Kullback- Leibler VAD for Robust Speech Recognition". IEEE Signal Processing Letters, 11, pp. 266, 2004.

[26] Vahid Majidnezhad "A novel hybrid of genetic algorithm and ANN for developing a high efficient method for vocal fold pathology diagnosis". EURASIP Journal on Audio, Speech, and Music Processing. DOI 10.1186/s13636-014-0046-1, 2015.

[27] Gelzinis A., Verikas A., Bacauskiene M. "Automated speech analysis applied to laryngeal disease categorization". Journal of Computer Methods and Programs in Biomedicine, 91, pp. 36-47, 2008.