



Faculty of Engineering



Cairo University

## ***Automatic Arabic Speech Syllables Segmentation***

By

**Mohamed Sayed Abdelmonem Abdo**

A Thesis submitted to the

Faculty of Engineering, Cairo University

In Partial Fulfilment of the Requirements for the Degree of

**Doctor of Philosophy**

**In**

**BIOMEDICAL ENGINEERING AND SYSTEMS**

FACULTY OF ENGINEERING, CAIRO UNIVERSITY  
GIZA, EGYPT  
2018

# ***Automatic Arabic Speech Syllables Segmentation***

By

**Mohamed Sayed Abdelmonem Abdo**

A Thesis submitted to the

Faculty of Engineering, Cairo University

In Partial Fulfillment of the Requirements for the Degree of

**Doctor of Philosophy**

**In**

**BIOMEDICAL ENGINEERING AND SYSTEMS**

**Under the supervision of**

**Prof. Ahmed Mohamed El-Bialy**

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

**Prof. Ahmed Hisham Kandil**

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

**Dr. Sahar Ali Fawzi**

Associate Professor Department of  
Systems and Biomedical Engineering  
Faculty of Engineering, Cairo University

FACULTY OF ENGINEERING, CAIRO UNIVERSITY  
GIZA, EGYPT  
2018

# ***Automatic Arabic Speech Syllables Segmentation***

By

**Mohamed Sayed Abdelmonem Abdo**

A Thesis submitted to the

Faculty of Engineering, Cairo University

In Partial Fulfillment of the Requirements for the Degree of

**Doctor of Philosophy**

**In**

**BIOMEDICAL ENGINEERING AND SYSTEMS**

**Approved by the Examining Committee:**

---

**Prof. Ahmed Mohamed El-Bialy**

**(Thesis Main Advisor)**

---

**Dr. Sahar Ali Fawzi**

**(Thesis Advisor)**

---

**Prof. Mohamed Riyad Elghoniemy**

**(Internal Examiner)**

---

**Prof. Samia Abdelrazik Mashali**

**(External Examiner)**

**(Professor, Computer and Systems Department – Electronic Research Institute)**

FACULTY OF ENGINEERING, CAIRO UNIVERSITY  
GIZA, EGYPT  
2018

**Engineer:** Mohamed Sayed Abd Elmonem Abdo  
**Date of Birth :** 29 / 1 / 1981  
**Nationality :** Egyptian  
**E-mail :** Bioengmsa@yahoo.com  
**Registration Date :** 1 / 3 / 2012  
**Awarding Date :** / / 2018  
**Degree :** Doctor of Philosophy  
**Department :** Biomedical Engineering and Systems



**Supervisors :** **Prof. Ahmed Mohamed El-Bialy**  
**Prof. Ahmed Hisham Kandil**  
**Dr. Sahar Ali Fawzi**

**Examiners :** **Prof. Ahmed Mohamed El-Bialy** (Thesis Main Advisor)  
**Dr. Sahar Ali Fawzi** (Thesis Advisor)  
**Prof. Mohamed Riyad Elghoniemy** (Internal Examiner)  
**Prof. Samia Abdel Razek Mashali** (External Examiner)  
(Professor, Computer and Systems Department – Electronic Research Institute)

**Title of Thesis :** *Automatic Arabic Speech Syllables Segmentation*

**Key Words:** Arabic language, Automatic segmentation, syllable boundaries allocation, Mel Frequency Cepstral Coefficients “MFCC”.

### **Summary :**

Syllables are the fundamental units of Arabic language. The proposed “Neural Network based Arabic Speech Segmentation System (NNASS)” is an adaptive Arabic speech syllable boundaries identifier that mainly serves as an automatic segmentation tool for speaker independent “Arabic speech verification (ASV)” and speech corpus/database construction systems. Cpestral peaks extracted from recorded speech signal within a certain validation thresholds assignment are considered probable boundaries. These probable boundaries are applied to NNASS to classify them into valid or invalid ones. An algorithm using neural networks is developed to train the features of valid boundaries/ cores. A program is developed to precisely identify the boundaries/cores from the test utterance, where the segmentation is done at their locations. The accuracy of NNASS was 87 % and 92.2 % identification rates with a semi-automatic labeling of the test dataset for verification within 10 and 20 milliseconds using two sample sizes. It will be shown that the system can be expanded to include more trained utterances for more than application.

## ***Acknowledgment***

First of all, I would like to thank **God** for selecting me to serve his words, always who guides me to the right path.

I am deeply indebted to my advisors and supervisors:

**Prof. Dr. Ahmed Hisham Kandil**, for his time and effort he devoted to the supervision and theoretical guidance to this work,

**Prof. Dr. Ahmed Mohamed El-Bialy**, for suggesting the subject of research, effort done in this work by his thoughts and references,

**Dr. Sahar Ali Fawzy**, for her sincere support, valuable help and useful discussions throughout the course of this work. They helped and pushed me in the proper directions.

In a special word of appreciation, I would like to extend my thanks and gratitude to (my **parents**, my **wife** and my **sons**) for their understanding support, encouragement and praying for me along these years.

# TABLE OF CONTENTS

	Page
<b>Acknowledgment</b>	<b>i</b>
<b>List of Figures</b>	<b>v</b>
<b>List of Tables</b>	<b>vii</b>
<b>List of Equations</b>	<b>viii</b>
<b>Abbreviations</b>	<b>ix</b>
<b>Abstract</b>	<b>x</b>
 <b>1. INTRODUCTION</b>	 <b>1</b>
1.1 Introduction	1
1.2 Basic Concepts in Speech	2
1.2.1 Utterances & Silence	2
1.2.2 Pronunciations	2
1.2.3 Speaker Dependence vs. Speaker Independence	2
1.2.4 Speech Recognition vs. Speech Verification	2
1.3 Neural Network Arabic Speech Segmentation (NNASS) Overview	3
1.4 Thesis Objectives	3
1.5 Thesis Organization	4
 <b>2. BACK GROUND AND RELATED WORKS</b>	 <b>5</b>
2.1 Introduction	5
2.2 Voice Acoustics	6
2.2.1 Introduction	6
2.2.2 Voice Classification	6
2.2.3 The Source-Filter Model	7
2.3 Segmentation Concepts	8
2.3.1 Metric-based segmentation	8
2.3.2 Phonetic Detection without Boundaries	9
2.3.3 Strict Phonetic Segmentation and Labeling	9
2.3.4 Energy-based segmentation	10
2.3.5 Model-based segmentation	10
2.3.6 The Phoneme Spotting Method	10
2.3.7 The On Line Segmentation	11

2.3.8	<i>The Off-line Segmentation</i>	11
2.4	<i>Related works</i>	12
2.4.1	<i>Speech segmentation for Arabic Language</i>	13
2.4.2	<i>Speech segmentation for Tamil Language</i>	14
2.4.3	<i>Speech segmentation for Maltese Language</i>	14
2.4.4	<i>Speech segmentation for Mandarin Language</i>	15
2.4.5	<i>Speech segmentation for French Language</i>	16
2.5	<i>Motivation for the proposed method</i>	16
<b>3.</b>	<b>SPEECH FEATURES</b>	<b>17</b>
3.1	<i>Introduction</i>	17
3.2	<i>Time Domain Features</i>	17
3.2.1	<i>Short-Term Energy</i>	17
3.2.2	<i>Zero Crossing Rate</i>	17
3.3	<i>Frequency Domain Features</i>	19
3.3.1	<i>Frequency Spectrum</i>	19
3.3.2	<i>Formants</i>	19
3.3.3	<i>Linear Predictive Analysis (LPC)</i>	21
3.3.4	<i>Log Area Ratio Coefficients (LAR)</i>	21
3.3.5	<i>Mel-Frequency Cepstral Coefficients (MFCCs)</i>	22
<b>4.</b>	<b>ARCHITECTURE OF ARABIC SPEECH SEGMENTATION SYSTEM</b>	<b>27</b>
4.1	<i>Introduction</i>	27
4.2	<i>The Proposed Segmentation System</i>	27
4.3	<i>Features Analysis</i>	28
4.3.1	<i>Delta 1<sup>st</sup> MFCC maxima analysis</i>	28
4.3.2	<i>13-MFCCs combination analysis</i>	29
4.4	<i>Data Collection and Preprocessing</i>	30
4.5	<i>Training Dataset</i>	32
4.5.1	<i>Semi-automatic allocation of candidate reference boundaries</i>	32
4.5.2	<i>Automatic allocation of the nearest maxima to each candidate reference boundary</i>	33
4.6	<i>Approaches of Segmentation</i>	34
4.7	<i>Building the Neural Network Segmenter Models</i>	34

4.8	<i>Training Neural Network Models with Reference Data</i>	34
4.9	<i>Classification of the Test Dataset</i>	36
4.9.1	<i>Determination of boundaries regions from test input</i>	36
4.9.2	<i>Validation and accuracy calculation</i>	36
4.9.2.1	<i>Semi-automatic Labeling</i>	37
4.9.2.2	<i>Verifications</i>	37
<b>5.</b>	<b>RESULTS AND DISCUSSIONS</b>	<b>39</b>
5.1	<i>Introduction</i>	39
5.2	<i>Scope and Limitation of the Study</i>	39
5.3	<i>Test data preparation and result analysis</i>	39
5.3.1	<i>Test data preparation using semi-automatic labeling</i>	39
5.3.2	<i>Experimental test results</i>	40
	<i>Experiment 1: Select the best speech processing variables</i>	40
	<i>Experiment 2: Select the best speech features of discrimination</i>	43
	<i>Experiment 3: Test effect of change number of MFCC on identification efficiency</i>	45
	<i>Experiment 4: Segmentation performance of the boundary based approach</i>	46
	<i>Experiment 5: Segmentation performance of the core based approach</i>	56
	<i>Experiment 6: Segmentation performance of the KNN based approach</i>	62
	<i>Experiment 7: Accuracy of automatic segmentation for speaker dependent</i>	65
	<i>Experiment 8: Reading's rate consistency measure for dataset readers</i>	66
5.3.3	<i>Work comparison</i>	76
<b>6.</b>	<b>CONCLUSION AND RECOMMENDATIONS</b>	<b>77</b>
6.1	<i>Introduction</i>	77
6.2	<i>Conclusion</i>	77
6.3	<i>Thesis Achievements</i>	78
6.4	<i>Recommendations</i>	78
	<b>REFERENCES</b>	<b>81</b>
	<b>APPENDIX-A</b>	<b>87</b>
	<i>Percentages of Syllable Boundaries Locations</i>	87
	<b>APPENDIX-B</b>	<b>91</b>
	<i>Time Periods of Quranic Arabic Syllables</i>	91



## List of Figures

Fig. 1.1	Maxima Extraction from Delta 1 <sup>st</sup> MFCC
Fig. 1.2	Brief diagram of proposed segmentation system principle
Fig. 2.1	Arabic Syllable Components
Fig. 2.2	The source filter model of speech.
Fig. 2.3	Output of the source filter model.
Fig. 2.4	Shape of voiced sound.
Fig. 2.5	Shape of unvoiced sound.
Fig. 3.1	Zero crossings of the signal.
Fig. 3.2	Short-term energy vs zero crossings for the word “seven”.
Fig. 3.3	Spectrum and smoothed spectrum of speech.
Fig. 3.4	Spectral peaks of the sound spectrum.
Fig. 3.5	5-points local maxima
Fig. 3.6	Acoustic tubes speech production model
Fig. 3.7	Overview of the MFCC process
Fig. 3.8	Spectrogram of the speech signal
Fig. 3.9	Triangular filters used to compute Mel-Cepstrum
Fig. 3.10	Diagram of computational steps of MFCCs
Fig. 4.1	Block Diagram of the Proposed System
Fig. 4.2	Maxima Extraction from Delta 1 <sup>st</sup> MFCC
Fig. 4.3	Extraction of Candidates’ Parameters
Fig. 4.4	Creation of Training Data
Fig. 4.5	Extraction of Nearest Maxima and their Features of Reference (HSARY)
Fig. 4.6	Neural Network Structure

- Fig. 4.7      Nearest Extraction at Boundaries Regions of Test Reader (BUKTR)
- Fig. 4.8      Validation of Resultant Boundaries
- Fig. 5.1      Diagram analysis for the identified boundaries from the test readers
- Fig. 5.2      Scheme of the second approach
- Fig. 5.3      The three types of the second approach
- Fig. 5.4      Sample of speaker dependent test result
- Fig.5.5      Gaussian distribution of CV syllables period's consistency for AFASY reader.
- Fig.5.6      Gaussian distribution of CV syllables period's consistency for AKHDA reader.
- Fig. 5.7      Gaussian distribution of CV syllables period's consistency for AUOOB reader.
- Fig. 5.8      Gaussian distribution of CV syllables period's consistency for BASFR reader.
- Fig. 5.9      Gaussian distribution of CV syllables period's consistency for BASIT reader.
- Fig. 5.10      Gaussian distribution of CV syllables period's consistency for BUKTR reader.
- Fig.5.11      Gaussian distribution of CV syllables period's consistency for GHMDY reader.
- Fig. 5.12      Gaussian distribution of CV syllables period's consistency for HSARY reader.
- Fig. 5.13      Gaussian distribution of CV syllables period's consistency for HZYFI reader.
- Fig. 5.14      Gaussian distribution of CV syllables period's consistency for JBRIL reader.
- Fig. 5.15      Gaussian distribution of CV syllables period's consistency for SUDIS reader.
- Fig. 5.16      Gaussian distribution of CV syllables period's consistency for SHTRY reader.
- Fig. 5.17      Gaussian distribution of CV syllables period's consistency for SHRIM reader.
- Fig. 5.18      Gaussian distribution of CV syllables period's consistency for REFAY reader.
- Fig. 5.19      Gaussian distribution of CV syllables period's consistency for QASIM reader.
- Fig. 5.20      Gaussian distribution of CV syllables period's consistency for MNSWY reader.
- Fig.5.21      Gaussian distribution of CV syllables period's consistency for MAQLY reader.
- Fig. 5.22      Gaussian distribution of CV syllables period's consistency for TBLWY reader.

## **List of Tables**

Table 4.1:	Dataset Build for this Study.
Table 4.2:	Description of Training Dataset.
Table 5.1:	Dataset used to select the best speech processing variables.
Table 5.2:	Trials of selecting the best speech processing variables
Table 5.3:	Test the best representation features of speech.
Table 5.4:	Effect of changing the number of MFC coefficients
Table 5.5:	Dataset used to experiment the first approach.
Table 5.4:	Effect of changing the number of MFC coefficients
Tables (5.6-5.20):	Validation of Results for the Test Utterances of the First Approach.
Table 5.21:	Validation of Results of a Sample Test Utterance from First Method of the Second Approach
Table 5.22:	Validation of Results of a Sample Test Utterance from Second Method of the Second Approach.
Table 5.23:	Validation of Results of a Sample Test Utterance from Third Method of the Second Approach.
Table 5.24:	Dataset used to experiment the third approach.
Table 5.25:	Validation of Results of a Sample Test Utterance from the 3 <sup>rd</sup> Approach.
Table 5.26:	Comparison of the performance of the related works with NNASS
Table A:	Percentages of Syllable Boundaries Locations.
Table B:	Time Periods of Quranic Arabic Syllables.

## **List of Equations**

- Equation (1): Short-term energy.
- Equation (2): Zero crossing rates.
- Equation (3): Fourier Transform.
- Equation (4): Spectrum from FFT of the estimated AR model parameters.
- Equation (5): The corresponding formula for the LPC model.
- Equation (6): The relationship between the LAR coefficients and the LPC.
- Equation (7): Mel-Frequency Transformation.
- Equation (8): The most common DCT definition of a 1-D sequence.
- Equation (9): The inverse transformation.
- Equation (10): Accuracy calculation.

## **Abbreviations**

(TTS)	Text to speech systems
(ANN)	Artificial Neural Networks
(CAPL)	Computer Aided Pronunciation Learning
(BRC)	Boundary Region Counter
(SBR)	Starting Boundary Region
(EBR)	Ending Boundary Region
(DFT)	Discrete Fourier Transform
(ASR)	Automatic Speech Recognition
(ASV)	Automatic Speech Verification
(MSA)	Modern Standard Arabic
(SROL)	Sounds of the Romanian Language
(FFT)	Fast Fourier Transform
(HMM)	Hidden Markov Model
(HTK)	Toolkit for the recognition
(LPC)	Linear Predictive Coefficients
(LAR)	Log Area Ratio
(MLE)	Maximum Likelihood Estimation
(MFCC)	Mel-Frequency Cepstral Coefficient
(ZCR)	Zero Crossing Rate
(STE)	Short-Term Energy
(KNN)	K- Nearest Neighbor.
(BRIV)	Boundary Regions Indicator Vector.

## ABSTRACT

Syllables are fundamental units of Arabic speech that play a vital role in different speech applications such as ASR, ASV and speech corpus/database construction systems. The speech utterance is a sequence of syllables. There is a significant difference in acoustic energy between syllables. The goal of this work is to develop a precise speaker independent system for the automatic segmentation of continuous speech into syllables. The proposed Neural Network Arabic Speech Segmentation system (NNASS) implements two approaches for Arabic speech segmentation using neural networks as an adaptive syllable boundaries identifier, boundaries features based approach and cores features based approach. The training set of NNASS is composed of a number of different candidate boundaries features from reference voices. NNASS behaves as a multiple classifier; it is capable of recognizing syllable boundaries of Arabic utterances irrelevant of their nature. The system was tested by applying continuous audio signals. Speech signal features and its cepstral peaks were extracted, and applied to NNASS to classify them into valid / invalid boundaries, through extracting the discriminating features for the syllable boundaries in Arabic speech, building the Neural Network for the identification of the boundaries and developing an algorithm for the automatic segmentation of the speech stream.

A set of 18 readers representing different Arabic countries was selected; each recited 15 continuous Quranic utterances “verses” constituting a total of 270 utterances containing 1908 boundaries. An analysis to select the best acoustical representation features for syllable boundaries was performed. An algorithm to train neural networks neurons was developed based on features of valid boundaries/cores, then a validation phase was achieved to locate syllables boundaries.

The accuracy of NNASS reached up to 87% and 92.2% identification rates with a semi-automatic labeling of the test dataset for verification within 10 and 20 milliseconds. This system proved the validity of the concept of using MFC difference feature as a mark for inter-syllables transitions that can be used in several applications.

## CHAPTER ONE INTRODUCTION

### 1.1 Introduction

The aim of this thesis is to develop a precise system for Arabic speech syllabification (i.e., segmentation of Arabic utterance into syllables units). Segmentation and labeling of a small sized acoustic corpus, of 2-3 hours recordings of uttered verses for the speech database construction, [1] is a time consuming (inconsistent) process. The proposed solution is an automatic segmentation algorithm based on neural networks models and embedded cues related to syllable boundaries to produce accurate syllable segmentation. This algorithm is used to segment Quranic verses and to form a corpus for recitation verification systems. Segmentation of speech at syllable level is an essential phase in many applications, such as text to speech systems (TTS), teaching the recitation rules of Holy Quran automatically, teaching Arabic language pronunciation for non-native speakers, correct the pronunciation and speech disorders for children and patients having defects in their speech production system. Applications such as speech verification and speech synthesis require highly accurate and consistent segmentation [2, 3]. Figure 1.1 shows a sample of Arabic utterance segmented at the syllable level using the delta first MFC coefficient as indicators to the syllable boundaries.

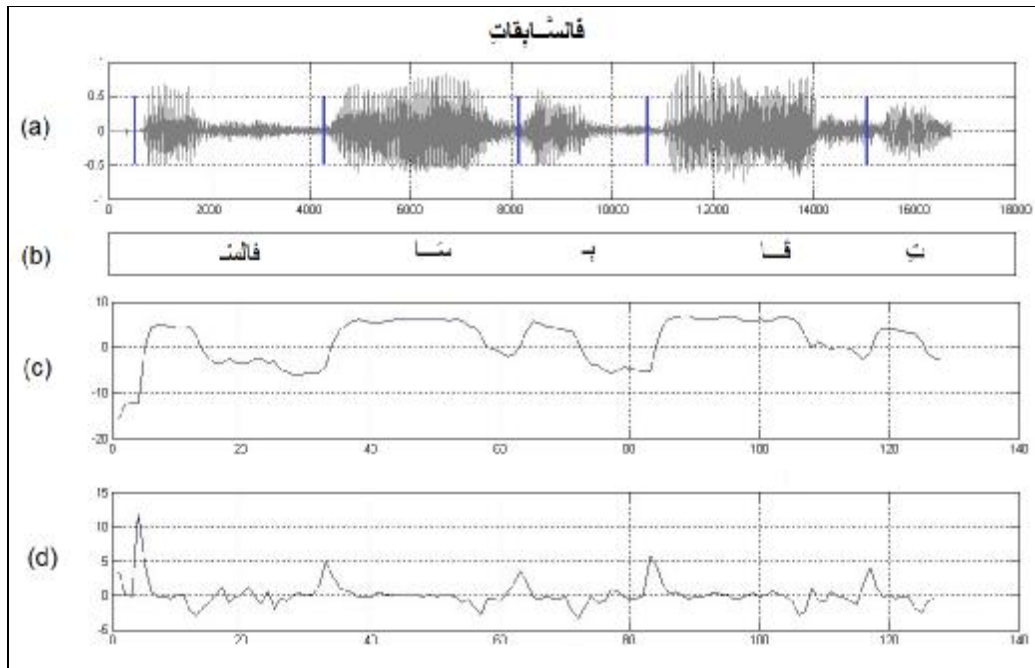


Fig. 1.1: Maxima Extraction from Delta 1<sup>st</sup> MFCC.

- (a) Input speech signal with marked locations of syllables boundaries. (b) Syllables transcriptions (c) 1<sup>st</sup> MFCC. (c) Delta 1<sup>st</sup> MFCC with local maxima.