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

List of Figures

List of Tables

List of Equations

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.

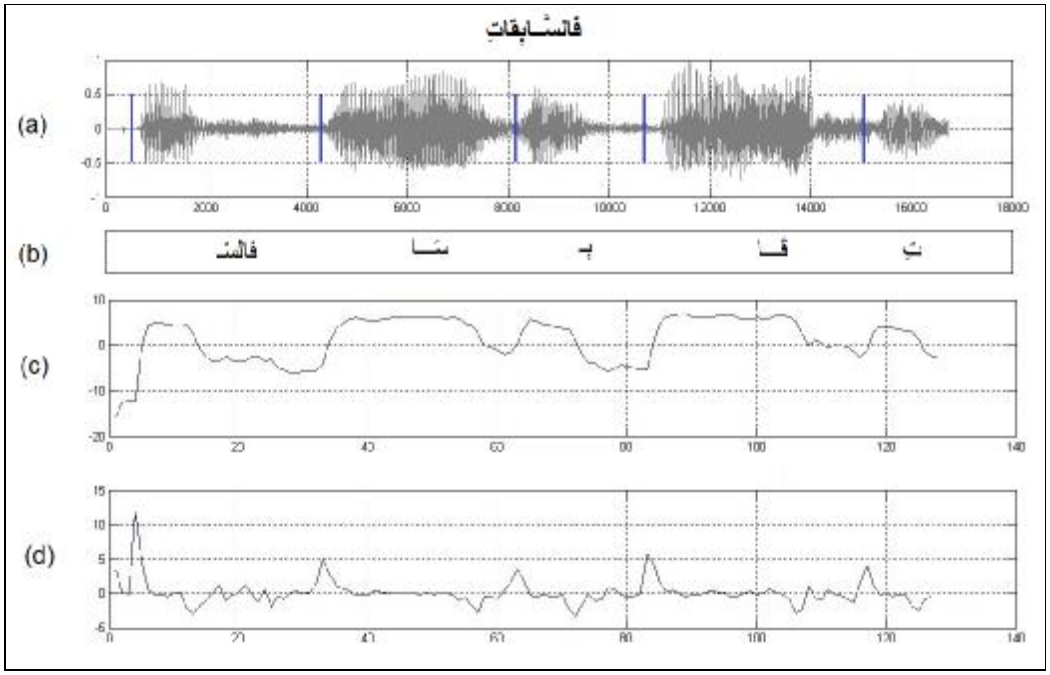

Fig. 1.1: Maxima Extraction from Delta 1st MFCC.

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