





Cairo University

# **SUBVOCAL SPEECH RECOGNITION USING ENGINEERED FEATURES AND DEEP LEARNING**

By

**Mohamed Said Elbially Elmahdy**

A Thesis Submitted to the  
Faculty of Engineering at Cairo University  
in Partial Fulfillment of the  
Requirements for the Degree of  
**MASTER OF SCIENCE**  
in  
**Biomedical Engineering and Systems**

FACULTY OF ENGINEERING, CAIRO UNIVERSITY  
GIZA, EGYPT  
2017



# **SUBVOCAL SPEECH RECOGNITION USING ENGINEERED FEATURES AND DEEP LEARNING**

By  
**Mohamed Said Elbially Elmahdy**

A Thesis Submitted to the  
Faculty of Engineering at Cairo University  
in Partial Fulfillment of the  
Requirements for the Degree of  
**MASTER OF SCIENCE**  
in  
**Biomedical Engineering and Systems**

Under the Supervision of

Dr. Ahmed A. Morsy

.....

Professor  
Biomedical Engineering and Systems  
Faculty of Engineering, Cairo University

FACULTY OF ENGINEERING, CAIRO UNIVERSITY  
GIZA, EGYPT  
2017



# **SUBVOCAL SPEECH RECOGNITION USING ENGINEERED FEATURES AND DEEP LEARNING**

By  
**Mohamed Said Elbially Elmahdy**

A Thesis Submitted to the  
Faculty of Engineering at Cairo University  
in Partial Fulfillment of the  
Requirements for the Degree of  
**MASTER OF SCIENCE**  
in  
**Biomedical Engineering and Systems**

Approved by the  
Examining Committee

---

Prof. Dr. Ahmed A. Morsy, Thesis Main Advisor

- Professor of Biomedical Engineering and Systems, Cairo University

---

Prof. Dr. Manal Abdel Wahed, Internal Examiner

- Professor of Biomedical Engineering and Systems, Cairo University

---

Prof. Dr. Hazem M. Abbas, External Examiner

- Professor of Computer and Systems Engineering, Ain Shams University

FACULTY OF ENGINEERING, CAIRO UNIVERSITY  
GIZA, EGYPT  
2017



**Engineer's Name:** Mohamed Said Elbially Elmahdy  
**Date of Birth:** 26/8/1991  
**Nationality:** Egyptian  
**E-mail:** m.elbially@eng1.cu.edu.eg  
**Phone:** 01123332586  
**Address:** 58 Elemam Ali St. Elzawia Elhamra  
**Registration Date:** 1/10/2013  
**Awarding Date:** 25/5/2017  
**Degree:** Master of Science  
**Department:** Biomedical Engineering and Systems



**Supervisors:**  
Prof. Ahmed A. Morsy

**Examiners:**  
Prof. Hazem M. Abbas (External examiner)  
Prof. Manal Abdel Wahed (Internal examiner)  
Prof. Ahmed A. Morsy (Thesis main advisor)

**Title of Thesis:**

Subvocal Speech Recognition Using Engineered Features and Deep Learning

**Key Words:**

Surface Electromyography; Deep Learning; Subvocal Speech.

**Summary:**

In this study we propose an end-to-end deep system for subvocal speech recognition. A single channel surface Electromyogram (sEMG) placed diagonally around the throat is used alongside a close-talk microphone for signal acquisition. The system was tested on a corpus of 20 words. The system classification was independent of the word level but smart enough to learn the mapping function from sound and sEMG sequences to letters, then extracting the most probable word from these letters. Different input signals and different depth levels were investigated using the deep learning model. The system was tested on ten healthy subjects (5 females, 5 males). The proposed system achieved a Word Error Rate (WER) of 9.44, 8.44 and 9.22 for speech, speech combined with single channel sEMG and speech with two channels of sEMG, respectively.

In order to compare the system with the results from literature, a wide range of hand crafted features were extracted and tested with Support Vector machine (SVM) and K-Nearest Neighbors. Results were comparable to those reported in literature.





## **Acknowledgments**

I would like to thank Eng. Ahmad Abdulkader, Facebook Inc., for his great insights in deep learning.

I would also like to extend my gratitude to Dr. Mohamed Nooman, IBM Inc., for his valuable thoughts and patience.



## **Dedication**

To my family who never stop supporting me and whose words of encouragement were pushing me to get the best out of myself. Thank you for never left my side.

To my mentor and supervisor Prof. Ahmed Morsy, thank you for your continuous support and great insights.

To Dr. Inas Yassine, thank you for being an amazing mentor and very supportive.

To my colleagues Mohamed Hisham, Eman Ayman, Tamim, Hady, and Attia, thank you for your support and encouragement.



# Table of Contents

<b>ACKNOWLEDGMENTS .....</b>	<b>I</b>
<b>DEDICATION.....</b>	<b>III</b>
<b>TABLE OF CONTENTS.....</b>	<b>V</b>
<b>LIST OF TABLES .....</b>	<b>VIII</b>
<b>LIST OF FIGURES .....</b>	<b>IX</b>
<b>NOMENCLATURE .....</b>	<b>XI</b>
<b>ABSTRACT .....</b>	<b>XIII</b>
<b>CHAPTER 1 : INTRODUCTION .....</b>	<b>1</b>
1.1.            MOTIVATION.....	1
1.2.            GOAL OF THIS RESEARCH .....	2
1.3.            STRUCTURE OF THE THESIS .....	2
1.4.            MAIN CONTRIBUTION OF THIS WORK.....	2
<b>CHAPTER 2 : LITERATURE REVIEW .....</b>	<b>3</b>
2.1.            SILENT SPEECH INTERFACE.....	3
2.1.1            Electroencephalogram (EEG) Signal .....	4
2.1.2.            Surface Electromyogram (sEMG).....	6
2.1.3.            Tongue based systems.....	13
2.1.3.1.            Electromagnetic Articulograph .....	13
2.1.3.2.            Ultrasound.....	16
2.1.4.            Non Audible Murmur (NAM) microphone.....	16
<b>CHAPTER 3 : MEDICAL AND TECHNICAL BACKGROUND .....</b>	<b>17</b>
3.1.            MUSCLE PHYSIOLOGY AND ANATOMY .....	17
3.2.            SURFACE ELECTROMYOGRAPHY ACQUISITION .....	18
3.2.1.            Equipment .....	18
3.2.2.            Electrodes.....	19
3.2.2.1.            Dry and gelled electrodes.....	19
3.2.2.2.            Electrode Properties .....	20
3.2.2.3.            Electrode Placement.....	21
3.2.3.            sEMG signal characteristics.....	21
3.2.3.1.            Factors affecting sEMG .....	21
3.2.3.2.            Noise affecting sEMG.....	22
3.2.4.            sEMG Preprocessing.....	23
3.2.4.1.            Filtering.....	23
3.2.4.2.            Normalization .....	23
3.3.            SPEECH PRODUCTION.....	23
3.3.1.            Human Speech Organs.....	23
3.4.            AUTOMATIC SPEECH RECOGNITION (ASR).....	23
3.5.            DEEP LEARNING.....	25

3.5.1.	Deep Neural Network .....	25
3.5.2.	Convolutional Neural Network .....	25
3.5.2.1.	Convolution Layer .....	26
3.5.2.2.	Filter Depth .....	26
3.5.2.3.	Filter Stride .....	26
3.5.2.4.	Zero Padding .....	26
3.5.3.	Non-Linear Activation Functions .....	26
3.5.3.1.	Sigmoid .....	27
3.5.3.2.	Tanh .....	28
3.5.3.3.	Rectified Linear Unit (ReLU) .....	28
3.5.4.	Fully Connected Layer .....	29
3.5.5.	Recurrent Neural Network (RNN) .....	30
3.5.6.	Long Short Term Memory Network .....	31
3.5.7.	Bidirectional LSTM .....	33
3.5.8.	Connectionist Temporal Classification Layer (CTC) .....	34
3.5.9.	Example For Training Deep Learning Model .....	34

## **CHAPTER 4 : SUBVOCAL SPEECH RECOGNITION USING ENGINEERED FEATURES VIA SURFACE ELECTROMYOGRAPH.....37**

4.1.	INTRODUCTION .....	37
4.2.	MATERIALS AND METHODS .....	38
4.2.1.	Corpus Design .....	38
4.2.2.	Subjects .....	38
4.2.3.	Signal Acquisition .....	39
4.2.4.	Experimental Protocol and Data Labeling .....	41
4.2.5.	Feature Extraction .....	41
4.2.5.1.	Time Domain Features .....	42
4.2.5.1.1.	Integrated EMG .....	42
4.2.5.1.2.	Mean Absolute Value (MAV) .....	42
4.2.5.1.3.	Simple Square Integral .....	42
4.2.5.1.4.	EMG Variance .....	42
4.2.5.1.5.	Root Mean Square .....	43
4.2.5.1.6.	Waveform Length .....	43
4.2.5.1.7.	Slope Sign Changes .....	43
4.2.5.1.8.	Skewness .....	43
4.2.5.1.9.	Kurtosis .....	43
4.2.5.1.10.	Auto Regressive Coefficients .....	44
4.2.5.2.	Frequency domain features .....	44
4.2.5.2.1.	Zero Momentum .....	45
4.2.5.2.2.	First, Second, and Third Momentum .....	45
4.2.5.2.3.	Central Frequency Variance .....	46
4.2.5.2.4.	Mean Power .....	46
4.2.5.2.5.	Total Power .....	46
4.2.5.2.6.	Median Frequency .....	46
4.2.5.2.7.	Mean Frequency .....	46
4.2.5.2.8.	Peak frequency .....	47
4.2.5.3.	Mel Frequency Cepstral Coefficients .....	48
4.2.6.	Classification Algorithms .....	49
4.2.6.1.	Support Vector Machine (SVM) .....	49
4.2.6.2.	K-Nearest Neighbors (KNN) .....	49
4.3.	RESULTS .....	49
4.3.1.	Time Domain Results .....	49
4.3.2.	Frequency Domain Results .....	54