UNIVERSITI TEKNOLOGI MARA

CONSONANTS RECOGNITION AND NOISE REDUCTION FOR ARABIC PHONEMES BASED MALAY SPEAKERS

ALI ABD ALMISREB

Thesis submitted in fulfillment
of the requirement for the degree of
Doctor of Philosophy

Faculty of Electrical Engineering

November 2015

AUTHOR'S DECLARATION

I declare that the work in this thesis was carried out in accordance with the regulations of Universiti Teknologi MARA. It is original and is the results of my own work, unless otherwise indicated or acknowledged as referenced work. This thesis has not been submitted to any other academic institution or non-academic institution for any degree or qualification.

I, hereby, acknowledge that I have been supplied with the Academic Rules and Regulations for Post Graduate, Universiti Teknologi MARA, regulating the conduct of my study and research.

Name of Student	:	Ali Abd Almisreb
Student I.D. No.	:	2012281702
Programme	:	Doctor of Philosophy (Electrical Engineering)
Faculty	:	Electrical Engineering
Thesis Title	:	Consonants Recognition and Noise Reduction for Arabic Phonemes based Malay Speakers
Signature of Student	:	....
Date	:	November 2015

ABSTRACT

Arabic phonemes can be categorised into 28 consonants. The variations in each phoneme and vowel cause difficulties for the non-native Arabic speakers, particularly the Malay speakers, to pronounce these letters correctly. Hence, in this thesis, noise reduction and consonants recognition are conducted among the Malay speakers. The Malay race has been chosen due to the high usage of the Arabic language for reciting Al-Quran. Generally, the study is divided into two parts, namely, the study of noise reduction and consonant recognition. First, two noise removal methods were developed. The first method is based on combining Negative function with Gamma correction function. The second noise reduction method is addressed by utilising 2D Gabor filter. Furthermore, the consonant study was conducted based on Automatic Speech Recognition (ASR) system concept. The ASR composes of feature extraction stage followed by speech recognition. On the other hand, the feature extraction was implemented by investigating three different methods, namely, Mel-Frequency Cepstrum Coefficients (MFCC), Linear Prediction Coefficients (LPC) and Perceptual Linear Prediction (PLP). Finally, the speech recognition process was conducted by utilising three methods: Dynamic Time Warping (DTW), Artificial Neural Network (ANN) and Deep Neural Network (DNN). Experimental analysis and results showed that the proposed noise reduction methods have advantages over the traditional methods in terms of the consonant waveforms enhancement quality and the computational time as well. The MFCC has shown better performance compare to LPC and PLP as a feature extraction technique. Additionally, the comparison between DTW and ANN has proven that the ANN more suitable for Arabic consonant recognition. On the other hand, the joining of ANN and DTW has worked optimally as well. Lastly, the DNN are the most suitable methods for recognition process of Arabic consonants based on Malay speakers' usage.

TABLE OF CONTENTS

CHAPTER ONE
INTRODUCTION

1.1 BACKGROUND STUDY

Speech is the physical process which produces sound using the tongue, lips, palate and respiratory system for transmitting thoughts. Humans develop speech by using three parts; tongue, teeth and lips to control and push the air through the oral cavity. Human lips, tongue and teeth all interact with each other to transform the air that comes from the lungs into speech sounds and ultimately the uttered words. The speech sounds created by the lips, tongue and teeth are called consonants, where they are made by redirecting the air stream produced by our tongue, teeth, and lips. In contrast, vowels are mostly produced with an open vocal tract [1]–[3].

Speech is characterised by features such as pitch, duration, amplitude (loudness, signal strength, power/energy), and the phase of each frequency component. As it occurs, only the first three features are relevant from a speech recognition perspective, given that the human ear is insensitive to phase [4]. Since phonemes, the basic linguistic unit, are characterised by frequency, time, and energy, it makes more sense to use the three-dimensional spectrograms rather than processing the raw (albeit filtered) time-varying speech waveform in this study.

Additionally, speech recognition is a technology commonly used in several applications daily. In other words, it is a subsequent step used in recognising speech, which has been investigated since the 1950s. Speech recognition can be considered as a new engineering science and with the current technological advancement, speech recognition technique can attain 98% to 99% accuracy under perfect environment conditions [5]. The importance of any speech recognition system relies on its straightforwardness, where this naivety together with ease of using a device-based speech has lots of preferences [6].