Classroom Speech Intelligibility Prediction System for Front-Rear Speech Amplified Classroom Based On Audio Features

by

Mohammad Ridhwan Bin Tamjis
(0930610371)

A thesis submitted in fulfillment of the requirements for the degree of Master of Science (Mechatronic Engineering)

School of Mechatronic Engineering
UNIVERSITI MALAYSIA PERLIS

2012

UNIVERSITI MALAYSIA PERLIS

DECLARATION OF THESIS

Author's full name : Mohammad Ridhwan Bin Tamjis

Date of birth : 1st June 1986

Title : Classroom Speech Intelligibility Prediction System for Front-Rear Speech Amplified Classroom Based on Audio Features

Academic Session : 2/2012

I hereby declare that the thesis becomes the property of Universiti Malaysia Perlis (UniMAP) and to be placed at the library of UniMAP. This thesis is classified as :

☐	**CONFIDENTIAL**	(Contains confidential information under the Official Secret Act 1972)*
☐	**RESTRICTED**	(Contains restricted information as specified by the organization where research was done)*
/	**OPEN ACCESS**	I agree that my thesis is to be made immediately available as hard copy or on-line open access (full text)

I, the author, give permission to the UniMAP to reproduce this thesis in whole or in part for the purpose of research or academic exchange only (except during a period of _____ years, if so requested above).

Certified by:

Ridhz

SIGNATURE

Paul MP

SIGNATURE OF SUPERVISOR

860601-38-5437

(NEW IC NO. / PASSPORT NO.)

Assoc. Prof. Dr. Paulraj Murugesa Pandiyan

NAME OF SUPERVISOR

Date : 24th July 2012

Date : 24th July 2012

NOTES : * If the thesis is CONFIDENTIAL or RESTRICTED, please attach with the letter from the organization with period and reasons for confidentially or restriction.

ACKNOWLEDGEMENT

There are a lot of people that helped in making this thesis a reality; however, I can only mention a few of them here.

First of all, all praise goes to the Almighty for the endless blessings and kindness that make me who I am today. Without HIS blessings, I may never find my path of glory.

I am very thankful to the Almighty to bless me with many of wonderful people in my life. The greatest blessing above all is my beloved parent and family, who have never tired of supporting me in finding my way to a better future, who have walked me through joy and sorrow, and taught me to never give up in everything that I do, no matter how much pain it will cost.

I am also fortunate to have been blessed with such supportive research supervisors and mentors namely Assoc. Prof. Dr. Paulraj Murugesa Pandiyan, Prof. Dr. Sazali Yaacob, Assoc. Prof. Ahmad Nazri Abdullah, and Prof. Dr. Raymond Boon Whee Heng. I owe them great deal in my path of success, and owe them great in surviving this adventure.

Throughout the study, I've met many kinds of people, friends and foe. I am grateful to them, as they brought joy and adventure to my life. I wish to extend my gratitude to my friends for accompanying me in the race to success.

And not to forget, I want to extend my appreciation to the appointed examiners for your willingness to evaluate my research works.

In the spirit of comradeship, I would like to convey my thanks to the Intelligent Research Cluster (formerly Acoustic Research Cluster) members for every help I received in completing my research.

And finally, I would like to thanks the administration of Universiti Malaysia Perlis, especially the School of Mechatronic Engineering for giving me the opportunities to further my study in this respected university, as well as for the hospitalities I've received.

May Allah bless you all with good and successful life

Yours sincerely,

Ridhz

Mohammad Ridhwan bin Tamjis

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

SI	Speech intelligibility
RT	Reverberation time
SC	Spectral centroid
SR	Spectral rolloff
SNR	Signal-to-noise ratio
LAeq	A-weighted equivalent sound level
AI	Articulation index
SIL	Speech interference level
PO	Power
ZCR	Zero-crossings rate
STE	Short time energy
FFNN	Feed forward neural network
EN	Elman network
UniMAP	Universiti Malaysia Perlis
ANOVA	Analysis of variance
GUI	Graphical user interface

LIST OF SYMBOLS

W	Width
L	Length
H	Height
$M(f)$	Modulation index
$H(t)$	Impulse response
N	Modulation frequency number
$Kurt$	Kurtosis
Var	Variance
γ_1	Skewness
$\bar{x}$	Mean
σ	Standard deviation
$Q1$	Lower quartile
$Q2$	Median quartile
$Q3$	Upper quartile
H_o	Null hypotheses
H_1	Alternative hypotheses
$X_i(n)$	Signal amplitude
C	Percent of the magnitude distribution of the discrete Fourier transform
F_c	Center frequency
X	Input neuron
V	Weight between input and hidden layer

w	Weight between hidden and output layer
Z	Hidden neuron
Y	Output neuron
X	x-coordinate of listening position
Y	y-coordinate of listening position

Sistem Ramalan Darjah Kejelasan Pertuturan Dalam Kelas Bagi Kelas Yang Diperkuatkan Ucapan Hadapan-Belakang Berasaskan Ciri-ciri Audio

ABSTRAK

Kebolehfahaman pertuturan bilik darjah adalah ukuran bagaimana ucapan atau perkataan difahami di dalam kelas. Ia adalah satu ukuran kualiti ucapan di dalam kelas. Pelbagai kaedah telah dicadangkan oleh beberapa penyelidik untuk meningkatkan kebolehfahaman pertuturan. Walau bagaimanapun, kaedah yang dicadangkan berkesan hanya dalam peringkat reka bentuk bilik darjah, kerana pelaksanaan dalam bilik darjah 'lama' adalah mahal dan memakan masa. Oleh itu, amplifikasi ucapan dilaksanakan untuk menangani masalah tersebut.Terdapat beberapa kaedah yang dicadangkan oleh pakar audio tentang bagaimana untuk menetapkan sistem di dalam kelas, untuk memaksimumkan kebolehfahaman pertuturan. Walau bagaimanapun, kaedah yang dicadangkan memakan masa dan agak rumit. Jadi, sebagai alternatif, penyelidikan ini telah mencadangkan audio-ciri ucapan kebolehfahaman sistem berasaskan ramalan. Matlamat kajian ini adalah untuk membangunkan sistem ramalan ucapan kebolehfahaman pintar dengan menggabungkan ciri-audio (rolloff spektrum (SR), sentroid spektrum (SC), kuasa (PO), kadar sifar-lintasan (ZCR), tenaga dalam masa yang singkat (STE)) dan Pengelas (rangkaian neural penghantar kehadapan (FFNN), rangkaian ELMAN (ENN)). Bagi mencapai matlamat tersebut, kajian ini telah mengumpul sampel data yang terdiri daripada rakaman ucapan dalam bilik darjah yang diperkuatkan ucapan, serta sifat-sifat fizikal. Pengukuran itu dilakukan dengan lapan bilik darjah yang berlainan di UniMAP, dan protokol pengukuran diperoleh daripada kajian terdahulu dan standard akustik. Data yang dikumpul telah dianalisis menggunakan kaedah statistik seperti analisis deskriptif dan ANOVA. Data telah diproses terlebih dahulu untuk membantu proses pengekstrakan ciri audio kemudian.Isyarat sebelum diproses kemudiannya menjalani proses pengekstrakan ciri untuk mengekstrak ciri-ciri audio. Dalam kajian ini, lima jenis ciri-ciri audio telah dipilih, dan setiap ciri ini kemudiannya digabungkan dengan data ciri fizikal bilik darjah sebagai input Pengelas bereksperimen. Hasilnya, didapati bahawa ciri-ciri audio PO menghasilkan ketepatan yang terbaik, tanpa mengira jenis Pengelas apabila dibandingkan dengan ciri-ciri lain. Di akhiran, sistem antara muka untuk ciri audio berasaskan bilik darjah ucapan ramalan sistem kebolehfahaman dibangunkan. Tambahan pula, pangkalan data kelas pengukuran kebolehfahaman pertuturan menggunakan mikrofon tunggal telah disusun.

Classroom Speech Intelligibility Prediction System for Front-Rear Speech Amplified Classroom Based On Audio Features

ABSTRACT

Classroom speech intelligibility is a measure of how well a speech or word is understood in the classroom. It is a measure of the speech quality in the classroom. Numbers of methods have been proposed by various researchers to improve the speech intelligibility. However, the proposed methods are effective only in the design stage of the classroom, as implementation in the 'old' classroom is costly and time consuming. Thus, speech amplification is implemented to tackle such problems. There are methods suggested by audio expert on how to properly setup the system in the classroom, in order to maximize the speech intelligibility. However, the methods are rather complicated and time consuming. So, as an alternative, this research has proposed an audio-feature based speech intelligibility prediction system. The goal of this research is to develop an intelligent speech intelligibility prediction system by combining audio-features (spectral rolloff (SR), spectral centroid (SC), power (PO), zero-crossings rate (ZCR), and short time energy (STE)) and classifiers (feed forward neural network (FFNN), Elman network (ENN)). To achieve the goal, this research has collected data samples which comprises of speech recordings in the speech amplified classrooms, as well as the physical properties. The measurement was done in eight different classrooms in UniMAP, and the measurement protocol was derived from the previous researches and acoustic standards. The data collected were then analyzed using statistical approach, such as descriptive analysis and ANOVA. The data were then pre-processed to assist the later feature extraction process. The pre-processed signals were then undergone feature extraction process to extract the audio features. In this research, five types of audio features have been selected, and each feature is then combined with the classroom's physical feature data as inputs of the experimented classifiers. As a result, it was found that audio feature PO yield the best accuracy, regardless the type of classifiers when compared to the other features. At the end, the interface system for the audio feature-based classroom speech intelligibility prediction system is developed. Moreover, a database of classroom speech intelligibility measurement using single microphone was compiled.

CHAPTER 1

INTRODUCTION

1.1 Research Overview

Speech intelligibility (SI) can be defined as the match between the intention of the speaker and the response of the listener to the speech passed through the transmission system (Kent, 1992). For any classroom or teaching facility, it is necessary to make sure that sound is distributed sufficiently to all listeners in order to have optimum SI (Nabelek & Nabelek, 1985). Optimum SI is considered achieved when a listener hears the words correctly uttered by a speaker and the word is not mistaken with any other word (Weil, 2003). For example, a person might speak with an accent but still be understood by a listener who is face-to-face with the talker. The accent might be a distortion in communication, but as long as the listener understands the message, the SI is established. Several studies reported that classrooms with high level of SI yields high performances students (Ross & Giolas, 1971; Elliot, 1979; Lukas et al., 1981; Bradley, 1986; Berg et al., 1996).

Methods have been introduced to enhance the SI in the classroom (ASHA, 1995; ANSI, 2002). However, most of the methods presented are only cost effective when implemented at the stage of classroom design (ATS&R, 2005). Renovation of an established classroom requires high expenditure. Moreover, the acoustical requirements such as the length, height, and width of the classroom must be chosen carefully to obtain classroom with optimal SI.

As acoustic renovation is expensive, people opted into speech amplification. The speech amplification system implements the use of microphone and loudspeakers to ‘amplify’ the speech, or any words spoken by the lecturer through the microphone. The ultimate goal of amplifying the speech in the classroom is to make sure that the broadcasted speech is equally distributed across the classroom (Berg, 1993). It is the simplest and the easiest way to enhance the learning experience in the classroom, and was found effective (Bess et al., 1984; Ross, 1986; Crandell & Smaldino, 1992). However, it is important to note that speech amplification does not solely amplify the speech, as it also amplifies the noise in the process (Crandell et al., 2005). Thus, a proper setup for the amplification system is required. Several methods have been suggested by audio experts (Davis et al. 1988; Infrastructure, 2004; Rives, 2008). However, the method is rather complicated and time consuming.

As an alternative, this research proposed an audio feature-based classroom SI prediction system for speech amplified classroom. In this research, audio samples of different signal-to-noise ratio level were collected using sound level meter. Statistical analysis and subjective measurement were done to validate the data collected. Extraction of audio features was done from the collected audio samples and was used as parts of the input features of experimented classifiers. In the end, a graphical user interface (GUI) system has been developed using the best trained classifier to represent the audio feature-based classroom SI prediction system for speech amplified classroom.

1.2 Problem Statement

There are number of problems found that motivates this study. First is the need for acoustic assessments in Universiti Malaysia Perlis (UniMAP). UniMAP campus is still under development. Currently, many of the teaching facilities are situated in the rental-basis building, which either a multifunctional hall or renovated shop. Next is the complexity of acoustic assessment technique. Acoustic assessment is rather complicated as it requires careful planning, considerations, the right equipments, and time consuming (ANSI, 2002). Thus, to assist in classroom design stage, intelligibility prediction technique is proposed. However, in literature, many of the studies concentrated on the conventional type classroom which does not require the speech amplification system (Hodgson et al., 1998; Bistafa & Bradley, 2000; Bistafa & Bradley, 2001).

1.3 Objectives of the Research

The objectives of the research are as follow:

(1) To assess the quality of speech in the sound amplified university classrooms by collecting real time data using single microphone based on acoustic standards.

(2) To identify the best audio feature for audio feature-based SI prediction, by measuring the performance of different type of audio feature.

(3) To develop a graphical user interface (GUI) system for audio feature-based classroom SI prediction using the best trained classifier, in MATLAB platform.

1.4 Scope

The main scope of this research is on the speech amplified classrooms, which can be found in most university classrooms. This research implemented the use of electro-acoustic equipments such as loudspeakers and amplifiers during the data collection. With the data collected, the research applied several signal processing technique to extract the audio features. Comparisons were made to determine the best audio features and for SI prediction.

1.5 Thesis Organization

This dissertation is organized into six chapters, and the summary of each chapter is as follows:

Chapter 1 provides a brief introduction of the research. This chapter also discuss on the problem statements, the objective of the research, the research contribution as well as overview on how the whole thesis is organized.

Chapter 2 describes the theoretical studies and research background for this thesis. It starts with the basic theory on human hearing to the perception of sound in the classroom. This chapter also discuss on the previous researches related to the thesis title.

Chapter 3 concentrates on the data collection process in the research. Every detail in the data collection such as measurement protocols and the instruments used is discussed in detail in this chapter. The subjective measurement is also described at the end of the chapter.

Chapter 4 describes the signal processing, feature extraction, and classification. In this chapter, the signal processing and feature extraction technique used in the research is briefly described. This chapter also describes the classifiers algorithms that were used in this research.

Chapter 5 discuss on the results obtained in the research. The results include data statistic and audio feature performances. Comparison on the performance of each audio feature was made and the developed Graphical User Interface (GUI) is briefly elaborated.

Chapter 6 concludes the research. In this chapter, the limitations and recommendations for future work is briefly discussed.

CHAPTER 2

INTRODUCTION TO CLASSROOM ACOUSTICS

2.1 Introduction

This chapter will describe the studies related to the studies on SI. The chapter starts with the fundamental of human hearing and speaking, where the structure and mechanism are briefly discussed. In addition, the fundamental of human speech is also described.

The chapter will then continue with the studies on SI. In this section, the problems in classrooms acoustics are highlighted and the previous studies on SI are discussed.

This chapter will also describe the factors that affect the SI and standards and guidelines that have been developed to control SI in the classrooms. The SI evaluation method is also described in this chapter, which focused on the calculation of Speech Transmission Index (STI) and Rapid Speech Transmission Index (RASTI).

The final part of this chapter will discuss on the loudspeaker arrangement in the classroom. There are two type of arrangements proposed, and both arrangement are compared. At the end, the summary of this chapter is presented.