



**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**

## 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 Murugesu 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 thank 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,



Mohammad Ridhwan bin Tamjis

© This item is protected by original copyright

## TABLE OF CONTENTS

	<b>PAGE</b>
<b>THESIS DECLARATION</b>	i
<b>ACKNOWLEDGEMENT</b>	ii
<b>TABLE OF CONTENTS</b>	iv
<b>LIST OF TABLES</b>	x
<b>LIST OF FIGURES</b>	xii
<b>LIST OF ABBREVIATIONS</b>	xiii
<b>LIST OF SYMBOLS</b>	xiv
<b>ABSTRAK</b>	xvi
<b>ABSTRACT</b>	xvii
<b>1. INTRODUCTION</b>	
1.1 Research Overview	1
1.2 Problem Statement	3
1.3 Objectives of the Research	3
1.4 Scope	4
1.5 Thesis Organization	5
<b>2. INTRODUCTION TO CLASSROOM ACOUSTICS</b>	
2.1 Introduction	6
2.2 Human Acoustic System	7
2.2.1 Human Hearing System	7

2.2.2	Voice Production System	9
2.2.3	Speech for Communications	10
2.3	Acoustic Related Problems in Classrooms	11
2.4	Studies on Classroom Speech Intelligibility	12
2.5	Classroom's Acoustical Measurements	15
2.6	Factors Affecting Speech Quality	17
2.7	Classroom Acoustic Standards	18
2.8	Speech Transmission Index (STI) and Rapid Speech Transmission Index (RASTI)	21
2.9	Classroom Speech Amplification	23
2.9.1	Loudspeaker Arrangement	24
2.10	Summary	25
<b>3.</b>	<b>DATA COLLECTION FOR CLASSROOM SPEECH INTELLIGIBILITY PREDICTION</b>	
3.1	Introduction	27
3.2	Objective Measurement	28
3.2.1	Measurement Procedure	28
3.2.1.1	Spectrum of the Speech Signal	29
3.2.1.2	Spectrum of the Environmental Noise	29
3.2.1.3	Classroom Reverberation	29
3.2.1.4	Associated selection of listener position	30
3.2.1.5	Evaluation of the resulting intelligibility score	30
3.3	Classrooms Descriptions	31

3.3.1	Classroom Criteria	31
3.3.2	Listening Positions	32
3.4	Measuring Instruments	32
3.4.1	01dB-Metravib Solo Sound Level Meter	32
3.4.2	PRE21W Preamplifier	34
3.4.3	Calibrators	34
3.4.4	Loudspeakers	35
3.4.5	Audio Mixer	36
3.4.6	Power Amplifier	37
3.4.7	Limitations in Classrooms and Measuring Instruments	38
3.5	Equipments Setup	38
3.5.1	Loudspeaker setup	38
3.5.2	Microphone setup	41
3.6	Measurement procedure	42
3.7	Subjective Measurement	46
3.7.1	Description of Volunteers	46
3.7.2	Placements	46
3.7.3	Measurement Procedure	47
3.8	Related Standards	48
3.9	Summary	49

<b>4.</b>	<b>DATA PREPROCESSING, FEATURE EXTRACTION, AND CLASSIFICATION</b>	
4.1	Introduction	51
4.2	Data Pre-processing	53
4.2.1	Data Analysis	54
4.2.2	Data Cleaning	56
4.2.3	Analysis of Variance (ANOVA)	58
4.3	Signal Pre-processing	58
4.3.1	Pre-emphasizing	59
4.3.2	Signal segmentation and windowing	61
4.4	Feature extraction	62
4.4.1	Short time energy	62
4.4.2	Zero-crossing rate	61
4.4.3	Spectral roll off	64
4.4.4	Spectral centroid	65
4.4.5	Power	66
4.5	Data Integration	67
4.6	Speech Intelligibility Classification	67
4.5.1	Data Preparation	68
4.4.2	Type of classifiers	69
4.4.2.1	Feed-forward neural network	69
4.4.2.2	Elman Neural Network	76
4.7	Summary	81

<b>5.</b>	<b>RESULTS AND DISCUSSIONS</b>	
5.1	Introduction	83
5.2	Classroom Characteristic	84
5.3	Data Analysis	87
5.3.1	Descriptive statistic	87
5.3.2	Data distribution	89
5.3.3	Outliers	89
5.3.4	Analysis of variance	92
5.4	Subjective Test Results	94
5.5	Speech Intelligibility Classification	95
5.5.1	Classification of speech intelligibility using FFNN	95
5.5.2	Classification of speech intelligibility using ENN	99
5.5.3	Performance comparison audio features	102
5.5	Classroom Speech Intelligibility Prediction System	104
5.6	Summary	105
<b>6.</b>	<b>CONCLUSIONS AND FUTURE WORK</b>	
6.1	Conclusion	107
6.2	Limitation	109
6.3	Research Findings	109
6.4	Recommendations for Future Work	110
	<b>REFERENCES</b>	112
	<b>APPENDIX A</b>	119
	<b>APPENDIX B</b>	120

**APPENDIX C**

122

**LIST OF PUBLICATIONS**

125

© This item is protected by original copyright

## LIST OF FIGURES

NO.		PAGE
2.1	Human ear structure (TutorVista.com <sup>tm</sup> , 2010)	7
2.2	Sectional view of human upper body, showing the important elements in voice mechanism. (Connelly, 2005)	9
2.3	Flow chart of the STI calculation scheme (Steeneken et al., 2002)	23
3.1	Sound level meter	33
3.2	CAL 21 sound calibrator	34
3.3(a)	10-inch loudspeaker	35
3.3(b)	DO12 Omni directional loudspeaker	35
3.4	The front and rear view of EUROPOWER PMH 2000 audio mixer	36
3.5	Interface scheme for audio mixer	37
3.6	The front and rear view of Inter-M 500 power amplifier	37
3.7(a)	Front-rear distribution	39
3.7(b)	Front-center distribution	39
3.8	Basic front-rear speaker connection and microphone placement	40
3.9	Process flow for the objective measurement procedure	44
3.10	The process flow of the subjective measurement	48
4.1	Schematic diagram for development stage methodology	52
4.2	Flowchart of data pre-processing stages	54
4.3(a)	Box plot for variable A	57
4.3(b)	Box plot for variable B	57
4.4	Frequency response curve of pre-emphasis filter	59
4.5	Plot of original and pre-emphasized signal	60

4.6	Plot of short time energy for frame reduction process of the word 'gam'	61
4.7	Plot of ZCR for sample signal 'gam'	63
4.8	Plot of SR for sample signal 'gam'	64
4.9	Plot of SC for sample signal 'gam'	65
4.10	Calculation scheme for PO	66
4.11	Block diagram of proposed system used for FFNN and ENN	67
4.12	Fully connected feed-forward network with one hidden layer and one output layer.	70
4.13	Binary sigmoid	72
4.14	Bipolar sigmoid	72
4.15	Fully connected EN with one hidden layer and one output layer.	77
5.1	Typical view of UniMAP classrooms	86
5.2(a)	Box-plot by group of cases STI	90
5.2(b)	Box-plot by group of cases RASTI	90
5.3	Box-plot of separate variables, where '+' mark represents outlier	91
5.4	Box plot of separate variables, where outliers have been removed	92
5.5	Mean score results for subjective speech intelligibility test	94
5.6	Performance comparisons of audio features	103
5.7	Initial layout of the developed GUI	104
5.8	Example of classroom speech intelligibility prediction using GUI	105
6.1	Flow chart of overall research methodology	108

## LIST OF TABLES

NO.		PAGE
2.1	Summaries of acoustic measurements done by previous researchers	16
2.2	Recommended minimal performance ratings for intelligibility and vocal effort in four applications (ISO, 2003)	19
2.3	STI and RASTI classification	21
2.4	Comparison of central and distributed cluster for loudspeaker placement (Foreman, 1987)	25
3.1	International sound level meter standards	33
3.2	Example of objective measurement results	45
5.1	Measured classrooms dimensions and average reverberation time	84
5.2	Descriptive statistic of the objective measurement	88
5.3	Analysis of skewness and kurtosis	89
5.4	Analysis of variance for STI and RASTI	93
5.5	Parameters and results for STE trained using FFNN	95
5.6	Parameters and results for ZCR trained using FFNN	96
5.7	Parameters and results for SR trained using FFNN	97
5.8	Parameters and results for SC trained using FFNN	97
5.9	Parameters and results for PO trained using FFNN	98
5.10	Parameters and results for STE trained using ENN	99
5.11	Parameters and results for ZCR trained using ENN	100
5.12	Parameters and results for SR trained using ENN	100
5.13	Parameters and results for SC trained using ENN	101
5.14	Parameters and results for PO trained using ENN	102

## LIST OF ABBREVIATIONS

SI	Speech intelligibility
RT	Reverberation time
SC	Spectral centroid
SR	Spectral rolloff
SNR	Signal-to-noise ratio
L <sub>Aeq</sub>	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
$\gamma_1$	Skewness
$\bar{x}$	Mean
$\sigma$	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

© This item is protected by original copyright

## **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.

© This item is protected by original copyright

## 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.

## 2.2 Human Acoustic System

Human acoustic system comprise of two major parts; the hearing and voice production system. Each of the system is connected to the brain, and co-related to each other. Combined with the ability of producing speech, these two parts became the most important tools in human communication. The mechanics of each system is described in the following subsections.

### 2.2.1 Human Hearing System

One of the most important mechanical structures in human communication is the ear. It is a delicate, and the most intricate of human body which consists of three main parts namely; the outer, middle and inner ears. Figure 2.1 depicts the structure of human ear.

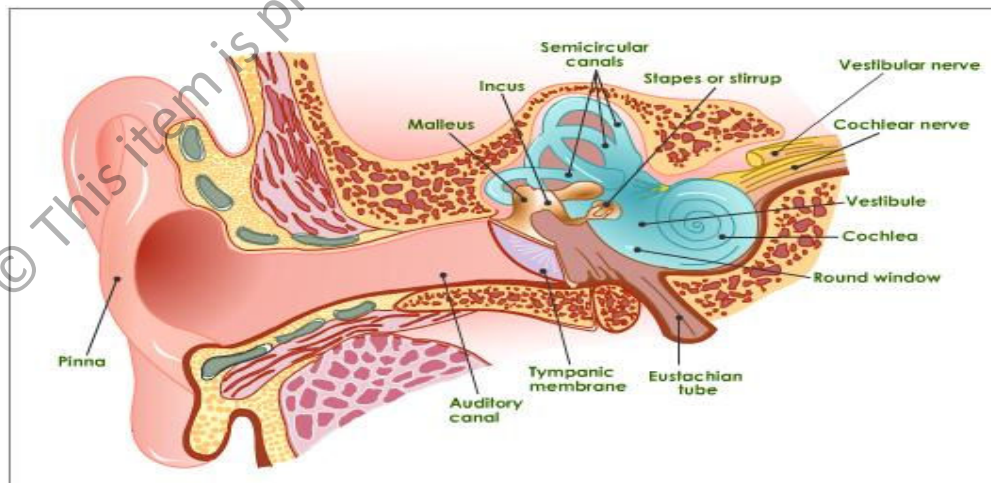


Figure 2.1: Human ear structure (TutorVista.com<sup>tm</sup>, 2010)

The outer part of ear is consists of pinna and ear lobe. These parts work as the horn collecting sound to the auditory canal. The ear lobe might help in warming up the