

ANALYSIS OF POWER SPECTRUM DENSITY OF MALE SPEECH AS  
INDICATORS FOR HIGH RISK AND DEPRESSED DECISION

By

Nik Nur Wahidah Nik Hashim

Thesis

Submitted to the Faculty of the  
Graduate School of Vanderbilt University  
in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in

Electrical Engineering

May, 2011

Nashville, Tennessee

Approved:

Associate Professor Mitch Wilkes, Ph.D.

Associate Professor Ronald Salomon, M.D.

# TABLE OF CONTENTS

ACKNOWLEDGEMENTS .....	iv
LIST OF TABLES .....	v
LIST OF FIGURES .....	vii
Chapter	
I. INTRODUCTION .....	1
1.1 Background .....	1
1.2 Statement of purpose .....	3
1.3 Description of chapters .....	5
II. BACKGROUND AND SIGNIFICANCE .....	7
2.1 Mechanism and physiological aspect of speech production .....	8
2.1.1 Voiced sound .....	13
2.1.2 Unvoiced sound .....	14
2.2 Speech production model .....	14
2.2.1 The source excitation .....	17
2.2.2 The vocal tract filter .....	18
2.2.3 Lips radiation .....	19
2.3 Source filter separation .....	20
III. LITERATURE REVIEW .....	22
IV. METHODOLOGY .....	26
4.1 Database collection .....	26
4.2 Data pre-processing .....	27
4.2.1 Voiced, unvoiced and silence detection .....	28
4.3 Feature extraction .....	30
4.3.1 Identifying number of spectral features/bands .....	31
4.3.2 Feature extraction for eight equal bands of energy spectrum .....	31
4.4 Feature analysis and classification .....	34
4.4.1 Quadratic and linear classifier .....	35
4.4.2 Methods of resampling .....	37

V. APPLICATIONS AND RESULTS .....	40
5.1 Analysis of Male Interview Speech for High Risk/Depressed Decision classification .....	40
5.1.1 PSD Statistical Analysis on Interview Speech .....	40
5.1.2 The number of frequency sub-bands effect on classification .....	41
5.1.3 Feature classification using all patients' data .....	42
5.1.4 Improving results for Jackknife and Cross-validation classification .....	46
5.2 Analysis of Male Reading for the High Risk/Depressed Decision .....	49
5.2.1 PSD Statistical Analysis of the Reading Speech .....	49
5.2.2 Feature classification using all patient's data .....	50
5.2.3 Improving results for Jackknife and Cross-validation classification .....	51
 VI. CONCLUSION AND DISCUSSION .....	 55
 APPENDIX .....	 59
A.1 Other results for male interview suicidal/depressed .....	59
A.2 Other results for male reading suicidal/depressed .....	64
A.3 Classification results on male interview suicidal/remitted and depressed/remitted .....	66
A.4 Matlab code .....	67
main.m .....	68
simplevuv.m .....	69
split20sec.m .....	71
mean_energy.m .....	72
pdgm_sum.m .....	73
ratio_collect.m .....	76
getalldepresseddata.m .....	77
getallhighriskdata.m .....	77
getallremitteddata.m .....	77
equal_test_train.m .....	78
jackknife.m .....	80
distance.m .....	84
getlabel.m .....	86
crossval.m .....	87
errorhist.m .....	91
percentMean.m .....	93
 REFERENCES .....	 94

## ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere gratitude to my adviser, Professor Mitch Wilkes for his continuous encouragement, motivation, ideas, advices, and patience to help me understand the material and helping me with finishing the thesis. The enthusiasm he has for his research and teaching was contagious and motivational for me. I could have never imagined having a better adviser than him.

Besides my adviser, my sincere thanks also go to Dr. Salomon and The Vanderbilt University Department of Psychiatry for providing valuable database for us to use in this study. I am also very grateful that Dr. Salomon was willing to spend his time to read this thesis even on such a short notice.

I would also like to thank my sponsor, Ministry of Higher Education Malaysia (MoHE) and Islamic International University of Malaysia (IIUM) for providing financial support and giving me this opportunity to pursue my studies here at Vanderbilt University.

I thank my fellow lab mate, Wan Ahmad Hasan for the wonderful support and stimulating discussions we had all throughout these two years.

Last but not least, I would like to thank my mother and father for their never-ending prayer, love and support. Without their trust, I could have never reach this far.

## LIST OF TABLES

Table	Page
4.1 The number of patients (male) according to groups .....	27
4.2 Frequency range for each band levels .....	19
4.3 The total number of 20-seconds segments according to groups .....	30
4.4 PSD based feature extraction .....	32
5.1 Mean and standard deviation collected from the spectral energy ratios in male interview speech .....	40
5.2 High risk/depressed features classification using all patients' data .....	43
5.3 Reasonable Equal-Test-Train classification results using all patients' data .....	45
5.4 High risk/depressed features classification after removing outliers .....	48
5.5 Mean and standard deviation collected from the spectral energy ratios in male reading Speech .....	50
5.6 High risk/depressed features classification using all patients' data .....	51
5.7 High risk/depressed features classification after removing two high risk patients and one depressed patients (outliers) .....	53
a1 Percentage of Equal-Test-Train classification for male interview suicidal/depressed using all patients' data .....	59
a2 Percentage of Jackknife classification for male interview suicidal/depressed using all patients' data .....	60
a3 Percentage of Cross-validation classification for male interview suicidal/depressed using all patients' data .....	61
a4 Percentage of Cross-validation classification for male interview suicidal/depressed after removing outliers .....	62
a5 Percentage of Jackknife classification for male interview suicidal/depressed after removing outliers .....	63

a6	Percentage of classification for male reading suicidal/depressed using all patients' data .....	64
a7	Percentage of Cross-validation and Jackknife classification for male reading suicidal/depressed after removing outliers.....	65
a8	Percentage of Equal-Test-Train classification for male interview suicidal/remitted and depressed/remitted using all patients' data .....	66
a9	Percentage of Jackknife classification for male interview suicidal/remitted and depressed/remitted using all patients' data .....	66
a10	Percentage of Cross-validation classification for male interview suicidal/remitted and depressed/remitted using all patients' data .....	66
a11	Steps for features extraction and analysis using Matlab .....	67

## LIST OF FIGURES

Figure	Page
2.1 Cross-sectional view of an anatomy structure for human vocal production.....	8
2.2 Simplified speech production model .....	9
2.3 Example of voice spectrum frequency.....	11
2.4 (a) Speech waveform produced by word “please” .....	12
(b) Voiced speech segment from word “please” .....	12
(c) Unvoiced speech segment from word “please” .....	12
2.5 EGG signal corresponding to opening and closing of the glottis, DEGG Derivative of the signal, and smoothed DEGG .....	13
2.6 Vocal tract organ pipe model.....	14
2.7 Linear filter of a voice production model .....	15
2.8 Source-filter model of speech production.....	16
2.9 Time and frequency domain representation of glottal pulses .....	17
2.10 Concatenation of lossless tubes for $N = 5$ .....	18
4.1 Flow chart for the PSD feature extraction .....	33
5.1 Plot of patient’s distribution on $PSD_2$ with respect to $PSD_3$ and a decision boundary using Equal-Test-Train quadratic classifier with all interview patients’ data.....	44
5.2 Error histogram from the interview speech for high risk patients using $PSD_2$ and $PSD_3$ .....	47
5.3 Error histogram from the interview speech for depressed patients using $PSD_2$ and $PSD_3$ .....	47
5.4 Plot of patient’s distribution on $PSD_2$ with respect to $PSD_3$ and a quadratic decision boundary using Equal-Test-Train quadratic classifier after removing two high risk patients (outliers) .....	49

5.5	Error histogram from reading speech for the high risk patients using PSD <sub>1</sub> to PSD <sub>3</sub> with linear classifier .....	52
5.6	Error histogram from reading speech for the depressed patients using PSD <sub>1</sub> to PSD <sub>3</sub> with linear classifier .....	52
5.7	Plot of patient's distribution on PSD <sub>1</sub> with respect to PSD <sub>2</sub> and a decision boundary using Equal-Test-Train linear classifier after removing two high risk patients and one depressed patient (outliers) .....	54
5.8	Plot of patient's distribution on PSD <sub>1</sub> with respect to PSD <sub>2</sub> and a decision boundary using Equal-Test-Train quadratic classifier after removing two high risk patients and one depressed patient (outliers) .....	54

# CHAPTER I

## INTRODUCTION

### *1.1 Background*

Over the past decades, suicide has been a major public health problem worldwide and it has now become one of the country's worst public health problems. Referring to suicide statistics for the year 2007, there were 34,598 reported suicide deaths in the United States. Suicide is the fourth leading cause of death for adults between the age of 18 and 65 years with a total of 28,628. Between the years 2000 to 2007, a significant increase of 8.7% percent occurred [1]. It has been reported that there are an estimated 8-25 attempted suicides for every suicide death. Among the suicide deaths, 60% are caused by major depression and this number rises to over 75% if depression caused by alcoholism is included. These percentages showed that untreated depression plays a major role in the cause of suicide [2]. The statistics show the need for an analysis to identify and detect patients with near term suicidal risk.

Recognizing that suicide has a significant effect on public health, a scientific strategy for evaluating suicide is an important preventive measure that needs to be developed. Current estimate of risk assessment is through a clinical judgment process that includes some objectifiable risk factors, and often also includes a subjective "gut feeling" of the clinicians regarding the patient's potential state of suicidal thoughts. Clinical judgment requires an evaluation of comprehensive information regarding a patient's demographic profile, health record, history, family evaluation, prior mental health

treatments, as well as suicidal ideation, behavior, planning, desire, and intent [3]. Personality psychometric instruments or suicide-specific scales also have a wide range of use for estimating suicide potential and treatment decisions. But these measures are not suitable for making judgments regarding a patient's level of risk for suicide [4]. Another method called the neurobiology of suicide is under investigation where the research involves the study of how suicidal behavior triggers neurochemical abnormalities, receptor abnormalities, gene marker, and cell alterations. But these methods still require the improvement in recognition of patient with a greater risk of suicide [5]. Plus, they are very time-consuming and require clinicians to be cautious when recording information in order to provide useful assessment and treatment information.

Obtaining collateral information on each patient would be difficult, thus reducing the accuracy for suicide risk assessment. The collection of comprehensive information requires a significant amount of time and in some cases the information obtained is insufficient for clinicians to perform the assessment. Clinicians without proper training in psychiatry are not proficient in recognizing the difference between depression and suicidal potential [11]. Misdiagnosed patients with a high risk suicide with major depression may lead to tragic outcome. Patients with imminent risk of committing suicide will be able to receive early hospitalization if proper and accurate diagnostic tools were used. In many aspects, it is the validity of the information from a second component to give better quantitative expression that may yield the greatest hint of imminent risk aside from the clinical judgment.

A lot of information regarding the psychological state is accessible through the human voice. Several studies have been conducted since 1984 on the effects of emotional

arousal on speech production based on indication of speech rate, voice articulation and respiration. Some methods based on acoustic features that were investigated are estimation of fundamental phonation or pitch, features based on a nonlinear model of speech production, relation of vocal tract features to emotional speech and speech energy estimation [12]. Research findings show evidence of specific vocal characteristics among patients at the level of near term suicide [6], [7], [8], [9], [10]. The investigation of correlation between vocal characteristic and psychological suicidal state was proposed by Drs. Stephan and Marilyn Silverman. The psychologists showed evidence of the possibility of transforming what was heard into measurable quantitative information. Although the vocal characteristic are unique, the relationships between voice parameter and psychological state are extremely complex and require thorough research in obtaining acoustic measurements that are robust in distinguishing near term suicide with ideation or major depression [11].

### *1.2 Statement of purpose*

The main purpose of this research was to continue the studies of vocal parameters that were done by France, Ozdas and Yingthawornsuk [6], [7], [8] in the investigation of determining near term suicidal risk. Previous studies separated patients into categories of high risk suicidal, depressed, and remitted. Suicidal patients and depressed patients are the two major groups that were being studied in depth in this research. Development of a robust instrument and independent distinguishable classifier between these two categories is crucial in achieving early detection, thus preventing suicidal behavior in later stages.

The second purpose of this research was to study the validity of using Equal-Test-Train, Jackknife and Cross-validation resampling techniques in assisting Linear and Quadratic Discriminant Analysis when performing classification between the suicidal and depressed states. Even though it is also important to be able to distinguish between depressed and remitted patients or suicidal and remitted patients, these two problems are not considered as critical. Recognizing between suicidal and remitted patients may be more obvious compared to other pairwise groups. Identifying suicidal and depressed patients is also one of the main focuses in this research and will be discussed in detail. Classifications on other groups of patients were also studied and the results were appended at the end of the chapters.

This study uses recordings from a better controlled environment compared to previous recordings provided by the Silvermans. Attaining accurate recordings were difficult due to the fact that patients may not always exhibit the state suicidal during the interview sessions. The third purpose of this study is to reproduce some of the results that were obtained by Yingthawornsuk [8]. In his research, the analysis on features obtained was done using statistical software called SYSTAT and also using an On-Line Pattern Analysis System, PcOLPARS. The research herein mainly uses code written for Matlab and its statistics toolbox. Classification methods were developed by extracting information contained in the Power Spectral Density of the speech samples. Yingthawornsuk, in his study showed that they are powerful and effective features to be used for discriminating between mental states.

Another purpose of this study is to recognize the effectiveness of using information extracted from interview speech compared to reading speech as a potential suicidal indication for male speech samples.

### *1.3 Description of chapters*

Chapter 2 will provide the readers with some basic concepts of mechanism and physiological aspects of speech production. The chapter continues with introducing the speech production model and very briefly discusses the source-filter separation model.

Chapter 3 briefly summarizes the previous work done by other researchers in the related field of study. This chapter briefly introduces the multiple features and analysis techniques that were used in previous studies for discriminating among the suicidal, major depressed and non-suicidal patients.

Chapter 4 started off with explaining how the database were collected and processed. The chapter continues explaining the features extraction analysis by showing a step-by-step procedure for obtaining the Power Spectral Density (PSD) from the voiced speech samples. Three types of resampling methods (Equal-Test-Train, Jackknife and Cross-validation) that were implemented together with the Linear and Quadratic Discriminant Analysis are explained in details.

Chapter 5 presents applications and results of the classification using the statistical resampling methods. This chapter investigates the effectiveness of using the PSD as a possible feature for identifying between the suicidal group and the depressed group. Further analysis on how to improve classification results was shown.

Finally, chapter 6 concluded the thesis with discussions and possible future work based on this study.

## CHAPTER II

### BACKGROUND AND SIGNIFICANCE

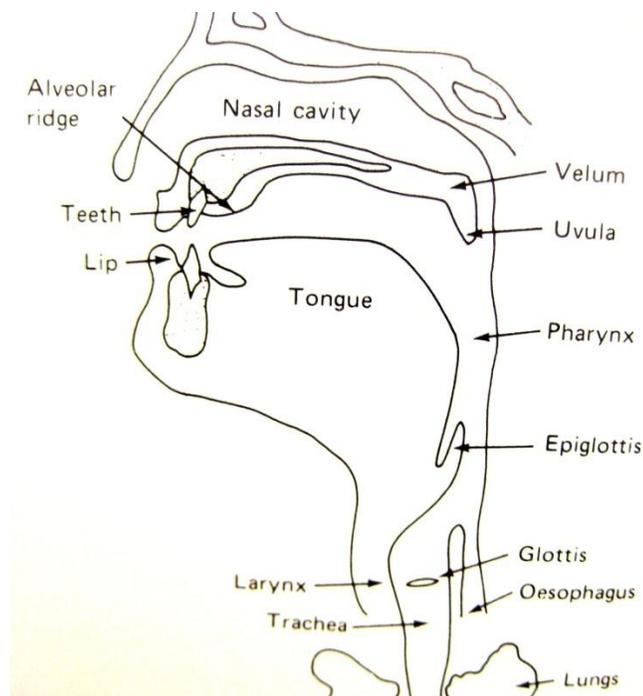
Voice is sound produce by humans using their lungs and vocal folds located in larynx while speech is thoughts, ideas and feelings that are being transferred orally, triggered by the muscle articulation that alters the tone produced by voice into a known decodable sound. Speech plays a significant role in everyday life. Humans use speech to express complex or abstract information such as emotions, thoughts and ideas. Communication between human beings involves not just the spoken words, but also other features corresponding to intonation, accent, loudness and speed.

Beginning in the early 20<sup>th</sup> century, psychiatrists began investigating on the effects of emotional arousal on human vocal emission and since then, many research studies had shown this relationship [17]. Over the years, research on this topic has grown, but compared to emotion estimation on speech, emotion estimation on facial expression is still far ahead. In particular, speech-based studies on suicidal detection are a fairly new topic that requires considerable analysis regarding the effect of the suicidal state on speech. The psychological states of depression and suicide were shown to exhibit somewhat different characteristics, and discriminating between them was shown attainable based on previous studies [6], [7], [8], [13]. A concise introduction on the mechanism and physiology of human speech will be discussed in this section for the purpose of understanding ways of representing speech signals using a digital signal

processing modeling. Once speech is generated and propagated, the perception of speech begins (speech recognition). Thus, the auditory system will also be briefly discussed.

## 2.1 Mechanism and Physiological Aspect of Speech Production

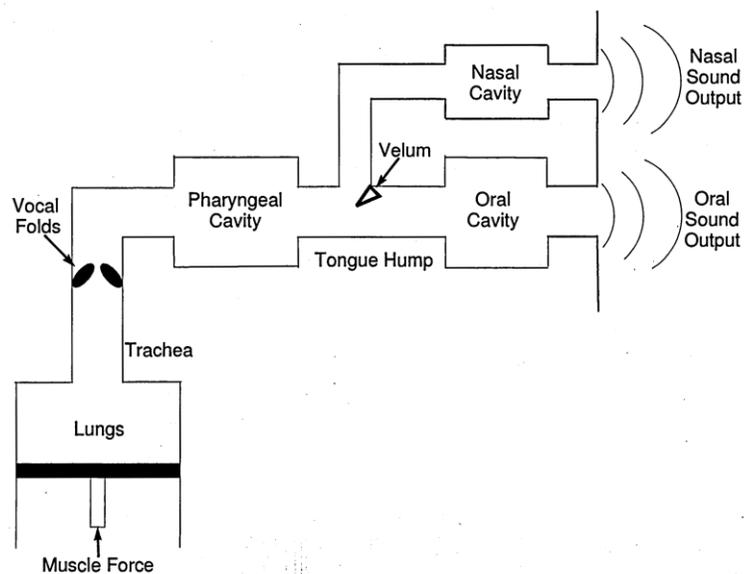
Figure 2.1 shows the left cross-sectional view of the upper portion of a human anatomical structure involved in the production of speech. Speech production can be viewed as four separate components, the respiratory system, laryngeal (vibration) system, resonance system and the articulatory system [19].



**Figure 2.1:** Cross-sectional view of an anatomy structure for human vocal production [19]

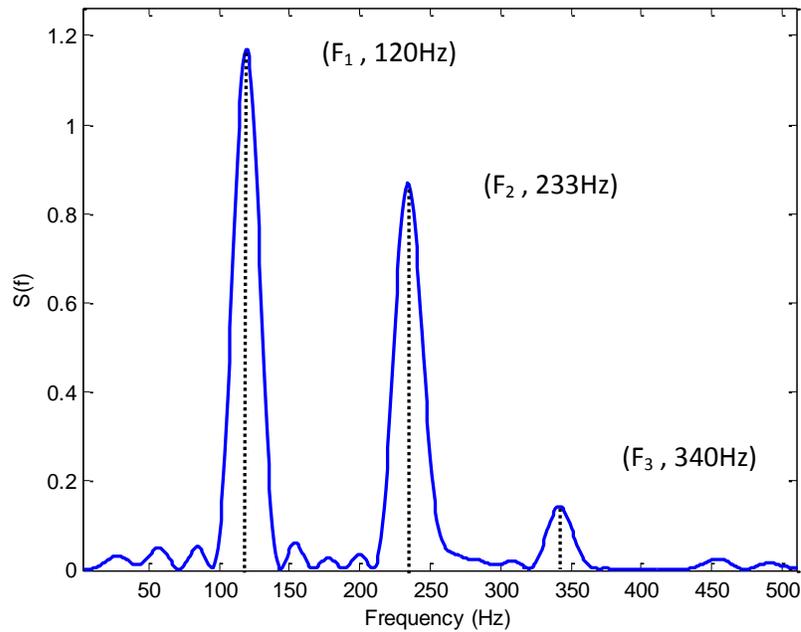
Respiratory system is comprised of lungs, bronchi, trachea and other associated muscles. It acts as the main energy source by supplying air, and is also responsible for the amplitude of the sound as the displacement of vocal chords changes with respect to

air flow energy. The diaphragm is the most powerful muscle use in producing the voice. Figure 2.2 shows a simplified block diagram of the speech production process. The lungs provide airflow and muscle force pushes air through trachea, bronchi, and through the glottis located between the vocal chords and larynx into three main cavities consisting of the vocal tract, the pharynx, and the oral and nasal cavities. The laryngeal system consists of the larynx and the vocal chords (vocal folds). The larynx is a tube consisting of cartilages and muscles, commonly called the voice box involved in the opening and closing of the glottis (the space in-between the vocal chords). The resonance system is the pharyngeal cavity where it is made of the pharynx (main resonating chamber for the voice), oral cavity and nasal cavity [19]. Pulses of sound are being manipulated here into a recognizable voice. The velum, jaw, lip, tongue, teeth and other structures that are mostly visible outside the human form make up the articulatory system. They alter the speech into a comprehensible utterance called speech. Air flow exits through the mouth and nose to become voiced or unvoiced sounds.



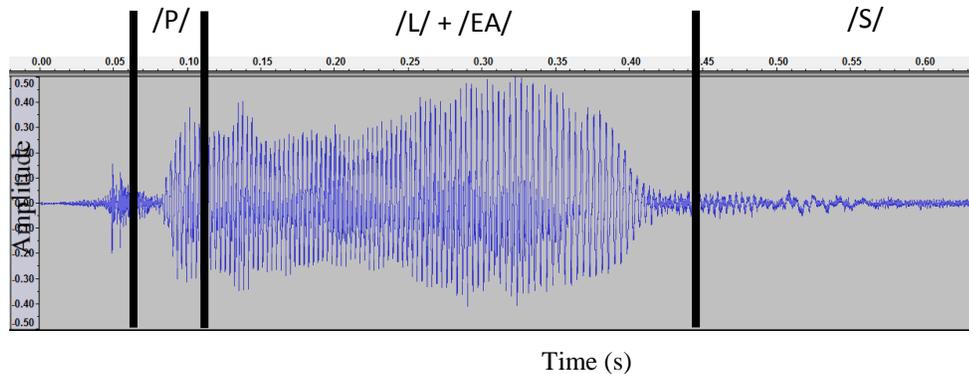
**Figure 2.2:** Simplified speech production model [21]

The most important system component in human speech production is primarily considered to be the vocal tract. The vocal tract is a tube-like passageway that begins at the opening of the vocal chords (glottis) and finishes at the lips (some literature may include the nose as well). The vocal chords vibrate and create the opening and closing of the glottal passage. The glottis is the opening of the vocal tract where pulses beginning to be filtered. Opening and closing of the vocal chords periodically controls air in a quasi-periodic manner. The vocal tract applies spectral shaping to the pulses produced and this shaping varies over time. The generated pulses are considered to be a quasi-stationary process where static parameters of speech remain reasonably constant over short time intervals, typically 10ms to 30ms. The vocal tract is frequently described in terms of its resonances (formants,  $F_i$ ). Figure 2.3 shows an example of formants located in a speech spectrum. Formants represent the spectral peaks of acoustic energy around a particular frequency in the speech wave depending on the shape of the vocal tract.  $F_1$  is often observed as the strongest formant because as the frequency increase, the power decreases due to the lowpass nature of glottal excitation [18].

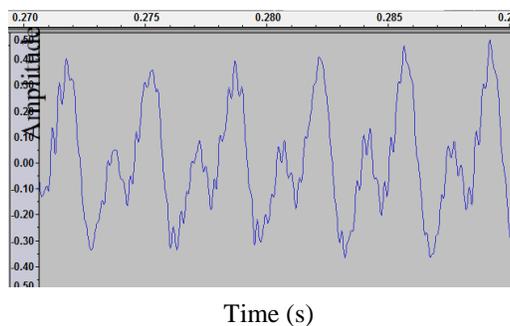


**Figure 2.3:** Example of voice spectrum frequency

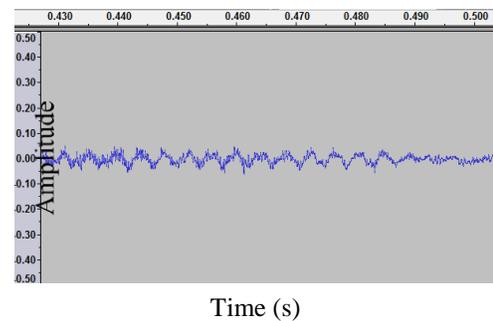
With air flowing through the vocal tract, the articulation of the velum is used to produce constriction. Lowering the velum causes air from the vocal tract region up to the lips to be restricted thus, allowing more openings towards the nose passage. For voiced speech, the velum is articulated upward, thus causing the nasal passage to be blocked temporarily in order for sound to be produced through the lips. The larynx functions as the airflow regulator into the vocal tract, which causes the formant frequencies to increase or decrease by altering the tract length via raising or lowering the larynx [18]. Besides the larynx and velum, the tongue and lips are the two other major organs in an articulatory system which produces various sounds.



**Figure 2.4 (a):** Speech waveform produced by word “please”

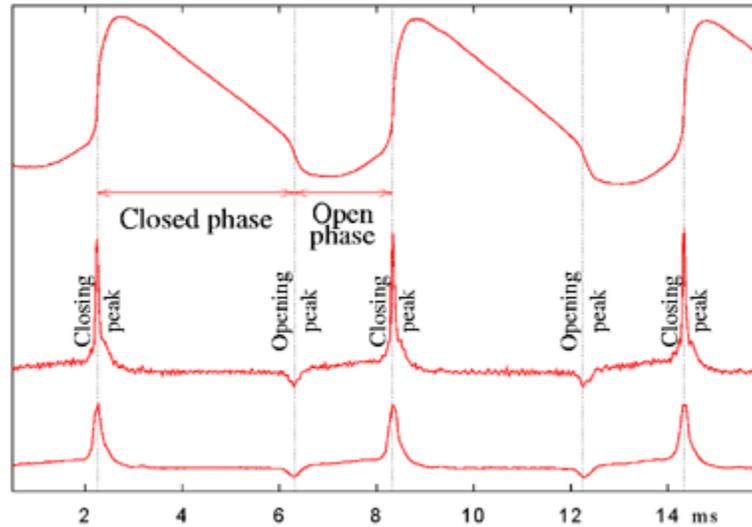


**Figure 2.4 (b):** Voiced speech segment from word “please”



**Figure 2.4 (c):** Unvoiced speech segment from word “please”

Production of speech can be divided in two main categories, voiced and unvoiced speech (or a combination of the two). Figure 2.4 shows the distinctions between voiced and unvoiced waveforms. The word “please” contains four phonemes at which /L/ and /EA/ can be categorized as voiced and /P/, /S/ as unvoiced. Voiced speech displays characteristics of large peaks, quasi-periodic waveforms and lower zero-crossings, while unvoiced speech displays multiple random small peaks, noise-like waveform and higher zero-crossings.



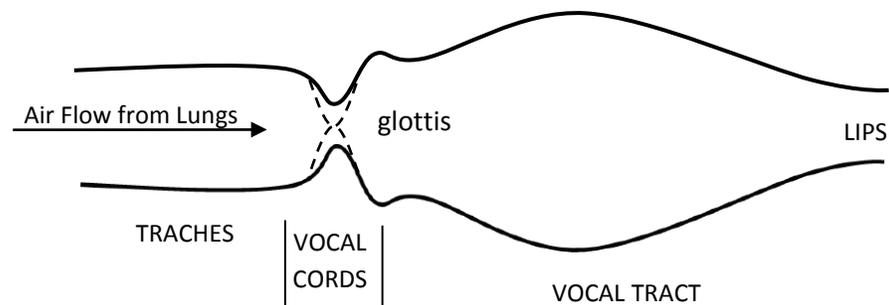
**Figure 2.5:** EGG signal corresponding to opening and closing of the glottis (top), DEGG Derivative of the signal (middle), smoothed DEGG (bottom) [23].

### 2.1.1 Voiced sound

Voiced sounds are formed through the vibration of the vocal chords (vocal folds). Its input can be modeled as a quasi-periodic excitation at the glottal passage caused by the opening and closing of the glottis. Figure 2.5 shows an electrogastrogram (EGG) recording of electrical signals that travel through the glottis. The derivative of EGG (DEGG) produced an alternating positive and negative peak, where positive peak corresponds to the immediate closing of the glottis and the negative peak corresponds to the opening of the glottis which is represented by the steep fall in EGG signal [23]. Through a process called adduction, Bernoulli force causes the vocal cords to be brought together and creates a closed air space under the glottis thus, provisionally blocking the air flow from the lungs. This process leads to an increase in sub-glottal pressure when air pressure from the lungs continues to build up below the vocal cords.

Figure 2.6 represents the schematic representation of the vocal tract model when the vocal chords are closed. Once this pressure becomes greater than the resistance of the

vocal chords, the vocal cords re-open and release a single waft of air. Due to elasticity, laryngeal muscle tension, and Bernoulli effect, the vocal chords rapidly close to its original position. This process is sustained by a continuous supply of pressurized air in a quasi-periodic manner. The cycle continues until thousands wafts of air are released and filtered through the vocal tract thus, producing sounds [22]. The fundamental period ( $F_0$ ) or pitch period ( $T_0$ ) corresponds to the time between consecutive vocal chords cessations (frequency pulses).



**Figure 2.6:** Vocal tract organ pipe model [20]

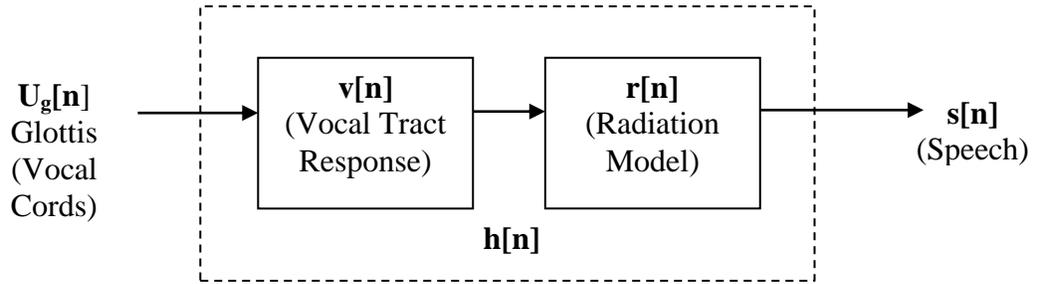
### 2.1.2 Unvoiced sound

Unvoiced sound has more of a noise-like quality that is created when the vocal chords make an opening and steady air flow passes through it. Articulatory organs provide constrictions in the passage, thus, causing turbulence or noise-like behavior in the air flow. The sound developed typically exhibit smaller amplitude and faster oscillation compared to voiced sound.

## 2.2 Speech Production Model

A conceptual representation of speech production is derived in order to extract important information from speech. According to source-filter theory, speech signals can

be viewed as a glottal source excitation (source) followed by a linear time-varying filter that shapes the resonance characteristic of the vocal tract. The assumption of speech production as a linear process is an oversimplified model of a more complex model as further described in [20]. Even so, such an approximate model permits an examination of the effects of glottal excitation and vocal tract independently. Therefore, modification of the properties in the vocal tract will not affect the properties in the source excitation and vice versa.



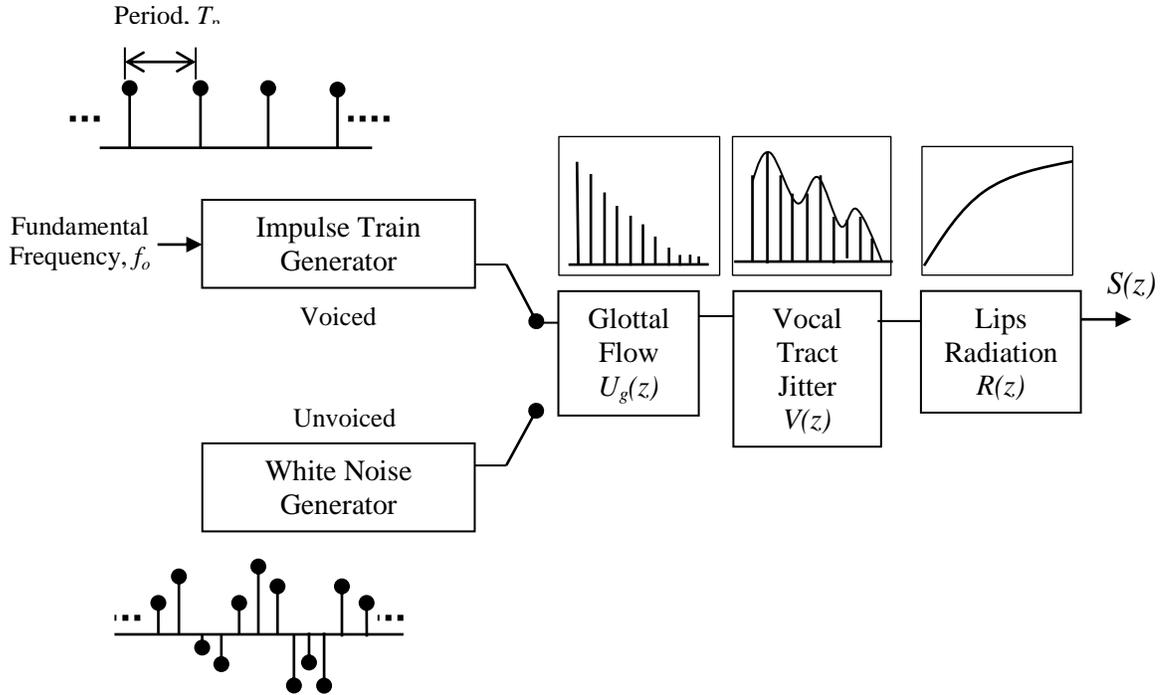
*Figure 2.7: Linear filter of a voice production model*

In most speech analysis, the main focus would be on the voiced part of the speech. Voiced speech can be defined to be the convolution of the input waveform with its impulse response in time-domain. Voiced speech is modeled as periodic pulses of air-flow with a desired fundamental frequency that is shaped by the glottis represented by  $U_g[n]$ . The glottal shaped pulse passes through a pulse shape modifier with a tube-like passage way called the vocal tract that is represented by  $v[n]$ . Finally, the produced sound is emitted to the surrounding air through radiation (lips),  $r[n]$ . For mathematical

purpose,  $v[n]$  and  $r[n]$  can be grouped together and represented as  $h[n]$ . Also, radiation (lips) only plays a role in shaping the quasi-periodic train of glottal pulses [16].

$$S(z) = Ug(z)V(z)R(z) \quad (2.1)$$

The voiced speech can also be viewed as multiplying the input spectrum by its frequency response which can be represented in the frequency-domain (z-transform) as shown in equation 2.1. The source-filter model allows the modeling of speech production as a linearly separable filter. In this acoustic system, the vocal tract is assumed to be approximately linear by disregarding the effect of vibrating walls or external radiation [18], thus allowing it to be characterized as a frequency response or impulse response.



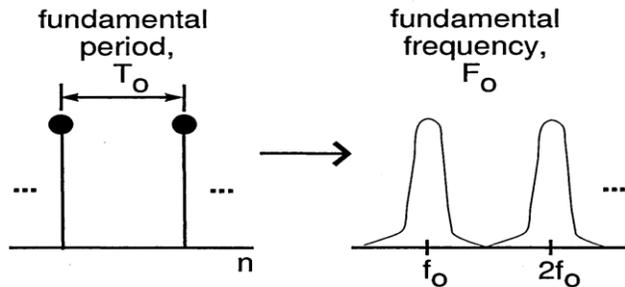
**Figure 2.8:** Source-filter model of speech production

### 2.2.1 The Source Excitation

When modeling the source of voice production, there is a difference between the acoustic model and the source-filter model. In the acoustic model, the glottal flow is dependent on the vocal tract shape due to the acoustic load above the glottis that is defined by the output of the vocal tract. Meanwhile, assuming that the source-filter model is independent of the vocal tract shaping variations, the glottal source is defined as a non-interactive signal description of the voice source [24]. Implementation of this model inputs a random white noise for unvoiced sound and a discrete-time periodic impulse train with a certain fundamental period between each pulse that acts as the source excitation signal for voiced sound.

Voiced speech is considered to be non-stationary over a large interval of time but the characteristics and information in the voiced speech can be measured to be relatively constant over a short period of time. Similarly, the glottal pulse can be represented by Equation 2.2 [16] where  $g[n]$  represents the discrete-time impulse train pulses and  $T_o$  is the fundamental period.

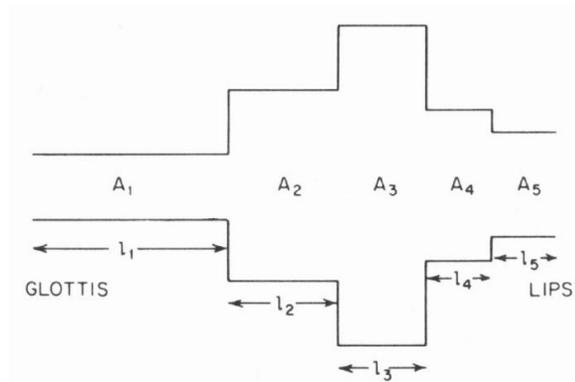
$$g[n] = \sum_k \delta[n - k T_o] \quad (2.2)$$



**Figure 2.9:** Time and frequency domain representation of glottal pulses [16]

### 2.2.2 The Vocal Tract Filter

The actual model of the vocal tract consists of varying the cross-sectional area based on the position across the tract as the wave propagates over time. These variations are caused by the alteration of the frequency content of the excitation signal. The continuous-time model of a vocal tract can be conveniently represented as a discrete-time model by transforming it into a concatenation of uniform lossless tubes of varying diameters. These tubes are considered “lossless” due to the assumption that no sound energy is absorbed by the walls. For an arbitrary shape of vocal tract, the area would vary with respect to time,  $A(x, t)$ . Referring to Figure 2.10, assuming that the vocal tract exhibits a uniform tube-like shape, the constant cross sectional area  $\{A_k\}$  and length  $\{l_k\}$  of  $N$ -sections are chosen to approximate the total area of the vocal tract,  $A(x)$ .



**Figure 2.10:** Concatenation of lossless tubes for  $N = 5$ . [20]

The output speech is related to the relationship between pressure and volume velocity which are determined by the cross-sectional area of the tube and the speed of air. At the joint of two tubes, continuity must be obtained in order to keep constant pressure on both sides as the waves traveling from one tube to the other. The excitation propagates through the series of tubes with some partially reflected and some waves partially

propagated across the two joint tubes. Besides the joint of two tubes, boundary conditions at the lips and glottis must also be taken into account [20]. A linear prediction (LP) analysis involves the prediction of signal parameters based on the previous values and is a technique that is used to model the vocal tract as an all-pole filter called an inverse filter as shown in Equation 2.2 where  $\{a_k, 1 \leq k \leq N\}$  are the predictor coefficient and the  $N$  order of the filter (number of poles).

$$V(z) = \frac{1}{A(z)} \quad \text{and} \quad A(z) = 1 - \sum_{k=1}^N a_k z^{-k} \quad (2.2)$$

When modeling the vocal tract by an all-pole filter, the nasal and unvoiced sounds are not taken into account. According to [25], inclusion of nasal and unvoiced sound into the current all-pole model can be achieved by including more poles rather than including zeros. All poles will remain inside the unit circle considering the areas of the concatenated tubes to be positive.

### 2.2.3 Lips Radiation

The opening of the lips marks the end of vocal tract tubes. The lip opening is modeled as an orifice in a sphere where the lips are represented as radiating sound waves and the head is represented by a spherical baffle that refracts the sound waves. If the opening of the lips is small enough compared to the size of the sphere, the radiating surface can be thought of as a radiation from an infinite plane baffle. Pressure is

measured from a given distance,  $l$  from the mouth and is proportional to the time-derivative of the lips flow.

### 2.3 Source Filter Separation

A signal and additive noise are inseparable in time domain due to the fact that they occur at the same point in time, whereas in the frequency domain, these signals are rather separable when applying appropriate filtering techniques to the resulted frequency responses. But according to Equation 3.1, speech is combined through a convolution of the glottal pulse excitation  $u_g[n]$ , vocal tract  $v[n]$ , and lips radiation  $r[n]$ , where  $h[n]$ , can be viewed as a combination of the vocal tract and lips radiation. In this case, the signal and noise will be overlap in both the time and frequency domains. By means of the Complex Cepstrum [7], the speech can be decomposed into a periodic pulse train and the vocal tract response.



The Complex Cepstrum,  $\hat{s}[n]$  of a discrete signal  $s[n]$  is defined by decomposing the signal through the Discrete-Time Fourier Transform (DTFT), the complex logarithm and the inverse DTFT which are shown respectively in Equations 2.4, 2.5 and 2.6. When computing the Complex Cepstrum, the problem lies in estimating  $\angle S(e^{j\hat{\omega}})$  which requires an efficient phase unwrapping method in order to prevent error.

$$s[n] = u_g[n] * h[n] \quad \xleftrightarrow{DTFT} \quad S(e^{j\hat{\omega}}) = U_g(e^{j\hat{\omega}})H(e^{j\hat{\omega}}) \quad (2.4)$$

$$\log_e[S(e^{j\hat{\omega}})] = \log_e|S(e^{j\hat{\omega}})| + j\angle S(e^{j\hat{\omega}}) \quad (2.5)$$

$$\hat{S}(e^{j\hat{\omega}}) = \hat{U}_g(e^{j\hat{\omega}}) + \hat{H}(e^{j\hat{\omega}}) \quad \xleftrightarrow{IDTFT} \quad \hat{s}[n] = \hat{u}_g[n] + \hat{h}[n] \quad (2.6)$$

## CHAPTER III

### LITERATURE REVIEW

The idea of recognizing distinctive patterns and tone of voice in patients with high risk suicide was introduced by two clinical psychologists, Drs. Stephen and Marilyn Silverman. Both had experience in treating patients with near term suicidal risk. They began research in the 1980s by collecting and analyzing suicidal tape recordings obtained through therapy sessions in an uncontrolled environment, and notes and interviews made shortly before suicide attempts [14]. They describe the similarity of vocal speech between depressed and suicidal patient but notice changes occur considerably in the tonal quality and acoustical characteristics when the patient enters the suicidal state [8]. Several other researchers continue to study the relation of vocal tract characteristic to depression and suicidal risk.

France et al [6] began the research by extracting and analyzing fundamental frequency ( $F_0$ ), amplitude modulation (AM), the formants and power distribution (PSD) on speech samples. Among these perceptual qualities, formant and PSD features appeared to be distinguishing vocal features when discriminating between suicidal and major depressed patients compared to the ones collected from control groups. Linear Predictive Coding (LPC) was preferred over Long-term-average (LTAS) to calculate the formant frequencies and bandwidths due to the volume of speech analyzed which made it computationally expensive even though the LTAS approach provides a more accurate representation of the formant properties. The classical Welch method was used when

extracting the PSD. The energy spectrum was investigated on the percentages of total energy in frequency sub-bands with a bandwidth of 500 Hz over the frequency range of 0-2000 Hz. It was reported that most energy are distributed in the range of 0-2000 Hz. Features were integrated when performing classification in order to obtain the best parameter combination in distinguishing between the suicidal and non-suicidal groups. Multi-parameter classification was shown to be more effective than classification of single parameters.

Ozdas et al [7] studied the discriminating power of lower order mel-cepstral coefficients (MFCC) among suicidal, major depressed and non-suicidal patients. Vocal tract characteristic using a non-model based approach for near term suicidal risk assessment was the focus of this study. The effects of source (excitation) and filter (vocal tract) on suicidal state were the two domains examined. Vocal jitter and slope of the glottal flow spectrum were two other excitation features that were further investigated related to the excitation signal, whereas in the filter domain, speech features are investigated through cepstral analysis. There were variations among different psychological states with the use of mel-cepstral filter bank coefficients and the results suggested that the use of MFCC features could provide useful measurements for identification of a possible suicidal state. The use of Gaussian mixture models yielded better class approximation when performing classification for individual diagnostic groups.

Yingthawornsuk et al [8] performed an acoustic discriminant analysis using PSD vocal features on male patients with suicidal, depressed and remitted speech. The study continued with the integration of vocal tract spectral characteristic from speech samples

obtained using PSD-based features with GMM-based spectral modeling features for discrimination. Similarly, this method was performed on groups of suicidal, depressed and remitted patients, but for both female and male patients. The features used for discriminating analysis were magnitude of maximum peak, frequency location of maximum peak, PSD<sub>1</sub>, PSD<sub>2</sub>, and PSD<sub>3</sub>. PSD<sub>4</sub> was disregarded due to its linear dependency on the sub-band energies. Among these features, the percentages of total power (PSD<sub>1</sub>, PSD<sub>2</sub>, and PSD<sub>3</sub>) were found to be the most significant features in effectively distinguishing between groups of suicidal, depressed and remitted patients.

Keskinpala et al [13] centers her work on analyzing vocal characteristic of high risk suicidal and depressed patients on both male and female speech samples using mel-cepstral coefficients and using energy in the frequency bands. This study introduced the method of cepstral mean normalization for compensating spectral variability due to different recording environments. The importance of environmental compensation was tested by performing classification with and without compensation and results demonstrated that using no compensation provided better results. Text-dependent speech samples were shown to provide better discrimination analysis in distinguishing suicidal patients compared to the interview speech sample. The cross-validation and testing with all data training were two methods of resampling used in attaining the classification qualitative measurements. Cross validation classifier based-method was demonstrated to perform well as an assessment approach in identifying high risk suicidal patients. Other studies involves changing the number of energy bands extracted from the spectral density, extending the frequency band from 2000 Hz to 3000 Hz and varying the size of band pass filter for each spectral energy band.

Tolkmitt et al [15] analyzed patients during the course of recovery by obtaining speech samples before and after receiving treatment by comparing fundamental frequency, spectral energy distribution and the formant frequencies of vowels found in identical phonetical content. The fundamental frequency reflects the tension in the muscle tone of the vocal chords and the patient's speech samples tend to show decrease in the fundamental frequency when going from the depressive state to recovery. Patients with depressed speech samples experience a decrease in spectral energy below 500 Hz and an increase in spectral energy between 500-1000 Hz after receiving treatments.

Moore et al [16] gathered multiple features from three main categories of prosodics, vocal tract and glottal measures. An optimal set of classifiers were chosen based on discriminant analysis techniques and features were selected according to a set of classifiers that were observed to be the most optimized classifier in separation of depressed and control groups. Formant structures and power spectral density measures were found to be the most prominent discriminators in the creation of statistical separation between patient groups. The best discriminators identified for male depressed patients were related to the vocal tract with regards to glottal and formant features. On the other hand, female depressed patients exhibited vocal tract feature related to glottal and energy features as the best discriminators.

## CHAPTER IV

### METHODOLOGY

#### **4.1 Database Collection**

The audio recordings were obtained from an ongoing research project in the Vanderbilt University Department of Psychiatry. Recordings were made during a one-to-one treatment session with a psychiatrist. Patients were between 25 to 65 years of age. Male and female recordings were divided into three diagnostic groups of high risk suicidal, depressed and remitted. Patients were initially categorized into these groups according to analysis and assessment performed by experienced psychiatrists. Remitted patients were patients with depressed symptoms, but who recovered after going through a series of treatments.

Two types of recordings were gathered from each patient labeled as interview and reading. Interview recordings are considered as spontaneous speech where patients are being interviewed by a psychiatrist with a given guidelines of questions and reading sessions are considered as controlled speech where patients were asked to read from a standardized “rainbow passage”. The “rainbow passage” was used due to the fact that it contains every sound in the English language and is considered to be phonetically balanced [26]. The brain acts differently when patient is engaged in spontaneous speech (interview) compared to controlled speech (reading). In the “rainbow passage” the patient would just have to read what’s given and the process of thinking of what to say is not involved when the patient is reading. For the purpose of this study, only male recordings

were analyzed in detail. The numbers of male patients gathered by groups are listed in table 4.1. Most of the data used in this research were also used previously by Yingthawornsuk [8] and Keskinpala [13].

*Table 4.1: The number of patients (male) according to groups*

Group	Number of patients	
	Interview	Reading
High Risk	10	9
Depressed	11	12
Remitted	10	11

The recordings were made in a closed-room. Before the interview session, patients were asked to count from one to thirty while the psychiatrist manually adjusted the volume intensity for the recording. The recordings were saved as a .wav files with an identification number that represents the date of the recordings and types of recordings (interview or reading). An example would be 010111\_interview.wav for an interview recorded on Jan 1<sup>st</sup> 2011.

#### **4.2 Data Pre-processing**

All speech recordings were made on mono channel. Previous voice recordings used in [7], [8] and [13] were sampled at 10 kHz but in this study, all voice recordings were digitized with a higher sampling rate of 44.1 kHz for better sound quality. These recordings were then edited using the Audacity audio digital editor by removing the

interviewer's voice, removing long pauses that are present for more than 0.5 second and removing background noises such as door slams, sneezing and paper rustling sounds. Each edited speech sample was detrended for better analysis of the data fluctuations and to compensate for possible variability that exists in the speech samples when the recordings were made. For analysis purposes, the sampled signals were divided into short time frames. The frames are chosen to be short enough in order for the signals to appear quasi-stationary but must also be long enough for the signal to contain at least one cycle of the desired frequency. Therefore, a window frame of 40ms that would comprise a few periods of the voiced waveform was chosen for this analysis.

#### *4.2.1 Voiced, unvoiced and silence detection*

Speech signals are comprised of voiced, unvoiced, and short silence segments that are mixed and combined together. Keskinpala [8] and Yingthawornsuk [13] used the method that was created by Ozdas [7] for voiced/unvoiced detection. According to Ozdas, voiced, unvoiced and silence speech samples can be estimated by segmenting the sampled signals based on their energy values at different levels of the Wavelet Transform (WT). Voiced speech samples exhibit a quasi-stationary behavior and are composed of low frequency characteristics. On the other hand, unvoiced speech samples exhibit noise-like behavior and contains more high frequencies. The sampled signals were separated into segments and for each segment, the energy was calculated for each of several different band levels.

*Table 4.2: Frequency range for each band levels [7]*

<b>Band level</b>	<b>-3dB band pass filter limits</b>
1	2500 – 5000
2	720 – 2340
3	320 – 1080
4	160 – 540
5	80 – 260

Energy for each band was obtained by calculating the Discrete Wavelet Transform (DWT) at each frequency band levels. These bands represent a set of band pass filters as shown in Table 4.2. Lower bands allow higher frequency content and filter out low frequency information while higher bands capture the lower frequency information.

For this study, a similar method was used for the voiced, unvoiced and silence classification but instead of using the WT to determine the energy bands, a set of third order band-pass filters was applied to each segment of the sampled signal. If a filtered speech segment has maximum energy equal to the total energy in band one, it was classified as unvoiced. The median of the total energy in band three was set as the threshold for separating voiced and silence where energy higher than or equal to the threshold was categorized as voiced otherwise it was categorized as silence. All unvoiced and silence terms were removed and only the voiced terms were collected and concatenated into one new speech signal for further analysis. Finally, the collected voiced signals were split into 20-seconds segments.

Table 4.3 shows the number of 20-second segments collected from male patients after performing the voiced/unvoiced detection. Note that for every patient, the last segment was removed if contained less than 20 seconds. The total length of voiced speech samples and the 20-seconds segments will vary for every patient depending on how long the interview or reading sessions were recorded initially.

*Table 4.3: The total number of 20-seconds segments according to groups*

Group	Number of 20-seconds segments	
	Interview	Reading
High Risk	123	26
Depressed	95	26
Remitted	92	23

### 4.3 Feature Extraction

No reference code was found on Yingthawornsuk [13] and Ozdas [7] because they did not use Matlab for the data analysis. Therefore, calculation to obtain the Power Spectral Density (PSD) had to be reverse engineered according to France's [6] Welch based PSD estimation. For this study, there were a few modifications made to France's method for attaining the PSD.

The PSD was extracted from the collected voiced speech samples using the method of the Periodogram. Each 20-seconds segment was analyzed using a non-overlapping window size of 40ms to produce 500 frames that consist of 1764

samples/frame. A simple Fast Fourier Transform based power spectrum estimation was applied on every frame to obtain the PSD.

#### *4.3.1 Identifying the number of spectral features/bands*

The features were collected by dividing the energy spectrum within 0-2000 Hz. The number of features needed affects how good the result will be. This research was conducted with two types of speech samples, interview and reading. For each sample, the numbers of features were collected starting with four 500 Hz equal bands within the 0-2000 Hz. Depending on the outcome, if the percentage of classification does not achieve satisfaction, a higher number of bands might need to be extracted with an increment of two bands per trial. Further method discussed in this chapter will be based on eight bands of spectral energy.

#### *4.3.2 Feature extraction for eight equal bands of energy spectrum*

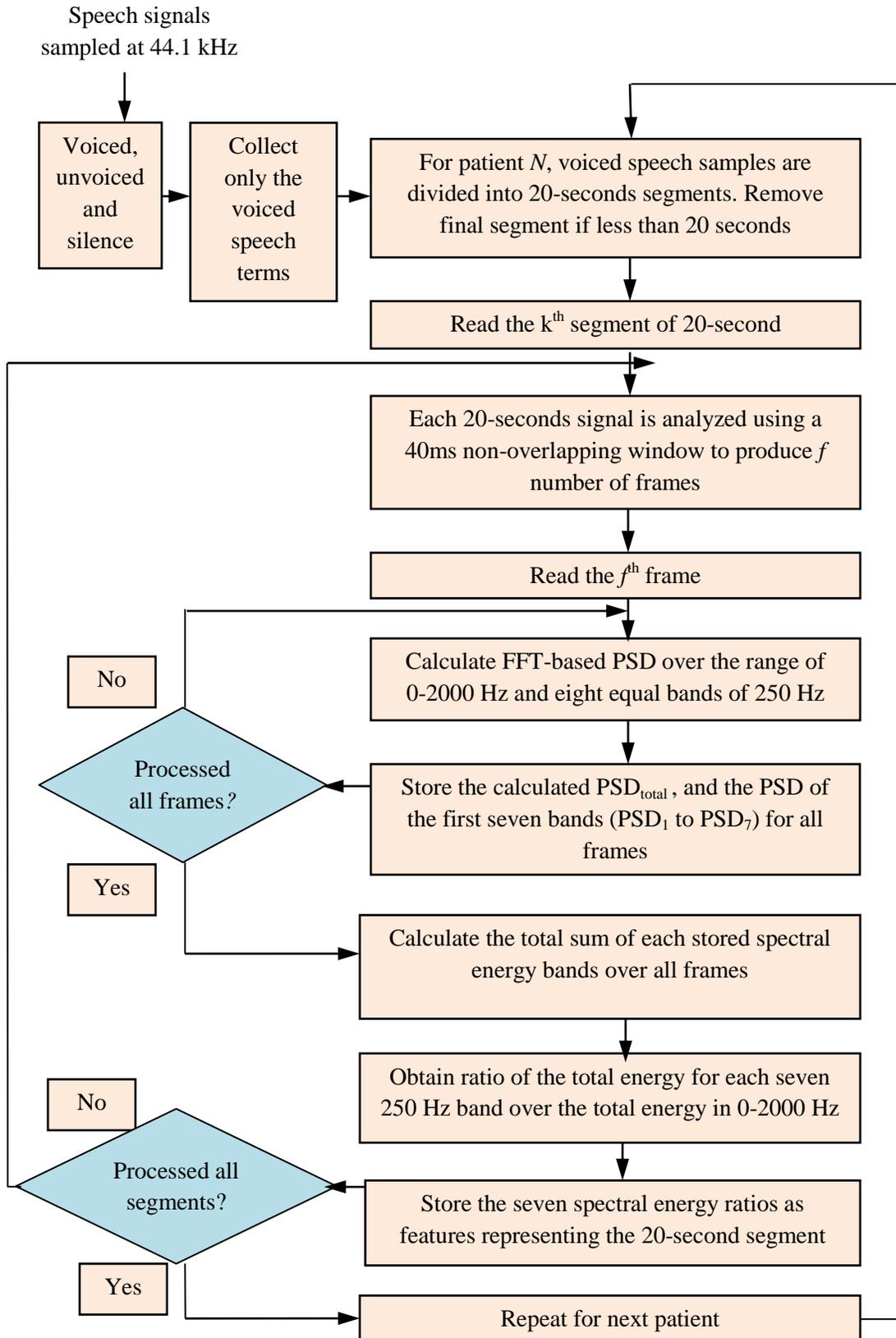
Eight equal 250 Hz bands of energy spectrum as shown in Table 4.4 were extracted from a frequency range of 0-2,000 Hz using the trapezoidal numerical integration. PSD<sub>8</sub> (1,750-2,000 Hz) was removed due to the fact that the contained information that is linear dependent on the other seven spectral energy bands and contained only a very small energy. PSD for a full frequency range of 0-2,000 Hz (PSD<sub>total</sub>) was also kept for further investigation.

Each estimated PSD band (PSD<sub>1</sub>, PSD<sub>2</sub>, PSD<sub>3</sub>, PSD<sub>4</sub>, PSD<sub>5</sub>, PSD<sub>6</sub>, PSD<sub>7</sub>) and the calculated PSD<sub>total</sub> obtained from all 500 frames were summed up. Ratios for individual spectral energy bands were calculated by dividing the total summed energy in each 250

Hz band over the total summed energy in 0-2000 Hz. A collection of features to represent each 20-seconds segments voiced speech samples were stored in a single row vector comprised of seven spectral energy ratios.

*Table 4.4: PSD based feature extraction*

<b>PSD Band</b>	<b>Frequency range (Hz)</b>
PSD <sub>1</sub>	0-250
PSD <sub>2</sub>	250-500
PSD <sub>3</sub>	500-750
PSD <sub>4</sub>	750-1000
PSD <sub>5</sub>	1000-1250
PSD <sub>6</sub>	1250-1500
PSD <sub>7</sub>	1500-1700
PSD <sub>total</sub>	0-2000



**Figure 4.1:** Flow chart for the PSD feature extraction

#### 4.4 Feature Analysis and Classification

Each patient,  $k$  number of 20-seconds segments where the total number of  $k$  segments varies from one patient to another. Features extracted from the PSD analysis for each patient were collected into a matrix of  $k$  rows by  $P$  column where  $P$  represents the seven spectral energy band ratios (PSD<sub>1</sub> to PSD<sub>7</sub>). These feature vectors were saved as a .mat file and separated into high risk suicidal, depressed and remitted group by an indication of the letter “h”, “d” and “r” that were attached at the beginning of the file name.

The discriminant analyses performed on the acquired features were done on the basis of pairwise analysis classification consisting of high risk/depressed, high risk/remitted and depressed/remitted. The decision boundaries for the two-class classification were obtained using a quadratic classifier and a linear classifier. Because of having small data sets, a resampling method was necessary to be used when performing linear and quadratic classifications. The resampling methods that were adopted in this research were Equal Test-Train, Jackknife (Leave-One-Out) and Cross-Validation. The discriminant functions were applied using the “classify” command provided in the MATLAB statistical tool.

The “classify” command in MATLAB that was used in this study requires four input parameters; test, train, class label and type of discriminant function. In the two-class classification, the class labels are made of a bunch of zeros and ones that are stacked together in a column vector. The size of the column vector depends on the number of training samples. The training data consists of a  $n$  row by  $P$  column matrix where the size of  $n$  varies depending on the training sample size.  $P$  represents the seven features

extracted from the PSD of each 20-second segment of voiced speech. Each training sample is associated with a class label, where in this case, the class label would be either zero or one. The test data are assumed to have an “unknown” class label. The classify command will output an estimated class label for the test data according to the training data distribution and classification. The two types of discriminant function that were used in this classification are linear and quadratic classifier.

#### 4.4.1 Quadratic and Linear Classifier

Quadratic Discriminant Analysis (QDA) assumes the observed feature vector  $\underline{x}$  follows a Gaussian conditional density distribution. A multivariate Gaussian density function is given by,

$$p(\underline{x}|\omega_i) = \frac{1}{(2\pi)^{-\frac{p}{2}} |\Sigma_i|^{-\frac{1}{2}}} e\left[-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i)\right] \quad (4.1)$$

where  $\{p, \omega_i, \mu_i \text{ and } \Sigma_i\}$  respectively represent dimensional space,  $i^{\text{th}}$  class,  $i^{\text{th}}$  class mean vector and covariance matrix. QDA assumes that the covariance matrices and mean vectors are not identical for each class. The quadratic discriminant function may be formulated as [27],

$$g_i(x) = \ln\left(p(\underline{x}|\omega_i) p(\omega_i)\right) \quad (4.2)$$

And for the Gaussian distribution case, the discriminant function becomes,

$$g_i(x) = -\frac{1}{2}(\underline{x}^T \Sigma_i^{-1} \underline{x} - 2\mu_i^T \Sigma_i^{-1} \underline{x} + \mu_i^T \Sigma_i^{-1} \mu_i) + \ln(p(\omega_i)) + c_i \quad (4.3)$$

where  $c_i = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(|\Sigma_i|)$

The classification rule for choosing class  $\omega_i$  is when the estimated  $g_i(x)$  is the largest.

In comparison, Linear Discriminant Analysis (LDA) depends on the sample covariance matrix of the training data where a special case occurs when the class covariance matrices are assumed to be identical but with different mean values for every class. For this case,  $\Sigma_i = \Sigma$  and  $c_i$  becomes a constant. In LDA, different classes are assumed to exhibit the same shape but are shifted according to their mean values.

A linear classifier often performs better than a quadratic classifier for small data sets because of the pooled covariance matrix assumption. Averaging covariance matrices on the entire class can produce a higher quality estimate even if it may not be accurate. A quadratic classifier may have more flexibility when fitting the data but estimating covariance matrix for each class increases the variance of parameter estimation. The problem for QDA arises when having small data sets compared to increase in the number of dimensions and classes, because this may cause instability in the parameter estimation.

Nonetheless, there is a trade-off between having the best fit for the data and having a simpler model to work with. A simpler model may not fit the data as well as a complex model but the simpler method might perform better due to its robustness. Classifications using all resampling methods were performed on all possible combinations of features.

#### *4.4.2 Methods of resampling*

##### Equal Test-Train Data

Classifications were first performed using the quadratic and linear classifier with the resampling method of Equal Test-Train data where all data in the training set are also used for testing. In other words, this would be an optimistic estimate due to the fact that the testing data are duplicates of the training data. Therefore, to verify the accuracy of the classifier model, a resampling method was applied on the data when performing classification.

##### Jackknife (Hold-One-Out)

The jackknife method resamples data without replacement and so the training sample will not be duplicated when performing classification. The overall set consists of  $N$  number of patient from all classes to be classified and each patient has different number of row vectors. The implementation of the jackknife method in this research is on the basis of leave-one-out patient instead of per vector. The procedure involves leaving out one patient for testing data set and develops a training data set with the remaining  $N-1$  patients. For the purpose of this study, the class label for testing data set is assumed to be unknown. The classifier output will be a vector of all ones, all zeros or a mixture of both depending on how well the data is classified. This process is repeated by excluding the next patient from the overall set of data until all patients have been chosen as testing data. This is an optimum method for resampling in the sense that it uses the most amounts of data as training when doing classification.

### Cross-validation

When performing classification, it is best to have as much training data as possible to prevent instability and high variance in parameter estimation. Cross-validation is an effective resampling method without replacement for the problem of small data sets. When performing classification, the data sets are partitioned into two sets of samples for testing and training. For this study, the partitioned samples were chosen to be 30% testing data and 70% training data. The testing data are chosen randomly from the original data sets. Similar to the jackknife method, the sample data were chosen according to patients instead of by vectors. According to the available database in this study, each group of patients exhibits an approximately similar number of patients. Therefore, when performing a two-class classification, a 30-70 would result in around 6 patients chosen for testing data and the remainder  $N-6$  patients for training data. Using a random pick, 3 patients were chosen from class 1 ( $\omega_1$ ) and another 3 patients from ( $\omega_2$ ).

By using a cross-validation resampling, the output of the classifier will differ for every run. Therefore, this method was performed iteratively and the averages for all outputs were computed in order to obtain more accurate and stable parameter estimation. If the iteration run is too low, some patients may be randomly picked multiple times and some may not be picked at all. If the iteration run is too high, the computation time will increase. Thus, an optimum iteration of 100 runs was picked for implementing the analysis for this study.

For both the jackknife and cross-validation, the ratio of correctness for all data was calculated by incrementing a counter for every classified output that is equal to the original class labels for testing data and then divided over all vectors of  $N$  patients. The ratio of correctness for each class is calculated within the class itself. Therefore, for class  $i$  ( $\omega_i$ ), the ratio of correctness would be the sum of all classified outputs that are equal to the original class  $i$  testing data labels and then divided by the total number of vectors from  $N$  patient in  $\omega_i$ .

## CHAPTER V

### APPLICATIONS AND RESULTS

#### 5.1 Analysis of Male Interview Speech for High Risk/Depressed Decision

##### 5.1.1 PSD Statistical Analysis on Interview Speech

Table 5.1 summarizes the mean and standard deviation for each group collected from male interview speech. For all three groups, most of the energies contained inside PSD<sub>1</sub>, PSD<sub>2</sub> and PSD<sub>3</sub>. The depressed state exhibits a decreasing trend of mean energy ratios as the frequency sub-bands get higher. The high risk and remitted groups on the other hand, revealed a higher energy ratio in PSD<sub>2</sub> as compared to PSD<sub>1</sub>. Overall, the high standard deviation shows the results were not clustered near the mean but were more scattered around the mean.

*Table 5.1: Mean and standard deviation collected from the spectral energy ratios in male interview speech*

	<b>High risk suicidal</b>	<b>Depressed</b>	<b>Remitted</b>
<b>PSD<sub>1</sub> Ratio</b>	0.305 ± 0.153	0.383 ± 0.111	0.285 ± 0.112
<b>PSD<sub>2</sub> Ratio</b>	0.415 ± 0.099	0.360 ± 0.097	0.395 ± 0.119
<b>PSD<sub>3</sub> Ratio</b>	0.222 ± 0.080	0.185 ± 0.105	0.243 ± 0.068
<b>PSD<sub>4</sub> Ratio</b>	0.030 ± 0.039	0.027 ± 0.022	0.031 ± 0.015
<b>PSD<sub>5</sub> Ratio</b>	0.011 ± 0.012	0.020 ± 0.019	0.016 ± 0.009
<b>PSD<sub>6</sub> Ratio</b>	0.008 ± 0.006	0.013 ± 0.008	0.013 ± 0.008
<b>PSD<sub>7</sub> Ratio</b>	0.005 ± 0.004	0.006 ± 0.004	0.010 ± 0.005

The means for PSD<sub>2</sub> and PSD<sub>4</sub> of the high risk suicidal group is higher compared to the depressed group while other sub-bands display an increase in energy ratios when going from the suicidal to the depressive state. In comparing the high risk suicidal to the remitted state, the mean for PSD<sub>1</sub> and PSD<sub>2</sub> are higher for the high risk group compared to the remitted group and other sub-bands display an increase of energy ratios in the remitted group. Going from depressed to remitted, the mean energy ratios in PSD<sub>1</sub> and PSD<sub>5</sub> exhibit a decreasing trend, PSD<sub>6</sub> remains the same and other sub-bands reveal an increment in mean energy ratios.

The total summations of mean PSD<sub>1</sub> and PSD<sub>2</sub> for the high risk suicidal, depressed and remitted groups are 0.720, 0.743, and 0.680 respectively. The result illustrates that the high risk suicidal and depressed groups exhibit a larger mean value compared to the remitted group. On the other hand, the total summations of mean PSD<sub>3</sub> and PSD<sub>4</sub> for the high risk suicidal, depressed and remitted group are 0.252, 0.212 and 0.274 respectively. For the second summation, the remitted group exhibits the largest mean value, followed by high risk suicidal and depressed group. Other bands are too small to compare. These results agree with Yingthawornsuk [8].

### *5.1.2 The number of frequency sub-bands effect on classification*

The effect of using a higher number of frequency sub-bands were studied in this research. In the beginning of the research, the 2000 Hz spectral energy was divided into four equal 500 Hz bands. Classifications using an Equal-Test-Train data were performed on the basis of four bands. An optimal result of 83% correct classification for the high risk suicidal group and 60% correct classification for the depressed group were obtained

when doing classification using all four bands on male interview data. This result was not satisfying enough because an Equal-Test-Train is an overly optimistic resampling method. Therefore, continuing classification using cross-validation and jackknife resampling method would likely result in a lower percentage of correctness.

The frequency sub-bands were then increased to approximately six equal 333 Hz bands. Overall, the lower frequency bands in the range of 0-666 Hz (PSD<sub>1</sub> and PSD<sub>2</sub>) resulted in approximately 60% correct classification for both the high risk suicidal and depressed groups. The information contained in 667-999 Hz (PSD<sub>3</sub>) band gives 40% correct classification for the high risk suicidal group and 70% for the depressed group. On the other hand, higher frequency bands in the range of 1000-1667 Hz (PSD<sub>4</sub> and PSD<sub>5</sub>) exhibit an approximately 80% correct classification for high risk suicidal and 20% for depressed. These results were highly variable and showed an extreme difference in the percentage of correctness between the two classified groups. Therefore, eight 250 Hz equal bands were extracted from the PSD and analyzed.

### *5.1.3 Feature classification using all patients' data*

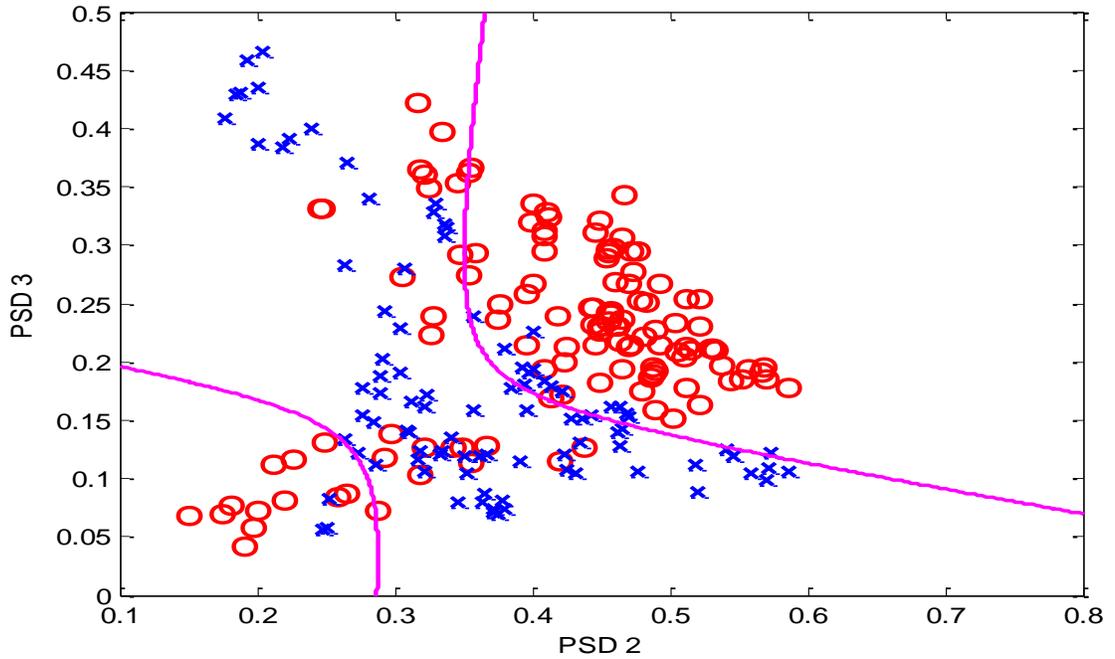
The acoustical features collected from all patients' data comprised of the high risk and depressed groups were classified using both linear and quadratic classifiers. Three types of resampling methods were implemented and the percentages of correctness were recorded in Table 5.2. All data percentages indicate feature vectors that are correctly classified over all classified groups. The high risk and depressed percentages represents the percentage of feature vectors that are correctly classified within the high risk suicidal group and the depressed group respectively. These optimal results of classification were

also obtained through a variety of band combinations and using either linear or quadratic classifiers.

*Table 5.2: High risk/depressed features classification using all patients' data.*

Percentage %	All Data	High Risk	Depressed	Band Combination	Classifier
<b>Equal-Test-Train</b>	81	81	81	PSD <sub>1</sub> , PSD <sub>3</sub>	QDA
<b>Jackknife</b>	70	77	61	PSD <sub>1</sub> , PSD <sub>6</sub> , PSD <sub>7</sub>	LDA
<b>Cross-validation</b>	66	66	61	PSD <sub>1</sub> , PSD <sub>6</sub>	LDA

Classification using an Equal-Test-Train resampling method resulted in 81% high risk and 81% depressed feature vectors that are correctly classified. This result implies that a quadratic discriminator using PSD<sub>2</sub> and PSD<sub>3</sub> features were equally effective in classifying both groups. A few other combinations of bands also yield better classification results but most of them used a higher number of bands (high dimensionality). Table 5.3 shows other good results of classification using the Equal-Test-Train resampling. According to a rule of thumb for finding an adequate sample size, a maximum number of bands that would be reasonable for classification would be five divided into the total number of patients (sample) that are being analyzed [28]. An optimal classification result that would fit the 22 patients would be approximately using four bands/features or less. Therefore, lower dimensional features may not produce a classification result as well as higher dimensional features do, but the result would be more robust in the sense that it justifies the rule of dimensionality of features with respect to number of variables.



**Figure 5.1:** Plot of patient's distribution on PSD<sub>2</sub> with respect to PSD<sub>3</sub> and a decision boundary using Equal-Test-Train quadratic classifier with all interview patients' data. 'o' indicates high risk suicidal patient and 'x' indicates depressed patients

Due to the fact that classification using PSD<sub>2</sub> to PSD<sub>7</sub> produced the best results as shown in table 5.3, classification using a lower dimensionality that consists of PSD<sub>2</sub> and PSD<sub>3</sub> would be the closest to representing the optimal result for the Equal-Test-Train classification. Additionally, two bands is easier to visualize and it also produced a result that was almost equally effective. Therefore, results from Equal-Test-Train classification using PSD<sub>2</sub> and PSD<sub>3</sub> were chosen compared to the three bands (PSD<sub>2</sub>, PSD<sub>3</sub> and PSD<sub>6</sub>). Figure 5.1 illustrates the patient's distribution for the plot of PSD<sub>2</sub> with respect to PSD<sub>3</sub> together with a plot of a quadratic classifier.

Jackknife classification correctly classified 77% of the high risk suicidal group and 61% of the depressed group with a linear discriminator classifier using PSD<sub>1</sub>, PSD<sub>6</sub>,

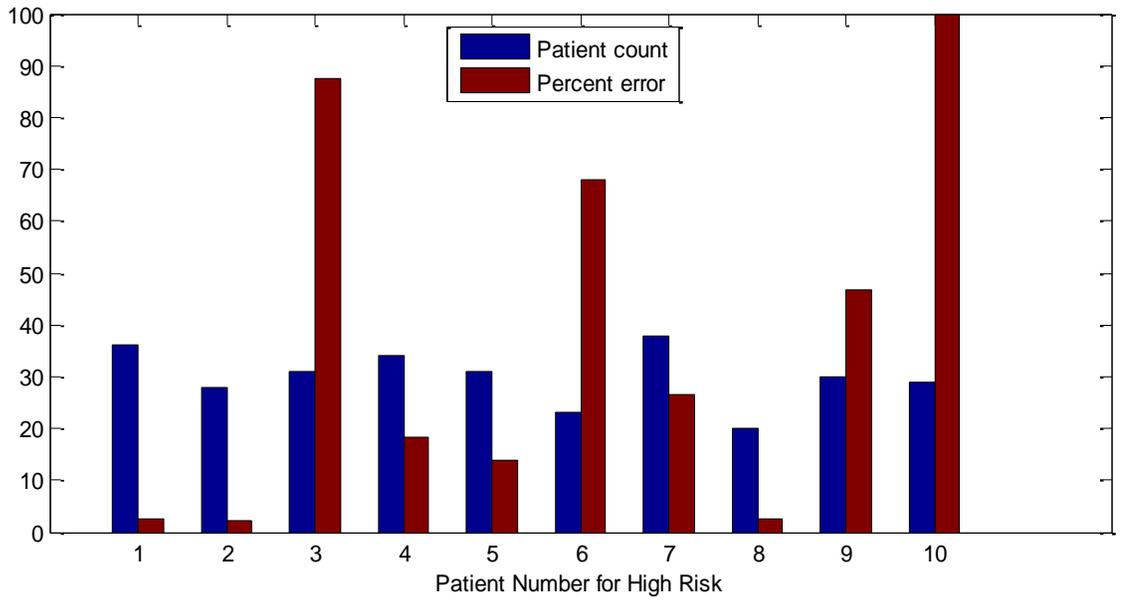
and PSD7. On the other hand, cross-validation managed to correctly classify 66% of the high risk suicidal group and 61% of the depressed group with a linear discrimination using PSD<sub>1</sub> and PSD<sub>6</sub>. These results indicate that the linear classifier works better in classifying the high risk suicidal group compared to the depressed group. Using all patients' data, these were the finest results that could be obtained from the Jackknife and Cross-validation classifications. In order to obtain a better percentage of correctness for both groups, further analysis was done on the Jackknife and Cross-validation classifications using information obtained from Equal-Test-Train classification.

**Table 5.3:** Reasonable Equal-Test-Train classification results using all patients' data

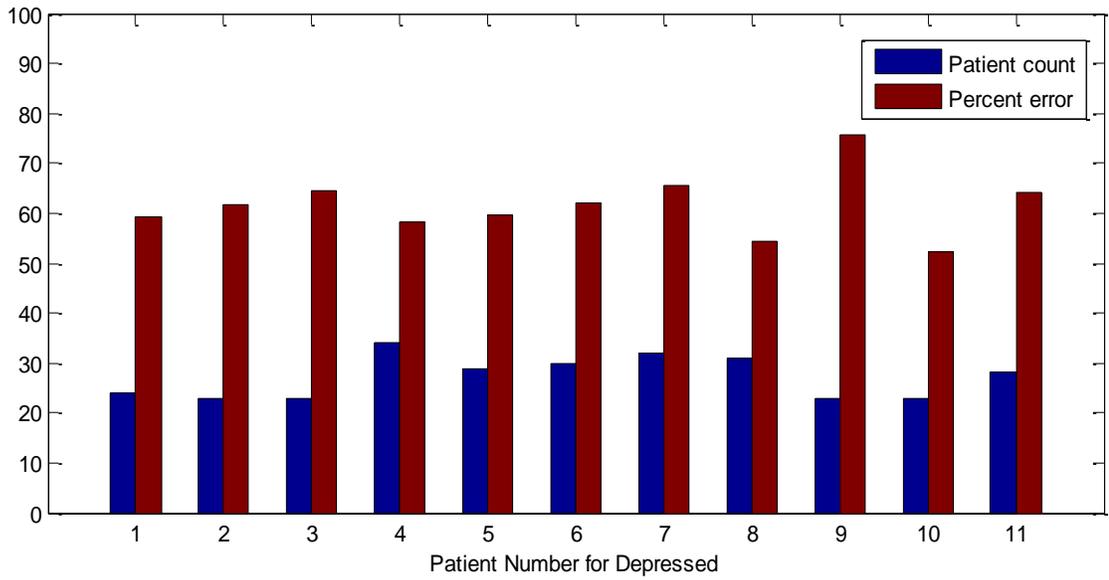
<b>Band Combination</b>	<b>All data %</b>	<b>High Risk %</b>	<b>Depressed %</b>	<b>Classifier</b>
PSD <sub>1</sub> – PSD <sub>6</sub>	85	85	84	QDA
PSD <sub>1</sub> – PSD <sub>3</sub>	83	80	85	QDA
PSD <sub>2</sub> , PSD <sub>3</sub>	81	81	81	QDA
PSD <sub>2</sub> – PSD <sub>6</sub>	83	82	86	QDA
PSD <sub>2</sub> – PSD <sub>7</sub>	89	89	88	QDA
PSD <sub>1</sub> , PSD <sub>3</sub>	82	82	81	QDA
PSD <sub>1</sub> , PSD <sub>2</sub> , PSD <sub>6</sub>	82	84	80	QDA
PSD <sub>1</sub> , PSD <sub>3</sub> , PSD <sub>6</sub>	81	81	81	QDA
PSD <sub>2</sub> , PSD <sub>3</sub> , PSD <sub>6</sub>	82	82	81	QDA

#### *5.1.4 Improving results for Jackknife and Cross-validation Classification*

In order to improve the results for classification using jackknife and cross-validation, an error histogram was constructed to identify which patients significantly affect the classification error. In particular, we are looking for outlier patients. The error histogram displays how many times each particular patient was chosen (patient count) and their percentage of error (percent error). Based on results from 100 iterations of classification with cross-validation and using PSD<sub>2</sub> and PSD<sub>3</sub>, an error histogram as shown in Figure 5.2 was produced. A further look into the result revealed that the third and the last patient from the high risk group show a significantly high percentage of error in classification with 89% and 100% respectively. To verify that the third and the last patient were truly affecting classification result, an error histogram was constructed for each band combination shown in Table 5.3. The outcome from each error histogram yielded the same two patients with highest error classification. For depressed patients, the error distribution as shown in Figure 5.3 was consistent across all patients. Therefore, no particular patient was considered as an ‘outlier’.



**Figure 5.2:** Error histogram from interview speech for high risk patients using  $PSD_2$  and  $PSD_3$



**Figure 5.3:** Error histogram from interview speech for depressed patients using  $PSD_2$  and  $PSD_3$

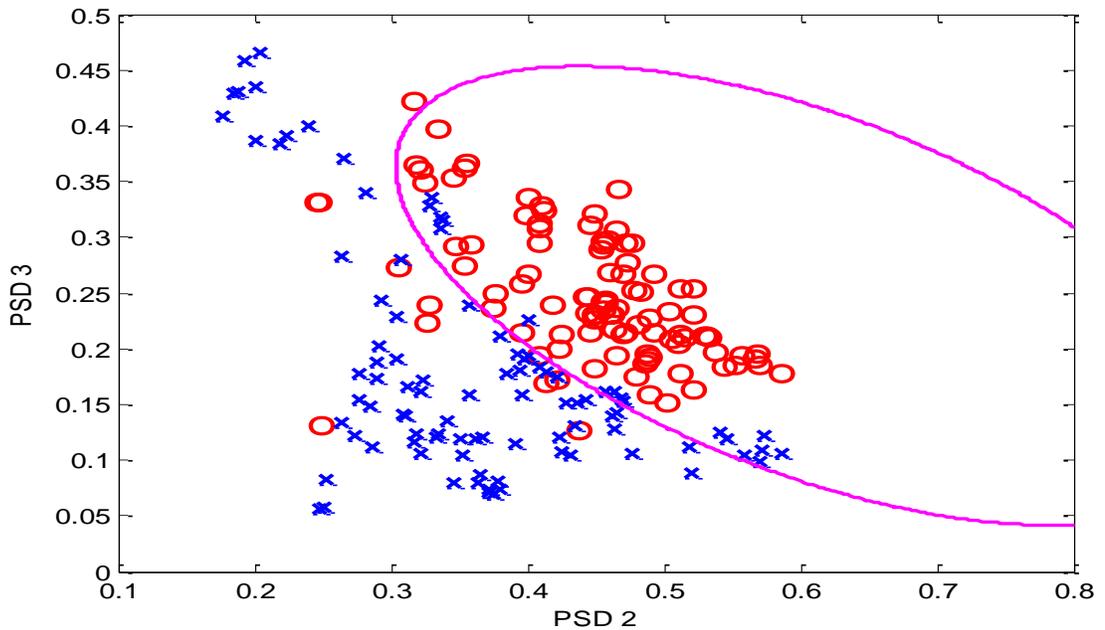
Aside from the error histogram result, performing jackknife classification using PSD<sub>2</sub> and PSD<sub>3</sub> also revealed the same outcome. The third patient and the last patient from the high risk suicidal group had a total of 7 and 14 vectors respectively and all feature vectors from these two patients were incorrectly misclassified.

Based on these results, these two patients were removed. Using the remaining 8 patients with 102 vectors from the high risk suicidal group and 11 patients with 95 vectors from the depressed group, classifications using all three resampling methods were repeated.

*Table 5.4: High risk/depressed features classification after removing outliers*

Percentage %	All Data	High Risk	Depressed	Band Combination	Classifier
<b>Equal-Test-Train</b>	86	89	82	PSD <sub>2</sub> , PSD <sub>3</sub>	QDA/LDA
<b>Jackknife</b>	83	88	79	PSD <sub>1</sub> , PSD <sub>2</sub> , PSD <sub>6</sub>	LDA
<b>Cross-validation</b>	82	88	74	PSD <sub>1</sub> , PSD <sub>2</sub> , PSD <sub>6</sub>	LDA

According to the result shown in Table 5.4, the percentage of correctness for Equal-Test-Train classification using a quadratic classifier increases significantly for the high risk while also effectively classifying depressed patient. Figure 5.4 the patient's distribution after removing the two outlier patients with a plot of a quadratic classifier representing the result from Equal-Test-Train classification. The percentage of correctness for classification using Jackknife and Cross-validation also improved significantly after removing the two high risk patients. Both methods produce the optimal result with a linear classifier and using three features of PSD<sub>1</sub>, PSD<sub>2</sub> and PSD<sub>6</sub>.



**Figure 5.4:** Plot of patient's distribution on  $PSD_2$  with respect to  $PSD_3$  and a quadratic decision boundary using Equal-Test-Train quadratic classifier after removing two high risk patients (outliers). 'o' indicates high risk suicidal patient and 'x' indicates depressed patients

## 5.2 Analysis of Male Reading for the High Risk/Depressed Decision

For the male reading speech samples, four bands of features were extracted and results from the analysis showed a consistent percentage of correctness for both classified groups. Therefore, further analysis was performed using four equal 500 Hz bands of spectral energy.

### 5.2.1 PSD Statistical Analysis of the Reading Speech

Table 5.5 summarizes mean and standard deviations of the spectral energy ratios collected from male reading speech for all mental states. Depressed patients exhibit the highest energy ratio for frequencies less than 500 Hz and the lowest energy ratio for a

frequency range of 500-1000 Hz compared to other patient groups. Greater energy was observed for remitted patients compared to high risk suicidal patients for both the PSD<sub>1</sub> and PSD<sub>2</sub> bands. The result does not agree with Yingthawornsuk [8] where he showed that remitted patients yielded higher energy than high risk suicidal patients.

*Table 5.5: Mean and standard deviation collected from the spectral energy ratios in male reading speech*

	<b>High risk suicidal</b>	<b>Depressed</b>	<b>Remitted</b>
<b>PSD<sub>1</sub> Ratio</b>	0.708 ± 0.128	0.790 ± 0.119	0.715 ± 0.131
<b>PSD<sub>2</sub> Ratio</b>	0.260 ± 0.118	0.179 ± 0.117	0.248 ± 0.128
<b>PSD<sub>3</sub> Ratio</b>	0.018 ± 0.008	0.018 ± 0.009	0.024 ± 0.018

### 5.2.2 Feature classification using all patient's data

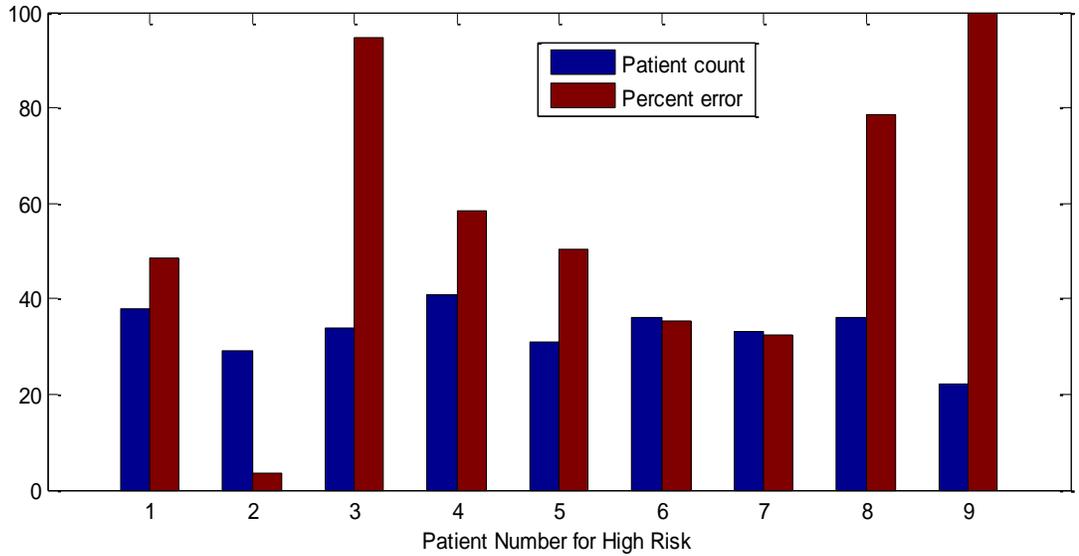
For each resampling method of classification using all patients' data, even though the results did not achieve the level of satisfaction, the results revealed a consistent percentage of correctness for all types of patients. Table 5.6 illustrates the best outcome for classification that was performed from all possible combination of bands. Compared to interview results using all patients' data, the classifier performed better in classifying depressed patients than high risk suicidal patients for reading speech samples.

*Table 5.6: High risk/depressed features classification using all patients' data.*

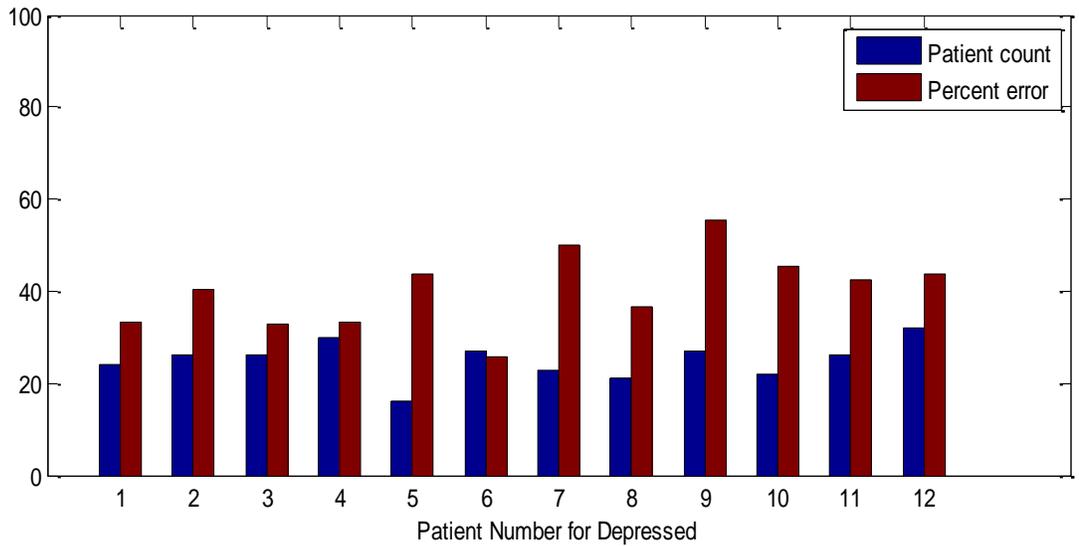
<b>Percentage %</b>	<b>All Data</b>	<b>High Risk</b>	<b>Depressed</b>	<b>Band Combination</b>	<b>Classifier</b>
<b>Equal-Test-Train</b>	71	65	77	PSD <sub>1</sub> , PSD <sub>2</sub> , PSD <sub>3</sub>	LDA
<b>Jackknife</b>	65	62	69	PSD <sub>2</sub>	LDA
<b>Cross-validation</b>	69	65	73	PSD <sub>1</sub>	LDA

### *5.2.3 Improving results for Jackknife and Cross-validation Classification*

Further analysis was conducted on the reading sample to identify which patients were affecting the classification results. An error histogram was generated by doing 100 iterations on cross-validation results using PSD<sub>1</sub> to PSD<sub>3</sub>. Figure 5.5 showed the third and last high risk patient obtained the highest percentage of error when performing cross-validation classification with an error of 100% and 98% respectively. Figure 5.6 is the error histogram constructed from the depressed patients. The percentage of error for depressed was considered as average low and consistent. The patient with highest error for the depressed group was identified to be the ninth patient with an error of approximately 58%. The eighth patient in the high risk group was not considered as an outlier due to the fact that there were a limited number of high risk vectors in the database.



*Figure 5.5: Error histogram from reading speech for the high risk patients using  $PSD_1$  to  $PSD_3$  with linear classifier*



*Figure 5.6: Error histogram from reading speech for the depressed patients using  $PSD_1$  to  $PSD_3$  with linear classifier*

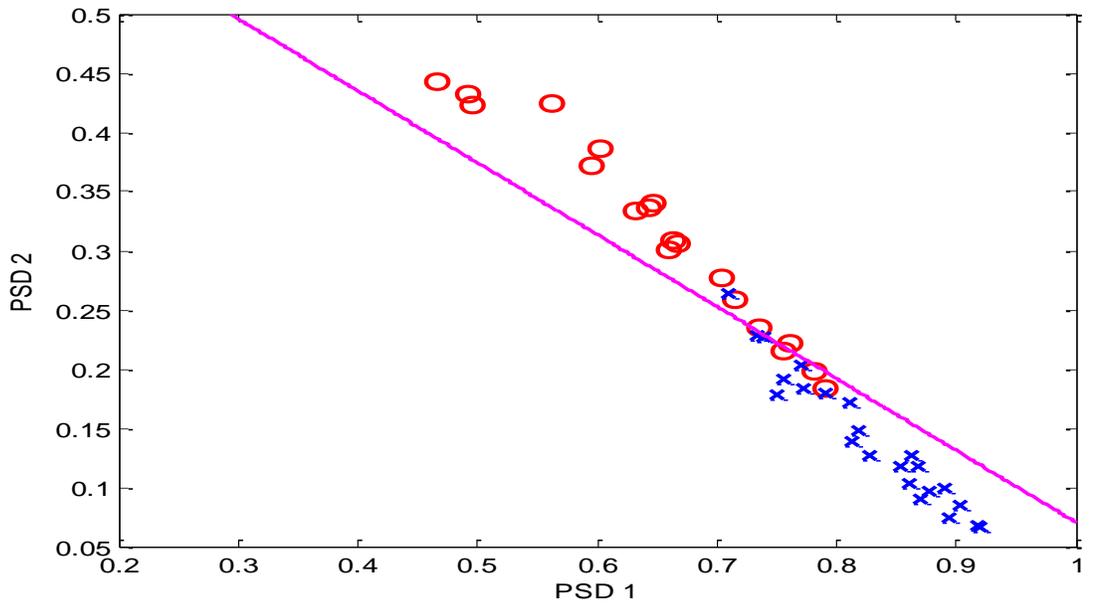
Multiple classifications using cross-validation was repeated by removing outlier patients one by one, two combinations, and finally removing all three patients. The best result was obtained when removing all outlier patients as summarized in Table 5.7.

*Table 5.7: High risk/depressed features classification after removing two high risk patients and one depressed patients (outliers)*

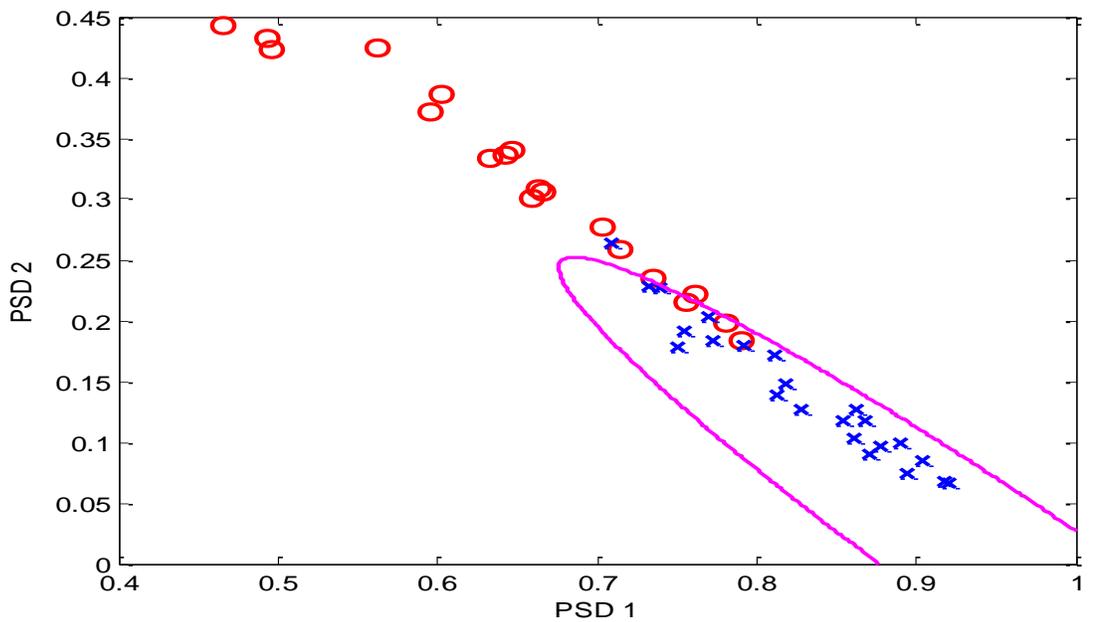
<b>Percentage %</b>	<b>All Data</b>	<b>High Risk</b>	<b>Depressed</b>	<b>Band Combination</b>	<b>Classifier</b>
<b>Equal-Test-Train</b>	90	84	95	PSD <sub>1</sub> , PSD <sub>2</sub>	LDA/QDA
<b>Jackknife</b>	86	84	87	PSD <sub>1</sub> , PSD <sub>2</sub>	LDA
<b>Cross-validation</b>	87	82	92	PSD <sub>1</sub> , PSD <sub>2</sub>	LDA

Removing all three patients resulted in 7 patients with 19 vectors for high risk and 11 patients with 23 vectors for depressed. The lower frequency band consisting of 0-1000 Hz (PSD<sub>1</sub> and PSD<sub>2</sub>) were found to be the most effective bands to be used for classification on the reading samples including the Jackknife and Cross-validation. Equal-Test-Train revealed 84% correct classification for the high risk suicidal group and 95% for the depressed group using both the linear and quadratic classifier. The performance measure for both the high risk and depressed classification using jackknife and linear classifier resulted in an almost equally effective percent of correctness. For Cross-validation, the linear classifier performed significantly better in classifying depressed than high risk with 92% and 82% respectively. Overall, the classifiers work better in classifying the depressed patients when using reading speech samples.

Figures 5.7 and 5.8 illustrate plots of PSD1 with respect to PSD2 produced by Equal-Test-Train classification using LDA and QDA on reading sample data after removing the two high risk patients and the one depressed patient. Both methods produced an optimal percentage of classification.



**Figure 5.7:** Plot of patient's distribution on  $PSD_1$  with respect to  $PSD_2$  and a decision boundary using Equal-Test-Train linear classifier after removing two high risk patients and one depressed patient (outliers). 'o' indicates high risk suicidal vectors and 'x' indicates depressed vectors



**Figure 5.8:** Plot of patient's distribution on  $PSD_1$  with respect to  $PSD_2$  and a decision boundary using Equal-Test-Train quadratic classifier after removing two high risk patients and one depressed patient (outliers). 'o' indicates high risk suicidal vectors and 'x' indicates depressed vectors

## CHAPTER 6

### CONCLUSION AND DISCUSSION

During the data pre-processing stage, speech recordings were sampled at 44.1 kHz compared to previous publications where the speech recordings were sampled at 10 kHz. Human hearing is in the range of 20 kHz and human speech frequency is in the range of 200 to 7000 Hz for typical speech activity such as talking, singing, laughing and crying [29]. The signal was sampled at least two times the highest frequency to satisfy the Nyquist Theorem. According to Katz [30] increasing sampling rates will also automatically provide signal-to-noise advantage. Therefore, speech recordings that are digitized with high sampling rate are able to effectively represent the information contained in the waveform.

An interview session would be considered as spontaneous speech because the patient is creating what he/she is saying whereas reading is speech that is controlled and the patient does not create the content of what is being said. Therefore, based on these properties, information contained in the energy spectrum might be distributed differently across all bands. According to PSD statistic for the interview (Table 5.1), when a depressed patient gets suicidal, energy in PSD<sub>1</sub>, PSD<sub>5</sub>, PSD<sub>6</sub> and PSD<sub>7</sub> (0-250Hz and 750-1750Hz) are reduced whereas the energy increases in the mid frequency bands of PSD<sub>2</sub>, PSD<sub>3</sub> and PSD<sub>4</sub> (250-750Hz). On the other hand, the PSD statistics for reading speech (Table 5.5) showed a decrease in energy for PSD<sub>1</sub> (0-500Hz) while increasing in energy for PSD<sub>2</sub> (500-1000Hz) when the patient goes from the depressed to the suicidal

state. These different properties of speech were also seen in the results where the interview speech was shown to work best as a suicidal indicator for all types of classifications (Equal-Test-Train, Jackknife and Cross-validation). However, the reading speech samples effectively classified depressed patients compared to suicidal patients and this finding agrees with previous work in [8].

Based on Figure 5.1, the distribution of the interview speech samples in  $PSD_2$  versus  $PSD_3$  were shown to have two peaks representing suicidal patients and one peak in the middle representing depressed patients. All three clouds of distributions were approximately centered at the mean. By observation analysis, one of the two peaks for the suicidal group had a significantly smaller number of samples compared to the other. After removing the sub-population, it can be seen that the covariance matrix for the two features were not significantly different for both distributions. There is a different case for the reading speech, a plot of  $PSD_1$  versus  $PSD_2$  in Figures 5.7 and 5.8 were shown to be stretched out with a decreasing slope characteristic. For both distributions, the covariance matrices for the two features were quite large but not significantly different.

The Jackknife and Cross-validation methods were applied to the classifiers in order to compensate for having a small sample size and for bias correction. Most of the high percentage of suicidal/depressed classification were contributed by  $PSD_1$ ,  $PSD_2$ ,  $PSD_3$  and  $PSD_6$  (0-750 Hz and 1250-1500 Hz) for the interview speech samples and  $PSD_1$  and  $PSD_2$  (0-1000 Hz) for the reading speech samples. The results are illustrated in Tables 5.2 and 5.7.

The percentage of correct classification increases significantly better after removing a small sub-population of patients with the highest rate of misclassification.

Having a larger sample size is particularly essential in order to accurately evaluate the effectiveness of the feature classifications. Due to the fact that this research uses a small number of features and a simple classification rule, results produced by the small sample size might be valid for larger populations. The sub-population may suggest that even with a larger sample size, there are just going to be a small group of people that are unusual. These people could possibly indicate that suicidality does not change the characteristic of the voice or perhaps their voice mechanism is damaged. Also, since the database was obtained from patients in a wide range of 18-65 years old, some of the speech recordings might belong to older people. Elderly people may experience changes in voice and thus contribute to the existence of the sub-population.

Compared to Yingthawornsuk [8], the results obtained by the Jackknife classification in this study were slightly lower but consistent when performing classifications using the entire database. These differences might be caused by different methods that were used for PSD extraction from speech samples and jackknife classification (hold-one-out) analysis. Previously related work used a statistical analysis called SYSTAT and used an On-Line Pattern Analysis System (PcOLPARS), while for this research, all analyses were done using Matlab and its Statistics Toolbox.

In this research, a small number of noisy recordings were removed. For future work, the effects of noisy speech recordings on classification can be verified. Will the classification still work the same way using noisy recordings? Using the original files, noise can be added to it and see how noisy it can get before misclassification becomes worse. Previous research including this study did not include a group called ideation in the classification analysis. The ideation group is patients with thoughts of suicidal

without actually making plans to commit suicide. In the future, the ideation group can also be included and analyzed to see if their characteristics represent more towards depressed, high risk suicidal or perhaps have distinct features from the rest of the group. Furthermore, each patient's mental state, especially the high risk suicidal patients should be labeled as accurately as possible. Previous recordings that were used by the Silvermans were strictly labeled according to the patient's actual condition where only patients who have attempted suicide will be labeled as high risk suicidal. Thus, the features obtained would result in a more definite discriminating characteristic that represents the high risk suicidal group.

The result of classification increases significantly after removing outlier patients. If the validity of removing these outliers is proven, then this study concluded that information contained in the PSD of male speech can be utilized as potential features for discriminating between high risk suicidal and depressed mental states. Obtaining more data and implementing better labeling would increase the accuracy and confidence of the classifications.

APPENDIX

A.1 Other Results for Male Interview Suicidal/Depressed

Table a1: Percentage of Equal-Test-Train classification for male interview suicidal/depressed using all patients' data

Equal-Test-Train		1:7	1:6	1:5	1:4	1:3	1:2	1	2:3	2:4	2:5	2:6	2:7	2	3:4
LINEAR	ALL	74.31	75.23	72.48	72.94	71.56	71.56	65.60	69.72	69.27	72.02	69.72	72.02	66.51	67.43
	High Risk	82.11	74.80	69.92	70.73	71.54	75.61	75.61	76.42	76.42	72.36	69.92	78.86	68.29	60.98
	Depressed	64.21	75.79	75.79	75.79	71.58	66.32	52.63	61.05	60.00	71.58	69.47	63.16	64.21	75.79
QUADRATIC	ALL	83.94	<b>84.86</b>	82.11	83.03	<b>82.57</b>	75.23	71.56	<b>81.19</b>	83.49	83.03	<b>83.95</b>	<b>88.99</b>	66.51	63.30
	High Risk	90.24	<b>85.37</b>	78.05	75.61	<b>80.49</b>	75.61	73.98	<b>81.30</b>	78.86	77.24	<b>82.11</b>	<b>89.43</b>	68.29	54.47
	Depressed	75.79	<b>84.21</b>	87.37	92.63	<b>85.26</b>	74.74	68.42	<b>81.05</b>	89.47	90.53	<b>86.32</b>	<b>88.42</b>	64.21	74.74

Equal-Test-Train		3:5	3:6	3:7	3	4:5	4:6	4:7	4	5	5:6	5:7	6	6:7	7
LINEAR	ALL	67.43	66.06	73.85	67.89	65.14	70.64	72.94	41.74	60.09	70.18	72.94	68.81	72.94	57.80
	High Risk	67.48	71.54	82.93	62.60	77.24	77.24	83.74	21.95	75.61	76.42	82.11	74.80	82.11	67.48
	Depressed	67.37	58.95	62.11	74.74	49.47	62.11	58.95	67.37	40.00	62.11	61.05	61.05	61.05	45.26
QUADRATIC	ALL	71.56	73.39	80.73	72.48	64.22	67.89	72.02	46.79	55.96	61.93	70.64	66.97	73.39	55.50
	High Risk	66.67	71.54	82.93	78.05	69.92	78.86	86.18	17.89	83.74	82.11	87.80	79.67	86.99	70.73
	Depressed	77.89	75.79	77.89	65.26	56.84	53.68	53.68	84.21	20.00	35.79	48.42	50.53	55.79	35.79

Equal-Test-Train		1,3	1,4	1,5	1,6	1,7	2,4	2,5	2,6	2,7	3,5	3,6	3,7	4,6	4,7	5,7
LINEAR	ALL	65.60							71.56							
	High Risk	75.61							69.11							
	Depressed	52.63							74.74							
QUADRATIC	ALL	<b>81.65</b>			76.61	72.94			71.56			71.10	71.56	69.27		
	High Risk	<b>82.11</b>			78.05	73.17			70.73			71.54	70.73	74.80		
	Depressed	<b>81.05</b>			74.74	72.63			72.63			70.53	72.63	62.11		

Equal-Test-Train		1,2,4	1,2,5	1,2,6	1,2,7	2,3,5	2,3,6	2,3,7	2,4,5	2,4,6	2,4,7	2,5,6	2,5,7	2,6,7	3,4,6	3,4,7
LINEAR	ALL	71.56	71.10	70.18	72.94	70.18	71.56	73.85	70.00	72.00	69.00	71.00	69.00	76.61	66.00	67.00
	High Risk	75.61	71.54	69.92	73.98	71.54	70.73	76.42	70.00	70.00	66.00	69.00	68.00	80.49	71.00	62.00
	Depressed	66.3158	70.53	70.53	71.58	68.42	72.63	70.53	70.00	74.00	73.00	73.00	70.00	71.58	58.00	74.00
QUADRATIC	ALL	79.8165	76.61	<b>82.11</b>	75.23	77.06	<b>81.19</b>	80.28	69.00	69.00	66.00	71.00	71.00	77.06	74.00	64.00
	High Risk	73.9837	86.99	<b>83.74</b>	74.80	86.18	<b>81.30</b>	79.67	61.00	60.00	52.00	82.00	78.00	84.55	66.00	52.00
	Depressed	87.3684	63.16	<b>80.00</b>	75.79	65.26	<b>81.05</b>	81.05	78.00	80.00	85.00	56.00	62.00	67.37	84.00	81.00

Equal-Test-Train		3,5,6	3,5,7	3,6,7	4,5,7	4,6,7	1,3,4	1,3,5	1,3,6	1,3,7	1,4,5	1,4,6	1,4,7	1,5,6	1,5,7	1,6,7
LINEAR	ALL							70.64	70.18			72.02				
	High Risk							71.54	69.92			71.54				
	Depressed							69.47	70.53			72.63				
QUADRATIC	ALL						81.65	81.65	<b>81.65</b>	78.90		77.06	74.31			
	High Risk						77.24	88.62	<b>82.11</b>	77.24		74.80	73.17			
	Depressed						87.37	72.63	<b>81.05</b>	81.05		80.00	75.79			

\*Note: Yellow - the optimum classification percentage  
 Dark blue- other reasonable consistent and high percentage  
 stripe - bad results

**Table a2: Percentage of Jackknife classification for male interview suicidal/depressed using all patients' data**

Jackknife		1:7	1:6	1:5	1:4	1:3	1:2	1	2:3	2:4	2:5	2:6	2:7	2	3:4
LINEAR	ALL	60.09	51.38	50.92	56.42	57.34	59.63	65.60	64.68	53.21	56.88	51.38	62.39	66.06	54.59
	High Risk	64.23	57.72	56.10	56.91	63.41	68.29	75.61	76.42	60.16	58.54	54.47	64.23	68.29	41.46
	Depressed	54.74	43.16	44.21	55.79	49.47	48.42	52.63	49.47	44.21	54.74	47.37	60.00	63.16	71.58
QUADRATIC	ALL	51.38	47.25	45.87	47.25	55.05	59.17	63.30	62.39	55.96	49.08	48.62	55.05	65.14	43.12
	High Risk	62.60	60.98	57.72	49.59	56.91	60.98	63.41	68.29	59.35	53.66	60.16	65.04	67.48	28.46
	Depressed	36.84	29.47	30.53	44.21	52.63	56.84	63.16	54.74	51.58	43.16	33.68	42.11	62.11	62.11

Jackknife		3:5	3:6	3:7	3	4:5	4:6	4:7	4	5	5:6	5:7	6	6:7	7
LINEAR	ALL	49.08	49.54	57.80	67.89	54.13	58.72	60.09	31.65	59.63	65.14	66.97	69.72	70.64	58.26
	High Risk	47.97	54.47	58.54	62.60	60.98	60.98	64.23	4.07	75.61	73.98	78.05	74.80	79.67	68.29
	Depressed	50.53	43.16	56.84	74.74	45.26	55.79	54.74	67.37	38.95	53.68	52.63	63.16	58.95	45.26
QUADRATIC	ALL	49.08	49.08	55.05	60.09	17.43	49.54	53.21	37.61	48.17	52.75	60.55	66.06	64.68	49.08
	High Risk	44.72	59.35	69.92	69.92	8.94	64.23	65.85	4.88	81.30	77.24	82.11	78.05	77.24	69.92
	Depressed	54.74	35.79	35.79	47.37	28.42	30.53	36.84	80.00	5.26	21.05	32.63	50.53	48.42	22.11

**TWO COMBINATION**

Jackknife		1,3	1,4	1,5	1,6	1,7	2,4	2,5	2,6	2,7	3,5	3,6	3,7	4,6	4,7	5,7
LINEAR	ALL	61.01		67.43	67.43			64.68	69.27							
	High Risk	73.17		69.92	69.11			63.41	66.67							
	Depressed	45.26		64.21	65.26			66.32	72.63							
QUADRATIC	ALL	58.72			66.51			50.46	65.14							
	High Risk	64.23			68.29			69.92	67.48							
	Depressed	51.58			64.21			25.26	62.11							

**THREE COMBINATION**

Jackknife		1,2,4	1,2,5	1,2,6	1,2,7	2,3,5	2,3,6	2,3,7	2,4,5	2,4,6	2,4,7	2,5,6	2,5,7	2,6,7	3,4,6	3,4,7
LINEAR	ALL			66.51										70.18		
	High Risk			68.29										71.54		
	Depressed			64.21										68.42		
QUADRATIC	ALL															
	High Risk															
	Depressed															

Jackknife		3,5,6	3,5,7	3,6,7	4,5,7	4,6,7	1,3,4	1,3,5	1,3,6	1,3,7	1,4,5	1,4,6	1,4,7	1,5,6	1,5,7	1,6,7
LINEAR	ALL															66.06
	High Risk															69.92
	Depressed															61.05
QUADRATIC	ALL															
	High Risk															
	Depressed															

\*Note: Yellow - the optimum classification percentage  
 Dark pink- other reasonable consistent percentage  
 stripe - bad results

**Table a3: Percentage of Cross-validation classification for male interview suicidal/depressed using all patients' data**

Cross-validation		1:7	1:6	1:5	1:4	1:3	1:2	1	2:3	2:4	2:5	2:6	2:7	2	3:4
LINEAR	ALL	54.65	53.96	51.23	54.76	55.71	59.08	64.15	61.70	56.19	55.57	54.00	58.67	59.68	50.19
	High Risk	56.60	55.11	53.46	52.38	60.37	63.76	73.16	70.20	59.28	58.65	55.45	61.94	59.12	41.81
	Depressed	49.87	50.49	44.78	53.29	47.51	54.62	52.29	50.54	53.26	47.75	47.12	51.50	59.31	63.10
QUADRATIC	ALL	51.61	48.77	47.20	45.44	49.51	51.28	63.69	59.60	54.61	49.99	49.78	52.39	56.42	47.88
	High Risk	59.58	55.92	54.28	47.22	56.83	59.24	65.18	66.47	59.50	53.21	53.60	60.93	57.20	41.66
	Depressed	40.11	37.74	36.24	41.36	39.20	40.89	64.17	50.06	48.79	42.77	41.40	39.71	54.00	57.97

Cross-validation		3:5	3:6	3:7	3	4:5	4:6	4:7	4	5	5:6	5:7	6	6:7	7
LINEAR	ALL	47.50	50.13	55.98	61.04	48.39	57.00	57.73	39.24	56.02	60.81	65.02	66.66	66.96	48.54
	High Risk	42.36	52.47	54.04	57.57	51.56	59.79	59.08	31.98	64.56	68.22	71.69	73.31	69.51	56.51
	Depressed	51.38	41.13	54.97	64.95	40.64	47.96	50.11	52.16	39.29	44.30	52.20	51.43	57.25	33.50
QUADRATIC	ALL	43.78	47.98	54.20	58.25	36.68	47.34	51.54	42.28	47.93	53.09	58.83	64.42	66.90	45.52
	High Risk	42.51	52.30	57.76	63.36	35.41	51.59	55.66	22.87	67.74	70.92	74.43	75.62	74.77	53.72
	Depressed	46.93	37.92	47.36	50.54	38.53	38.60	41.00	69.14	16.62	21.77	35.01	42.81	50.18	30.08

**TWO COMBINATION**

Cross-validation		1,3	1,4	1,5	1,6	1,7	2,4	2,5	2,6	2,7	3,5	3,6	3,7	4,6	4,7	5,7
LINEAR	ALL				66.50											
	High Risk				66.92											
	Depressed				61.28											
QUADRATIC	ALL				65.48											
	High Risk				65.27											
	Depressed				61.42											

\*Note: Yellow - the optimum classification percentage  
 Dark green - other reasonable consistent percentage  
 stripe - bad results

**THREE COMBINATION** (all three combinations gave bad results)

**Table a4: Percentage of Cross-validation classification for male interview suicidal/depressed after removing outliers**

**INTERVIEW - Redo CROSS VALIDATION**

Redo analysis on cross validation after removing two patients from HR- 2 outliers: h030905nt2, h103105nt2

\* these outliers were taken out based on results from error histogram in alldata 2,3

LINEAR	1:7	1:6	1:5	1:4	1:3	1:2	1	2:3	2:4	2:5	2:6	2:7	2	3:4
ALL	71.90	71.80	71.63	72.88	71.60	72.67	73.15	77.83	73.42	73.20	72.74	72.25	69.39	58.72
High Risk	77.59	76.30	77.27	73.68	76.97	83.84	83.18	86.83	77.70	79.45	76.64	76.86	71.60	47.14
Depressed	62.17	64.48	61.62	69.42	62.83	57.20	60.45	65.00	68.26	63.70	64.53	64.83	67.49	75.21
QUADRATIC														
ALL	66.16	64.36	60.42	61.23	65.63	68.51	72.49	78.21	67.53	61.63	63.92	66.80	68.84	57.08
High Risk	67.59	65.26	62.01	60.80	68.98	76.44	83.11	83.35	71.62	65.30	63.76	68.06	74.57	50.54
Depressed	62.72	62.12	57.13	60.97	61.02	58.06	59.06	70.61	64.72	57.38	61.47	64.72	62.32	68.10

LINEAR	3:5	3:6	3:7	3	4:5	4:6	4:7	4	5	5:6	5:7	6	6:7	7
ALL	51.38	54.37	59.56	69.15	44.37	44.37	55.75	38.19	49.11	63.74	62.91	67.13	63.69	50.70
High Risk	40.65	51.61	52.76	64.54	38.08	38.08	57.32	25.08	56.27	71.64	69.24	69.59	67.62	56.13
Depressed	67.77	54.39	66.27	77.38	48.95	48.95	51.15	55.83	35.24	45.87	48.85	59.07	54.11	40.99
QUADRATIC														
ALL	53.59	58.33	66.82	73.10	38.53	38.53	52.89	41.34	45.24	56.95	61.52	63.64	62.22	46.75
High Risk	46.78	57.06	64.55	74.59	31.91	31.91	57.12	22.65	60.28	71.29	73.67	70.84	69.73	53.04
Depressed	65.84	57.99	67.79	71.87	46.71	46.71	44.29	66.89	21.15	29.34	38.41	48.69	47.91	37.85

**TWO COMBINATIONS**

LINEAR	1,3	1,4	1,5	1,6	1,7	2,4	2,5	2,6	2,7	3,5	3,6	3,7	4,6	4,7	5,7
ALL	69.62	71.33	75.31	82.04	73.18	65.96	63.80	71.54	71.13	63.10	68.75	60.15	56.38	37.44	45.64
High Risk	81.28	75.38	83.95	85.90	82.49	63.83	65.62	66.94	67.11	59.39	72.05	54.58	50.68	31.44	51.39
Depressed	55.20	66.34	61.75	74.83	59.25	71.95	59.14	72.51	74.37	65.16	61.46	67.40	61.15	46.36	33.07
QUADRATIC															
ALL	68.90	72.29	70.02	76.88	69.50	56.27	58.19	68.18	66.29	59.54	69.01	60.10	46.41	36.59	42.15
High Risk	74.89	77.04	79.40	77.67	75.61	50.70	63.60	63.22	64.00	61.19	65.42	56.89	36.41	23.98	58.57
Depressed	63.75	65.61	55.40	73.46	59.63	67.29	48.90	70.00	66.85	55.74	73.30	63.64	59.90	55.64	15.27

**THREE COMBINATIONS**

LINEAR	1,2,4	1,2,5	1,2,6	1,2,7	2,3,5	2,3,6	2,3,7	2,4,5	2,4,6	2,4,7	2,5,6	2,5,7	2,6,7	3,4,6	3,4,7
ALL	64.39	74.71	82.45	73.17	75.10	78.74	73.72	68.51	70.63	64.50	69.64	63.30	71.31	58.03	49.16
High Risk	70.03	84.26	87.96	82.74	81.68	82.80	78.57	66.00	66.92	59.37	71.26	60.67	71.73	53.35	37.76
Depressed	57.33	60.19	73.73	59.69	64.68	70.87	64.34	70.60	72.31	71.07	64.15	65.46	67.52	61.87	67.14
QUADRATIC															
ALL	63.83	64.87	74.73	64.53	67.59	75.26	67.88	52.21	54.25	50.64	65.80	58.20	68.41	55.48	51.47
High Risk	68.35	74.48	75.72	66.69	71.95	75.55	66.59	46.65	46.37	42.89	66.35	58.58	67.69	51.29	44.00
Depressed	59.30	52.39	72.99	61.11	61.84	73.14	66.54	59.28	63.01	62.74	61.48	57.30	67.09	63.45	64.55

LINEAR	3,5,6	3,5,7	3,6,7	4,5,7	4,6,7
ALL	66.29	58.44	69.62	42.94	56.23
High Risk	70.04	55.80	68.73	42.88	56.78
Depressed	56.64	59.57	67.31	42.61	51.75
QUADRATIC					
ALL	65.40	56.43	73.51	39.60	52.47
High Risk	65.89	59.11	69.77	37.16	51.50
Depressed	62.63	52.80	77.60	42.43	50.53

\*Note: Yellow - the optimum classification percentage  
Dark green- other reasonable consistent and high percentage

**Table a5: Percentage of Jackknife classification for male interview suicidal/depressed after removing outliers**

**INTERVIEW - Redo JACKKNIFE**

Redo analysis on cross validation after removing two patients from HR- 2 outliers: h030905nt2, h103105nt2

\* these outliers were taken out based on results from error histogram in alldata 2,3

<b>LINEAR</b>		<b>1:7</b>	<b>1:6</b>	<b>2:3</b>
ALL		75.13	73.10	77.67
High Risk		81.37	80.39	88.24
Depressed		68.42	65.26	66.32
<b>QUADRATIC</b>				
ALL		67.51	67.01	<b>78.17</b>
High Risk		70.59	71.57	<b>84.31</b>
Depressed		64.21	62.11	<b>71.58</b>

**TWO COMBINATIONS**

<b>LINEAR</b>		<b>1,6</b>	<b>2,3</b>
ALL		<b>81.73</b>	77.67
High Risk		<b>84.31</b>	88.24
Depressed		<b>78.95</b>	66.32
<b>QUADRATIC</b>			
ALL		78.68	78.17
High Risk		78.43	84.31
Depressed		78.95	71.58

**THREE COMBINATIONS**

<b>LINEAR</b>		<b>1,2,6</b>	<b>2,3,6</b>	<b>1,3,6</b>	<b>1,4,6,</b>
ALL		<b>82.74</b>	<b>79.19</b>	<b>80.7107</b>	<b>81.2183</b>
High Risk		<b>87.25</b>	<b>84.31</b>	<b>84.3137</b>	<b>84.3137</b>
Depressed		<b>77.89</b>	<b>73.68</b>	<b>76.8421</b>	<b>77.8947</b>
<b>QUADRATIC</b>					
ALL		78.17	77.67	74.6193	74.6193
High Risk		77.45	79.41	73.5294	76.4706
Depressed		78.95	75.79	75.7895	72.6316

\*Note: Yellow - the optimum classification percentage  
Dark pink- other reasonable consistent and high percentage

## A.2 Other Results for Male Reading Suicidal/Depressed

*Table a6: Percentage of classification for male reading suicidal/depressed using all patients' data*

LINEAR		1:2	1:3	2:3	1	2	3	1,3
Equal-Test-Train	ALL	69.2308	<b>71.15</b>	69.23	67.31	69.23	53.85	69.23
	High Risk	61.54	<b>65.38</b>	61.54	57.69	61.54	57.69	61.54
	Depressed	76.92	<b>76.92</b>	76.92	76.92	76.92	50.00	76.92
jackknife	ALL	57.69	51.92	59.62	63.46	<b>65.38</b>	17.31	57.69
	High Risk	46.15	42.31	57.69	57.69	<b>61.54</b>	23.08	53.85
	Depressed	69.23	61.54	61.54	69.23	<b>69.23</b>	11.54	61.54
crossval	ALL	54.59	49.42	59.46	<b>68.97</b>	66.11	40.37	58.52
	High Risk	46.65	39.73	49.55	<b>65.43</b>	60.46	38.10	52.85
	Depressed	65.76	63.07	72.54	<b>73.16</b>	74.58	44.48	67.13

QUADRATIC		1:2	1:3	2:3	1	2	3	1,3
Equal-Test-Train	ALL	65.38	61.54	<b>69.23</b>	69.23	67.31	53.85	69.23
	High Risk	50.00	42.31	<b>65.38</b>	57.69	57.69	76.92	61.54
	Depressed	80.77	80.77	<b>73.08</b>	80.77	76.92	30.77	76.92
jackknife	ALL	42.31	34.62	50.00	61.54	63.46	25.00	48.08
	High Risk	23.08	19.23	42.31	53.85	57.69	46.15	38.46
	Depressed	61.54	50.00	57.69	69.23	69.23	3.85	57.69
crossval	ALL	44.46	38.72	49.60	60.74	<b>64.77</b>	35.04	47.33
	High Risk	29.95	25.23	40.86	52.60	<b>60.29</b>	37.38	39.02
	Depressed	64.66	58.22	61.65	70.74	<b>71.72</b>	33.06	59.70

\*Note: Yellow - the optimum classification percentage  
Other colored boxes - other reasonable consistent and high percentage

**Table a7: Percentage of Cross-validation and Jackknife classification for male reading suicidal/depressed after removing outliers**

1. from 4 bands male reading, do all data for all possible bands
2. do error histogram on the best all data result (band 1:3 Linear) HR9 - 1031
3. identified that HR9 (100%) and HR3 (98%) and DEP9 (55%) HR3 - 0309
4. for crossval: DEP9 - 0901
  - do : a) delete HR9 only then do crossval
  - b) delete HR9 and DEP9 then do crossval
  - c) delete DEP9 only then do crossval
  - d) delete HR9, HR3 then do crossval
  - e) delete all then do crossval

**Cross-validation**

**a) delete HR9 only then do crossval**

crossval		1:2	1:3	2:3	1	2	3	1,3
LINEAR	ALL	64.53	54.41	65.74	69.69	71.85	37.82	68.21
	High Risk	55.96	46.01	58.83	62.89	66.36	29.65	59.37
	Depressed	76.07	67.80	75.39	79.83	80.96	49.35	80.71
QUAD	ALL	50.41	41.93	56.05	62.71	69.60	30.10	52.71
	High Risk	34.17	27.48	48.53	55.82	66.28	20.04	42.24
	Depressed	72.11	64.17	67.13	73.05	75.97	44.23	67.73

**b) delete HR9 and DEP9 then do crossval**

crossval		1:2	1:3	2:3	1	2	3	1,3
LINEAR	ALL	79.19	73.92	78.58	72.53	76.00	41.31	77.29
	High Risk	76.77	67.01	72.75	65.88	69.03	39.11	72.82
	Depressed	82.93	83.48	86.50	82.19	86.19	44.12	83.55
QUAD	ALL	67.43	58.68	71.76	72.02	75.79	36.71	71.35
	High Risk	57.31	50.24	63.46	64.14	68.74	42.55	66.37
	Depressed	81.24	72.21	83.43	83.02	86.02	30.31	78.79

**c) delete DEP9 only then do crossval**

crossval		1:2	1:3	2:3	1	2	3	1,3
LINEAR	ALL	70.66						72.68
	High Risk	65.29						69.56
	Depressed	77.62						76.31
QUAD	ALL							67.26
	High Risk							60.89
	Depressed							75.12

**d) delete HR9, HR3 then do crossval**

crossval		1:2	1:3	2:3	1	2	3	1,3
LINEAR	ALL	70.96			77.72	78.00		75.74
	High Risk	60.47			72.29	71.43		67.84
	Depressed	85.24			85.05	85.63		86.27
QUAD	ALL	60.21			68.60	74.99		64.19
	High Risk	44.67			62.17	68.92		51.71
	Depressed	81.21			77.45	82.46		80.58

**e) delete all then do crossval**

crossval		1:2	1:3	2:3	1	2	3	1,3
LINEAR	ALL	86.87	83.12	84.95	81.77	84.61		84.37
	High Risk	82.46	76.19	81.31	74.67	81.84		79.18
	Depressed	92.26	90.43	88.76	90.71	88.10		89.83
QUAD	ALL			79.34	81.57	85.19		77.65
	High Risk			74.76	77.39	84.43		72.55
	Depressed			85.00	87.05	85.81		83.57

\*Note: Yellow - the optimum classification percentage  
Dark green - other reasonable consistent and high percentage

**Jackknife**

jackknife		1:2	1:3	2:3	1	2	3	1,3
LINEAR	ALL	65.31	57.14	69.39	67.35	71.43		69.39
	High Risk	52.17	39.13	60.87	60.87	65.22		60.87
	Depressed	76.92	73.08	76.92	73.08	76.92		76.92
QUAD	ALL	51.02	44.90	48.98	65.31	67.35		51.02
	High Risk	30.43	21.74	39.13	60.87	65.22		39.13
	Depressed	69.23	65.38	57.69	69.23	69.23		61.54

jackknife		1:2	1:3	2:3	1	2	3	1,3
LINEAR	ALL	71.43		69.39				
	High Risk	78.26		73.91				
	Depressed	73.91		73.91				
QUAD	ALL	61.22		67.35				
	High Risk	56.52		69.57				
	Depressed	73.91		73.91				

jackknife		1:2	1:3	2:3	1	2	3	1,3
LINEAR	ALL							
	High Risk							
	Depressed							
QUAD	ALL							
	High Risk							
	Depressed							

jackknife		1:2	1:3	2:3	1	2	3	1,3
LINEAR	ALL	66.67	68.89	71.11	71.11	71.11		63.27
	High Risk	63.16	63.16	73.68	73.68	73.68		73.68
	Depressed	69.23	73.08	69.23	69.23	69.23		65.38
QUAD	ALL	51.11	44.44	57.78	57.78	64.44		46.94
	High Risk	47.37	26.32	57.89	57.89	73.68		57.89
	Depressed	53.85	57.69	57.69	57.69	57.69		46.15

jackknife		1:2	1:3	2:3	1	2	3	1,3
LINEAR	ALL	85.71	83.33	83.33	76.19	78.57		
	High Risk	84.21	78.95	78.95	73.68	78.95		
	Depressed	86.96	86.96	86.96	78.26	78.26		
QUAD	ALL	78.57	71.43	76.19	78.57	78.57		
	High Risk	68.42	57.89	78.95	78.95	78.95		
	Depressed	86.96	82.61	73.91	78.26	78.26		

\*Note: Yellow - the optimum classification percentage  
Dark pink - other reasonable consistent and high percentage

### A.3 Classification Results on Male Interview Suicidal/Remitted and Depressed/Remitted

*Table a8: Percentage of Equal-Test-Train classification for male interview suicidal/remitted and depressed/remitted using all patients' data*

	Percentage of classification %			Band	Classifier
	All data	suicidal	remitted		
<b>Suicidal/remitted</b>	<b>All data</b>	<b>suicidal</b>	<b>remitted</b>	PSD <sub>4</sub> , PSD <sub>5</sub> ,	QDA
	82	84	80	PSD <sub>6</sub>	
<b>Depressed/remitted</b>	<b>All data</b>	<b>depressed</b>	<b>remitted</b>	PSD <sub>3</sub> , PSD <sub>4</sub>	QDA
	82	80	85		

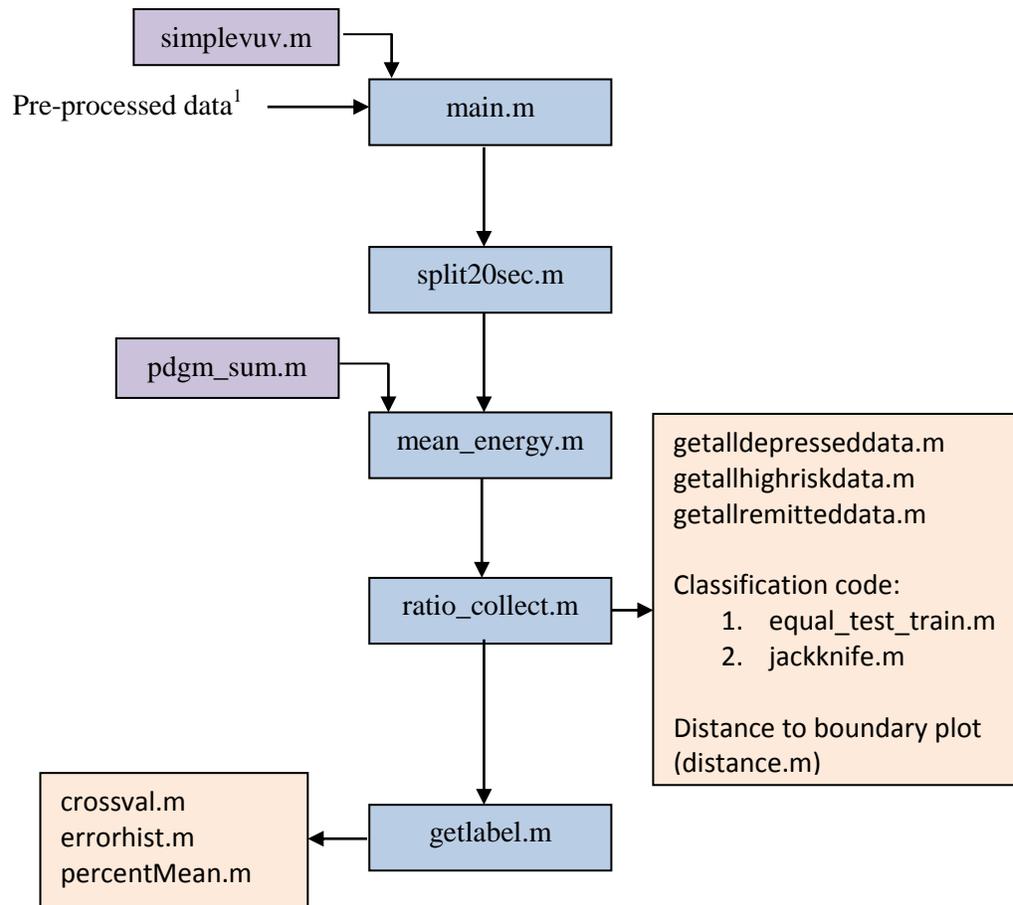
*Table a9: Percentage of Jackknife classification for male interview suicidal/remitted and depressed/remitted using all patients' data*

	Percentage of classification %			Band	Classifier
	All data	suicidal	remitted		
<b>Suicidal/remitted</b>	<b>All data</b>	<b>suicidal</b>	<b>remitted</b>	PSD <sub>4</sub> , PSD <sub>5</sub> ,	QDA
	71	73	67	PSD <sub>6</sub>	
<b>Depressed/remitted</b>	<b>All data</b>	<b>depressed</b>	<b>remitted</b>	PSD <sub>3</sub>	QDA
	72	75	68		

*Table a10: Percentage of Cross-validation classification for male interview suicidal/remitted and depressed/remitted using all patients' data*

	Percentage of classification %			Band	Classifier
	All data	suicidal	remitted		
<b>Suicidal/remitted</b>	<b>All data</b>	<b>suicidal</b>	<b>remitted</b>	PSD <sub>4</sub> , PSD <sub>7</sub>	QDA
	66	60	70		
<b>Depressed/remitted</b>	<b>All data</b>	<b>depressed</b>	<b>remitted</b>	PSD <sub>3</sub>	QDA
	71	68	71		

## A.4 Matlab Code



**Figure a11:** Steps for features extraction and analysis using Matlab

<sup>1</sup>Pre processed data are voiced speech samples that were collected after removing interviewer's voice, cutting long pauses and cutting other noises such as door slam, mouth noise or sneezing)

## main.m

```
% Code written by Nik Nur Wahidah and Wan Ahmad Hasan, Spring 2011

% function
[X,justvoiced,sumjustvoiced,sumunvoiced,sumsilent,cepsvoiced] = main %
ceps = main
function [sumjustvoiced] = main(filename) % ceps = main
[s,fs] = wavread(filename);
s = s(:,1); % for stereo typed files
s = s - mean(s);
Twin = 0.040;
Nwin = round(Twin*fs);

% Each column of X is a non overlapping frames of size Nwin.
% Justvoiced consist of only voiced part of the signal with each column
is
% the Nwin frame size of the voiced part.

[X,justvoiced] = simplevuv(s,Nwin,fs);

% collect all the voiced terms into one row
sumjustvoiced = [];
[r,c] = size(justvoiced);
for m = 1:c
    sumjustvoiced = [sumjustvoiced justvoiced(:,m)'];
end

% %collect all unvoiced terms into one row
% sumunvoiced = [];
% [r,c] = size(unv);
% for a = 1:c
%     sumunvoiced = [sumunvoiced unv(:,a)'];
% end
%
% %collect all silence terms into one row
% sumsilent = [];
% [r,c] = size(sil);
% for l = 1:c
%     sumsilent = [sumsilent sil(:,l)'];
% end

% reads in MFCC to give coef for each frames (all and voiced only)

% mfcc for only voiced collected signals per frame
% mfcc using Malcolm Slaney
% cepsvoiced = mfcc(sumjustvoiced,fs,Nwin); % using fftsize = 2048
```

## simplevuv.m

```
% Code written by Mitch Wilkes, modified by Nik Nur Wahidah and Wan
Ahmad Hasan, Spring 2011

% function [X,justvoiced,unv,sil] = simplevuv(s,Nwin,fs)
function [X,justvoiced] = simplevuv(s,Nwin,fs)

% Set the frame length
% Nwin = 200;

% Compute the number of non-overlapping windows
Nlen = length(s);
Nwins = floor(Nlen/Nwin);

% Force the signal, x, to have exactly Nwins frames

x = s(1:(Nwins*Nwin));
Nlen = length(x);

%This is main part of the voiced/unvoiced/silence detection

[B1,A1] = butter(3, [2500 5000]/(fs/2));
[B2,A2] = butter(3, [720 2340]/(fs/2));
[B3,A3] = butter(3, [320 1080]/(fs/2));
[B4,A4] = butter(3, [160 540]/(fs/2));
[B5,A5] = butter(3, [80 260]/(fs/2));

% Put the signal, x, into a matrix, X, where each column is a
% frame. The frames are not overlapping.

X = reshape(x, Nwin, Nwins);

% For each frame, compute the energy in each of the frequency bands.
% The result is a vector of energies for each frequency band.
% These vectors are row vectors.

E1 = zeros(1,Nwins);
E2 = zeros(1,Nwins);
E3 = zeros(1,Nwins);
E4 = zeros(1,Nwins);
E5 = zeros(1,Nwins);

for i=1:Nwins
    E1(i) = sum(filter(B1,A1, X(:,i)).^2);
    E2(i) = sum(filter(B2,A2, X(:,i)).^2);
    E3(i) = sum(filter(B3,A3, X(:,i)).^2);
    E4(i) = sum(filter(B4,A4, X(:,i)).^2);
    E5(i) = sum(filter(B5,A5, X(:,i)).^2);
end

% Combine the energy band vectors into a matrix where each row
```

```

% is an energy band vector

E = [E1 ; E2 ; E3 ; E4 ; E5];

% Results of the analysis are the vectors that indicate which frames
are voiced, unvoiced, and silence.
% These are the vectors computed below: unvoiced, voiced and silent

unvoiced = max(E) == E1; % unvoiced(i) = 1 means ith frame is unvoiced
thresh = median(E3);
voiced = (E3 >= thresh) & (1 - unvoiced); % voiced(i) = 1 means ith
frame is voiced
silent = (E3 < thresh) & (1 - unvoiced); % silent(i) = 1 means ith
frame is silence (background noise only)

%This is the end of the main part. The rest is for plotting results.

nnn = 0:(Nlen -1);
mmm = (0:(Nwins-1))*Nwin;
maxscale = max(abs(x));
figure(1),
plot(nnn,x,mmm,silent*maxscale,mmm,voiced*maxscale,mmm,unvoiced*maxscal
e)

% -----
% collecting the voiced part
justvoiced = zeros(Nwin,1);
k = 1;
for j = 1:Nwins
    if voiced(j) == 1
        justvoiced(:,k) = X(:,j);
        k = k+1;
    end
end

%collecting unvoiced part for testing
% m = 1;
% for n = 1:Nwins
%     if unvoiced(n) == 1;
%         unv(:,n) = X(:,n);
%         m = m+1;
%     end
% end

% %collecting silence part for testing
% u = 1;
% for t = 1:Nwins
%     if silent(t) == 1;
%         sil(:,t) = X(:,t);
%         u = u+1;
%     end
% end

```

## split20sec.m

```
% Code obtain from:
% http://www.mathworks.com/matlabcentral/newsreader/view_thread/292920
% modified by Nik Nur Wahidah and Wan Ahmad Hasan, Spring 2011

function split20sec

files = dir('*.wav');

for i = 1:length(files)
    [path, name, ext] = fileparts(files(i).name);
    [fileName b] = strread(name, '%s %s', 'delimiter', '.');
    fileName = char(fileName);
    fileNamewav = char(strcat(fileName, '.wav'));

    format long

    %fileName='112105nt2_readingVUV.wav';
    [y, Fs, nbits] = wavread(fileNamewav);
    [size_r, size_c]=size(y);
    j=[];
    k=0;
    wavefilesplit=[];

    for i=1:20*Fs:size_r, %build array of the desired range
        j(end+1,:)=i;
    end;

    j(end+1,:)=size_r; % adds the end of the sound file to the end of
the j
        % array

    [size_rj, size_cj]=size(j); %used to get size of j array

    for i=1:1:size_rj-1, k=k+1;
        wavefilesplit=y(j(k):j(k+1),:); % get range from j array
        wavefn=strcat(fileName, num2str(k)); % create file name
        wavwrite([wavefilesplit], Fs, 32, strcat('D:\niknwan\3vuv 20 sec
segments only\male interview\', wavefn));
    end;
end
```

## mean\_energy.m

```
% Code written by Nik Nur Wahidah and Wan Ahmad Hasan, Spring 2011
function mean_energy

%mean_energy = [];
files = dir('*.wav');

for i = 1:length(files)
    [s,fs] = wavread(files(i).name);

    [mean_energy] = pdgm_sum(s,fs);
    [path, name, ext] = fileparts(files(i).name);
    filename = fullfile(path, [name []]);
    filename = strcat('D:\NIKWAN\4psd ratio\male 8bands
interview\HR\',filename);
    save(filename, 'mean_energy');
end
```

## pdgm\_sum.m

```
% Code written by Nik Nur Wahidah and Wan Ahmad Hasan, Spring 2011
function [mean_energy] = pdgm_sum(s,fs)
% [s,fs] = wavread('011706ntl_readingVUV1.wav');

Twin = 0.040; %window size
Nwin = round(Twin*fs);

% Compute the number of nonoverlapping windows
Nlen = length(s);
Nwins = floor(Nlen/Nwin);

% Force the signal, x, to have exactly Nwins frames
x = s(1:(Nwins*Nwin));
Nlen = length(x);

% Each column w of X is a non overlapping frames of size Nwin.
X = reshape(x, Nwin, Nwins);    %1764x500

[Xr,Xc] = size(X);
k = 0;
Pxx = []; psd = [];
w = [];
f = []; freq = [];

for i = 1:Xc, k = k+1;
    % Pwelch
    % [Pxx,w] = pwelch (X (:,i),Nwin,0,fs);
    % psd = [psd Pxx];

    % periodogram
    Xmag = (abs(fft(X(:,i),fs)).^2)/Nwin;
    psd = [psd Xmag];
end

% 4 BANDS -----
% %assigning variables for 4 bands and full range
% total_band1=[];
% total_band2=[];
% total_band3=[];
% total_band4=[];
% total_area=[];
% total_ratio1=[];
% total_ratio2=[];
% total_ratio3=[];
% total_ratio4=[];
%
% fr1 = 2:501;
% fr2 = 502:1001;
% fr3 = 1002:1501;
% fr4 = 1502:2001;
% ftotal = 2:2001;
%
% bandareal = [];
```

```

% bandarea2 = [];
% bandarea3 = [];
% bandarea4 = [];
% allarea = [];

% 8 BANDS-----
total_band1 = [];
total_band2 = [];
total_band3 = [];
total_band4 = [];
total_band5 = [];
total_band6 = [];
total_band7 = [];
total_band8 = [];

total_area = [];

total_ratio1 = [];
total_ratio2 = [];
total_ratio3 = [];
total_ratio4 = [];
total_ratio5 = [];
total_ratio6 = [];
total_ratio7 = [];
total_ratio8 = [];

fr1 = 2:251;
fr2 = 252:501;
fr3 = 502:751;
fr4 = 752:1001;
fr5 = 1002:1251;
fr6 = 1252:1501;
fr7 = 1502:1751;
fr8 = 1752:2001;
ftotal = 2:2001;

bandarea1 = [];
bandarea2 = [];
bandarea3 = [];
bandarea4 = [];
bandarea5 = [];
bandarea6 = [];
bandarea7 = [];
bandarea8 = [];
allarea = [];

for j=1:Xc;
    psdtotal = psd(2:2001,j); %in index 48 for freq,the value is 2000Hz

% % 4 BANDS -----
%     psdr1=psd(2:501,j);
%     psdr2=psd(502:1001,j);
%     psdr3=psd(1002:1501,j);
%     psdr4=psd(1502:2001,j);
%

```

```

%      %Area calculation
%      totalarea=trapz(fttotal,psdttotal);
%      band1area=trapz(fr1,psdr1);
%      band2area=trapz(fr2,psdr2);
%      band3area=trapz(fr3,psdr3);
%      band4area=trapz(fr4,psdr4);
%
%      bandarea1 = [bandarea1 band1area];
%      bandarea2 = [bandarea2 band2area];
%      bandarea3 = [bandarea3 band3area];
%      bandarea4 = [bandarea4 band4area];
%      allarea = [allarea totalarea];

% % 8 BANDS -----
psdr1 = psd(2:251,j);
psdr2 = psd(252:501,j);
psdr3 = psd(502:751,j);
psdr4 = psd(752:1001,j);
psdr5 = psd(1002:1251,j);
psdr6 = psd(1252:1501,j);
psdr7 = psd(1502:1751,j);
psdr8 = psd(1752:2001,j);

%Area calculation
totalarea = trapz(fttotal,psdttotal);
area1 = trapz(fr1,psdr1);
area2 = trapz(fr2,psdr2);
area3 = trapz(fr3,psdr3);
area4 = trapz(fr4,psdr4);
area5 = trapz(fr5,psdr5);
area6 = trapz(fr6,psdr6);
area7 = trapz(fr7,psdr7);
area8 = trapz(fr8,psdr8);

bandarea1 = [bandarea1 area1];
bandarea2 = [bandarea2 area2];
bandarea3 = [bandarea3 area3];
bandarea4 = [bandarea4 area4];
bandarea5 = [bandarea5 area5];
bandarea6 = [bandarea6 area6];
bandarea7 = [bandarea7 area7];
bandarea8 = [bandarea8 area8];

allarea = [allarea totalarea];
end

ratio1 = sum(bandarea1)/sum(allarea);
ratio2 = sum(bandarea2)/sum(allarea);
ratio3 = sum(bandarea3)/sum(allarea);
ratio4 = sum(bandarea4)/sum(allarea);
ratio5 = sum(bandarea5)/sum(allarea);
ratio6 = sum(bandarea6)/sum(allarea);
ratio7 = sum(bandarea7)/sum(allarea);
ratio8 = sum(bandarea8)/sum(allarea);

mean_energy = [ratio1 ratio2 ratio3 ratio4 ratio5 ratio6 ratio7];

```

## ratio\_collect.m

```
% Code written by Nik Nur Wahidah and Wan Ahmad Hasan, Spring 2011
function ratio_collect

files = dir('*.mat');
ratiolist = [];
names = [];

load(files(1).name);
[path, name, ext] = fileparts(files(1).name);
[a b] = streadd(name, '%s %s', 'delimiter', '_');
names = [names;a];
ratiolist = [ratiolist; mean_energy];

for i = 2:(length(files)-1)
    load(files(i).name);
    [path, name, ext] = fileparts(files(i).name);
    [a b] = streadd(name, '%s %s', 'delimiter', '_');
    names = [names;a];
    c = char(names(i-1));
    a = char(a);

    if strcmp(c,a) == 1
        ratiolist = [ratiolist; mean_energy];
    else
        ratiolist = ratiolist;
        filename = strcat('D:\NIKWAN\5psd combine\male 8bands
reading\', 'h', c);
        save(filename, 'ratiolist');
        clear ratiolist;
        ratiolist = [];
        ratiolist = [ratiolist; mean_energy];
    end
end

load(files(length(files)).name);
[path, name, ext] = fileparts(files(length(files)).name);
[a b] = streadd(name, '%s %s', 'delimiter', '_');
c = char(names(length(files)-1));
ratiolist = [ratiolist; mean_energy];
ratiolist = ratiolist;
filename = strcat('D:\NIKWAN\5psd combine\male 8bands reading\', 'h', c);
save(filename, 'ratiolist');
```

### **getalldepresseddata.m**

```
% Code written by Mitch Wilkes, modified by Nik Nur Wahidah and Wan  
Ahmad Hasan, Spring 2011
```

```
dfiles = dir('d*.mat');  
Nd = length(dfiles);  
Dmean_energy = [];  
for i=1:Nd  
    load( dfiles(i).name );  
    Dmean_energy = [Dmean_energy ; ratiolist];  
end
```

### **getallhighriskdata.m**

```
% Code written by Mitch Wilkes, modified by Nik Nur Wahidah and Wan  
Ahmad Hasan, Spring 2011
```

```
hfiles = dir('h*.mat');  
Nh = length(hfiles);  
Hmean_energy = [];  
for i=1:Nh  
    load( hfiles(i).name );  
    Hmean_energy = [Hmean_energy ; ratiolist];  
end
```

### **getallremitteddata.m**

```
% Code written by Mitch Wilkes, modified by Nik Nur Wahidah and Wan  
Ahmad Hasan, Spring 2011
```

```
rfiles = dir('r*.mat');  
Nr = length(rfiles);  
Rmean_energy = [];  
for i=1:Nr  
    load( rfiles(i).name );  
    Rmean_energy = [Rmean_energy ; ratiolist];  
end
```

## equal\_test\_train.m

```
% Code written by Nik Nur Wahidah and Wan Ahmad Hasan, Spring 2011

getalldepresseddata
getallhighriskdata
getallremitteddata

% % one
% data = [Hmean_energy(:,7); Dmean_energy(:,7)];

% two
a=2 ; b=3;
data = [[Hmean_energy(:,a) Hmean_energy(:,b)]; [Dmean_energy(:,a)...
Dmean_energy(:,b)]];

% % three
% a=1 ; b=5; c=6;
% data = [Hmean_energy(:,a) Hmean_energy(:,b) Hmean_energy(:,c);
% Dmean_energy(:,a) Dmean_energy(:,b) Dmean_energy(:,c)];

lab = [ones(102,1); zeros(95,1)];

[idx1,err1,P1,logp1,coeff1] = classify(data,data,lab,'linear');
all = sum(idx1==lab)/length(lab)*100
hr = sum(idx1(1:102)==lab(1:102))/102*100
dep = sum(idx1(103:197)==lab(103:197))/95*100

[idxq,errq,Pq,logpq,coeffq] = classify(data,data,lab,'quadratic');
all = sum(idxq==lab)/length(lab)*100
hr = sum(idxq(1:102)==lab(1:102))/102*100
dep = sum(idxq(103:197)==lab(103:197))/95*100
% figure,plot(data(1:194,3), data(1:194,4),'bo',data(195:271,3),
% data(195:271,4),'ro');

% for plotting quadratic and linear discriminant boundaries
figure,plot(data(1:102,1),data(1:102,2),'ro',data(103:197,1),data(103:1
97,2),... 'bx','Markersize',8,'Linewidth',2)
hold on

% ----- plotting linear boundary -----
K = coeff1(1,2).const;
L = coeff1(1,2).linear;
% Q = coeffq(1,2).quadratic;
% Q = coeff1(1,2).quadratic;
% Function to compute K + L*v + v'*Q*v for multiple vectors
% v=[x;y]. Accepts x and y as scalars or column vectors.
f = @(x,y) K + [x y]*L;

% K + [x y]*L + sum(([x y]*Q) .* [x y], 2);
% f = @(x,y) K + [x y]*L + sum(([x y]*Q) .* [x y], 2);

h2 = ezplot(f,[0.1 0.8 0 0.5]);
set(h2,'Color','m','LineWidth',2)
hold off
```

```

% plotting quadratic boundary
K = coeffq(1,2).const;
L = coeffq(1,2).linear;
Q = coeffq(1,2).quadratic;
% Function to compute  $K + L*v + v'*Q*v$  for multiple vectors
%  $v=[x;y]$ . Accepts  $x$  and  $y$  as scalars or column vectors.
%  $K + [x y]*L + \text{sum}([x y]*Q .* [x y], 2)$ ;

f = @(x,y) K + [x y]*L + sum([x y]*Q .* [x y], 2);

h2 = ezplot(f,[0.1 0.8 0 0.5]);
set(h2, 'Color', 'm', 'LineWidth', 2)
hold off

```

## jackknife.m (example for HR/DEP)

```
% Code written by Mitch Wilkes, modified by Nik Nur Wahidah and Wan
Ahmad Hasan, Spring 2011

%% jackknife HR-DEP
clear;clc
getalldepresseddata
getallhighriskdata

% % one
% myData = [Hmean_energy(:,7); Dmean_energy(:,7)];

% two
a=1 ; b=3;
myData = [[Hmean_energy(:,a) Hmean_energy(:,b)]; [Dmean_energy(:,a)
Dmean_energy(:,b)]];

% % three
% a=4 ; b=6; c=7;
% myData = [Hmean_energy(:,a) Hmean_energy(:,b) Hmean_energy(:,c); %
Dmean_energy(:,a) Dmean_energy(:,b) Dmean_energy(:,c)];

labels = [ones(123,1); zeros(95,1)];

%linear
htotal=[];dtotal=[];

idxm = classify( myData(1:9,:), myData( 10:218,:), labels(10:218) );
h=sum(idxm==1);htotal=[htotal h];
idxm = classify( myData(10:18,:), [myData( 1:9,);
myData(19:218,)],... [labels(1:9) ; labels(19:218)]
);h=sum(idxm==1);htotal=[htotal h];
idxm = classify( myData(19:25,:), [myData( 1:18,); myData(26:218,)],
... [labels(1:18) ; labels(26:218)] );h=sum(idxm==1);htotal=[htotal h];
idxm = classify( myData(26:32,:), [myData( 1:25,); myData(33:218,)],
... [labels(1:25) ; labels(33:218)] );h=sum(idxm==1);htotal=[htotal h];
idxm = classify( myData(33:51,:), [myData( 1:32,); myData(52:218,)],
... [labels(1:32) ; labels(52:218)] );h=sum(idxm==1);htotal=[htotal h];
idxm = classify( myData(52:69,:), [myData( 1:51,); myData(70:218,)],
... [labels(1:51) ; labels(70:218)] );h=sum(idxm==1);htotal=[htotal h];
idxm = classify( myData(70:84,:), [myData( 1:69,); myData(85:218,)],
... [labels(1:69) ; labels(85:218)] );h=sum(idxm==1);htotal=[htotal h];
idxm = classify( myData(85:108,:), [myData( 1:84,);
myData(109:218,)], ... [labels(1:84) ; labels(109:218)]
);h=sum(idxm==1);htotal=[htotal h];
idxm = classify( myData(109,:), [myData( 1:108,); myData(110:218,)],
... [labels(1:108) ; labels(110:218)] );h=sum(idxm==1);htotal=[htotal
h];
idxm = classify( myData(110:123,:), [myData( 1:109,);
myData(124:218,)], ... [labels(1:109) ; labels(124:218)]
);h=sum(idxm==1);htotal=[htotal h];
```

```

idxm = classify( myData(124:132,:), [myData( 1:123,:) ;
myData(133:218, :)], ... [labels(1:123) ; labels(133:218)]
);d=sum(idxm==0);dtotal=[dtotal d];
idxm = classify( myData(133:146,:), [myData( 1:132,:) ;
myData(147:218, :)], ... [labels(1:132) ; labels(147:218)]
);d=sum(idxm==0);dtotal=[dtotal d];
idxm = classify( myData(147:154,:), [myData( 1:146,:) ;
myData(155:218, :)], ... [labels(1:146) ; labels(155:218)]
);d=sum(idxm==0);dtotal=[dtotal d];
idxm = classify( myData(155:163,:), [myData( 1:154,:) ;
myData(164:218, :)], ... [labels(1:154) ; labels(164:218)]
);d=sum(idxm==0);dtotal=[dtotal d];
idxm = classify( myData(164:175,:), [myData( 1:163,:) ;
myData(176:218, :)], ... [labels(1:163) ; labels(176:218)]
);d=sum(idxm==0);dtotal=[dtotal d];
idxm = classify( myData(176:178,:), [myData( 1:175,:) ;
myData(179:218, :)], ... [labels(1:175) ; labels(179:218)]
);d=sum(idxm==0);dtotal=[dtotal d];
idxm = classify( myData(179:183,:), [myData( 1:178,:) ;
myData(184:218, :)], ... [labels(1:178) ; labels(184:218)]
);d=sum(idxm==0);dtotal=[dtotal d];
idxm = classify( myData(184:196,:), [myData( 1:183,:) ;
myData(197:218, :)], ... [labels(1:183) ; labels(197:218)]
);d=sum(idxm==0);dtotal=[dtotal d];
idxm = classify( myData(197:206,:), [myData( 1:196,:) ;
myData(207:218, :)], ... [labels(1:196) ; labels(207:218)]
);d=sum(idxm==0);dtotal=[dtotal d];
idxm = classify( myData(207,:), [myData( 1:206,:) ; myData(208:218, :)],
... [labels(1:206) ; labels(208:218)] );d=sum(idxm==0);dtotal=[dtotal
d];
idxm = classify( myData(208:218,:), myData( 1:207,:), labels(1:207) );
d=sum(idxm==0);dtotal=[dtotal d];

hall = sum(htotal);
dall = sum(dttotal);
all = hall + dall;

per_all = all/length(labels)*100
per_hr = hall/123*100
per_dep = dall/95*100

% quadratic
htotal=[];dttotal=[];

idxm = classify( myData(1:9,:), myData( 10:218,:),
labels(10:218), 'quadratic' ...);
h=sum(idxm==1);htotal=[htotal h];
idxm = classify( myData(10:18,:), [myData( 1:9,:) ; myData(19:218, :)],
... [labels(1:9) ; labels(19:218)], 'quadratic' );
h=sum(idxm==1);htotal=[htotal h];
idxm = classify( myData(19:25,:), [myData( 1:18,:) ; myData(26:218, :)],
... [labels(1:18) ; labels(26:218)], 'quadratic' );
h=sum(idxm==1);htotal=[htotal h];
idxm = classify( myData(26:32,:), [myData( 1:25,:) ; myData(33:218, :)],
... [labels(1:25) ; labels(33:218)], 'quadratic' );

```

```

h=sum(idxm==1);htotal=[htotal h];
idxm = classify( myData(33:51,:), [myData( 1:32,:) ; myData(52:218,:)],
... [labels(1:32) ; labels(52:218)], 'quadratic' );
h=sum(idxm==1);htotal=[htotal h];
idxm = classify( myData(52:69,:), [myData( 1:51,:) ; myData(70:218,:)],
... [labels(1:51) ; labels(70:218)], 'quadratic' );
h=sum(idxm==1);htotal=[htotal h];
idxm = classify( myData(70:84,:), [myData( 1:69,:) ; myData(85:218,:)],
... [labels(1:69) ; labels(85:218)], 'quadratic' );
h=sum(idxm==1);htotal=[htotal h];
idxm = classify( myData(85:108,:), [myData( 1:84,:) ;
myData(109:218,:)], ... [labels(1:84) ; labels(109:218)], 'quadratic' );
h=sum(idxm==1);htotal=[htotal h];
idxm = classify( myData(109,:), [myData( 1:108,:) ; myData(110:218,:)],
... [labels(1:108) ; labels(110:218)], 'quadratic' );
h=sum(idxm==1);htotal=[htotal h];
idxm = classify( myData(110:123,:), [myData( 1:109,:) ;
myData(124:218,:)], [labels(1:109) ; labels(124:218)], 'quadratic' );
h=sum(idxm==1);htotal=[htotal h];
idxm = classify( myData(124:132,:), [myData( 1:123,:) ;
myData(133:218,:)], ... [labels(1:123) ; labels(133:218)], 'quadratic'
);
d=sum(idxm==0);dtotal=[dtotal d];
idxm = classify( myData(133:146,:), [myData( 1:132,:) ;
myData(147:218,:)], ... [labels(1:132) ; labels(147:218)], 'quadratic'
);
d=sum(idxm==0);dtotal=[dtotal d];
idxm = classify( myData(147:154,:), [myData( 1:146,:) ;
myData(155:218,:)], ... [labels(1:146) ; labels(155:218)], 'quadratic'
);
d=sum(idxm==0);dtotal=[dtotal d];
idxm = classify( myData(155:163,:), [myData( 1:154,:) ;
myData(164:218,:)], ... [labels(1:154) ; labels(164:218)], 'quadratic'
);
d=sum(idxm==0);dtotal=[dtotal d];
idxm = classify( myData(164:175,:), [myData( 1:163,:) ;
myData(176:218,:)], ... [labels(1:163) ; labels(176:218)], 'quadratic'
);
d=sum(idxm==0);dtotal=[dtotal d];
idxm = classify( myData(176:178,:), [myData( 1:175,:) ;
myData(179:218,:)], ... [labels(1:175) ; labels(179:218)], 'quadratic'
);
d=sum(idxm==0);dtotal=[dtotal d];
idxm = classify( myData(179:183,:), [myData( 1:178,:) ;
myData(184:218,:)], ... [labels(1:178) ; labels(184:218)], 'quadratic'
);
d=sum(idxm==0);dtotal=[dtotal d];
idxm = classify( myData(184:196,:), [myData( 1:183,:) ;
myData(197:218,:)], ... [labels(1:183) ; labels(197:218)], 'quadratic'
);
d=sum(idxm==0);dtotal=[dtotal d];
idxm = classify( myData(197:206,:), [myData( 1:196,:) ;
myData(207:218,:)], ... [labels(1:196) ; labels(207:218)], 'quadratic'
);
d=sum(idxm==0);dtotal=[dtotal d];
idxm = classify( myData(207,:), [myData( 1:206,:) ; myData(208:218,:)],
... [labels(1:206) ; labels(208:218)], 'quadratic' );

```

```
d=sum(idxm==0);dtotal=[dtotal d];
idxm = classify( myData(208:218,:), myData( 1:207,:),
...labels(1:207), 'quadratic' );
d=sum(idxm==0);dtotal=[dtotal d];
```

```
hallq = sum(htotal);
dallq = sum(dttotal);
allq = hallq + dallq;
```

```
per_all = allq/length(labels)*100
per_hr = hallq/123*100
per_dep = dallq/95*100
```

## distance.m

```
% Code written by Nik Nur Wahidah and Wan Ahmad Hasan, Spring 2011

%% Individual patient distance calculation using jackknife classifier
with hyperplane boundary (3D)
clear;clc;close all;
getalldepresseddata
getallhighriskdata
X = [Hmean_energy; Dmean_energy];
labels = [ones(77,1); zeros(194,1)];

[class,err,POSTERIOR,logp,coeff] =
classify(X(23:28,:),[X(1:22,:);X(29:271,)],[labels(1:22);
labels(29:271)]);

A = coeff(1,2).linear(1);
B = coeff(1,2).linear(2);
C = coeff(1,2).linear(3);
D = coeff(1,2).const;

% -----
----
% plot all (plane and dep vs hr)
[xx, yy] = meshgrid(0.5:0.1:1,0:0.1:.5);
zz = (-A * xx - B * yy - D)/C;
surf(xx,yy,zz)
hold on

plot3(X(1:77,1), X(1:77,2), X(1:77,3), 'ro',X(78:271,1), X(78:271,2),
X(78:271,3), 'bo');
xlabel('PSD 1');
ylabel('PSD 2');
zlabel('PSD 3');
title('Male interview dep vs highrisk');
hold off
grid on

% -----
----
% plotting of one patient data with plane (using each patient coeff
hyperplane)

[xx, yy] = meshgrid(0.5:0.1:1,0:0.1:.5);
zz = (-A * xx - B * yy - D)/C;
surf(xx,yy,zz)
hold on
plot3(X(1:5,1),X(1:5,2),X(1:5,3), 'ro'); %ratiolist
grid on
hold off
xlabel('PSD 1');
ylabel('PSD 2');
zlabel('PSD 3');
title('Male interview: patient h020806nt1');
```

```

% -----
% -----
% calculating distance using equation  $(A*xo + B*yo + C*zo + D)/\sqrt{A^2 + B^2 + C^2}$ 
% -----

dist = [];
[r,c] = size(ratiolist);
for i = 1:r
    xo = ratiolist(i,1);
    yo = ratiolist(i,2);
    zo = ratiolist(i,3);

    pointdist =  $-(A*xo + B*yo + C*zo + D)/\sqrt{A^2 + B^2 + C^2}$ ;

    dist = [dist pointdist];
end
dist = dist';

% -----
% -----
% plotting distance to boundry vs. time (20 sec segments)
t = 1:r;
figure,stem(t,dist)
xlabel('time (20 second segments)')
ylabel('distance to boundary of classifier')
title('Male interview (d101805ab1)')

```

## getlabel.m

```
% Code written by Nik Nur Wahidah and Wan Ahmad Hasan, Spring 2011
```

```
clear;clc;
%total = 22;
%test = 6;

Hvec = dir('h*.mat');

for i = 1:length(Hvec)
    load(Hvec(i).name);
    [r,c] = size(ratiolist);
    ratiolist = [ratiolist ones(r,1)];
    filename = strcat('D:\NIKWAN\6crossval\male 8bands
reading\',Hvec(i).name);
    save(filename,'ratiolist');
end

Dvec = dir('d*.mat');

for i = 1:length(Dvec)
    load(Dvec(i).name);
    [r,c] = size(ratiolist);
    ratiolist = [ratiolist zeros(r,1)];
    filename = strcat('D:\NIKWAN\6crossval\male 8bands
reading\',Dvec(i).name);
    save(filename,'ratiolist');
end

Rvec = dir('r*.mat');

for i = 1:length(Rvec)
    load(Rvec(i).name);
    [r,c] = size(ratiolist);
    ratiolist = [ratiolist zeros(r,1)];
    filename = strcat('D:\NIKWAN\6crossval\male 8bands
reading\',Rvec(i).name);
    save(filename,'ratiolist');
end
```

## crossval.m

```
% Code written by Nik Nur Wahidah and Wan Ahmad Hasan, Spring 2011

function
[percent_all,percent_hr,percent_dep,percent_allq,percent_hrq,percent_de
pq,idxHR,idxDEP,roHR,roDEP,sumHRerr,sumDEPerr] = crossval % for error
histogram evaluation

clear;clc;

testlength = 3;
trainstart = testlength + 1;

Hfiles = dir('h*.mat');

Hlength = length(Hfiles);
Htemp = randn(Hlength,1);
[a,idxH] = sort(Htemp);

Hname = [];
for i = 1:Hlength
    Hname = [Hname; Hfiles(i).name];
end

% choose 3 random HR files for testing
Htest = [];
countHR = zeros(1,Hlength);

for t = 1:testlength
    load(Hname(idxH(t),:));
    Htest = [Htest; ratiolist];

    for m = 1:Hlength
        compare = strcmp(Hfiles(idxH(t)).name,Hfiles(m).name);
        if compare == 1
            countHR(1,m) = countHR(1,m) + 1;
        else
            end
        end
    end

end
Htestlabel = Htest(:,8);
% ----
Htest = Htest(:,1:7);
% ----

Htrain = [];
for t = trainstart:Hlength
    load(Hname(idxH(t),:));
    Htrain = [Htrain; ratiolist];
end
```

```

% % ----
% Htrain = [Htrain(:,1:3) Htrain(:,4)];
% % ----

% -----
----
Dfiles = dir('r*.mat');

Dlength = length(Dfiles);
Dtemp = randn(Dlength,1);
[a,idxD] = sort(Dtemp);

Dname = [];
for i = 1:Dlength
    Dname = [Dname; Dfiles(i).name];
end

% choose 3 random DEP files for testing
Dtest = [];
countDEP = zeros(1,Dlength);

for t = 1:testlength
    load(Dname(idxD(t),:));
    Dtest = [Dtest; ratiolist];

    for n = 1:Dlength
        compare = strcmp(Dfiles(idxD(t)).name,Dfiles(n).name);
        if compare == 1
            countDEP(1,n) = countDEP(1,n) + 1;
        else
            end
        end
    end

end

Dtestlabel = Dtest(:,8);
% ----
Dtest = Dtest(:,1:7);
% ----

Dtrain=[];
for t = trainstart:Dlength
    load(Dname(idxD(t),:));
    Dtrain = [Dtrain; ratiolist];
end
% % ----
% Dtrain = [Dtrain(:,1:3) Dtrain(:,4)];
% % ----
% -----
----
%classify
alltrain = [Htrain ; Dtrain];
alltest = [Htest; Dtest];
testlabel = [Htestlabel; Dtestlabel];

```

```

% band = 2:3;
% class = classify(alltest(:,band), alltrain(:,band), alltrain(:,8));
% classq = classify(alltest(:,band), alltrain(:,band),
alltrain(:,8),'quadratic');

% a=6;
% b=7;
% class = classify([alltest(:,a) alltest(:,b)], [alltrain(:,a)
alltrain(:,b)], alltrain(:,8));
% classq = classify([alltest(:,a) alltest(:,b)], [alltrain(:,a)
alltrain(:,b)], alltrain(:,8),'quadratic');

a=2; b=4; c=7;
class = classify([alltest(:,a) alltest(:,b) alltest(:,c)],
[alltrain(:,a) alltrain(:,b) alltrain(:,c)], alltrain(:,8));
classq = classify([alltest(:,a) alltest(:,b) alltest(:,c)],
[alltrain(:,a) alltrain(:,b) alltrain(:,c)],
alltrain(:,8),'quadratic');

percent_all = sum(class == testlabel)/length(testlabel)*100;
percent_hr = sum(class(1:length(Htestlabel)) ==
testlabel(1:length(Htestlabel)))/length(Htestlabel)*100;
percent_dep = sum(class(length(Htestlabel)+1:length(testlabel)) ==
testlabel(length(Htestlabel)+1:length(testlabel)))/length(Dtestlabel)*1
00;

percent_allq = sum(classq == testlabel)/length(testlabel)*100;
percent_hrq = sum(classq(1:length(Htestlabel)) ==
testlabel(1:length(Htestlabel)))/length(Htestlabel)*100;
percent_depq = sum(classq(length(Htestlabel)+1:length(testlabel)) ==
testlabel(length(Htestlabel)+1:length(testlabel)))/length(Dtestlabel)*1
00;

%-----HR error calculation-----
roHR = [];
counterrHR = [];
for t = 1:3
    load(Hfiles(idxH(t)).name)
    [r c] = size(ratiolist);
    roHR = [roHR; r];
end

for j=1:length(Htestlabel)
    if classq(j) == 0
        counterrHR = [counterrHR; 1];
    elseif classq(j) == 1
        counterrHR = [counterrHR; 0];
    end
end

rH1 = roHR(1); rH2 = roHR(2); rH3 = roHR(3);

```

```

sumHRerr = [sum(counterrHR(1:rH1, 1)); sum(counterrHR(rH1+1:rH1+rH2,
1)); sum(counterrHR(rH1+rH2+1:rH1+rH2+rH3, 1))];

%-----DEP error calculation-----
roDEP = [];
counterrDEP = [];
for t = 1:3
    load(Dfiles(idxD(t)).name)
    [r c] = size(ratiolist);
    roDEP = [roDEP; r];
end

for j=1:length(Dtestlabel)
    if classq(j) == 1
        counterrDEP = [counterrDEP; 1];
    elseif classq(j) == 0
        counterrDEP = [counterrDEP; 0];
    end
end

rD1 = roDEP(1); rD2 = roDEP(2); rD3 = roDEP(3);

sumDEPerr = [sum(counterrDEP(1:rD1, 1)); sum(counterrDEP(rD1+1:rD1+rD2,
1)); sum(counterrDEP(rD1+rD2+1:rD1+rD2+rD3, 1))];

%-----
-----

idxHR = idxH(1:3);
idxDEP = idxD(1:3);

```

## errorhist.m

```
% Code written by Nik Nur Wahidah and Wan Ahmad Hasan, Spring 2011

function [CHR,percentErrorHR,cDEP,percentErrorDEP] = errorhist
clear;clc;

testrun = 100;

patientHR = [];
vectorHR = [];
sumerrorHR = [];

patientDEP = [];
vectorDEP = [];
sumerrorDEP = [];

for i = 1:testrun

[percent_all,percent_hr,percent_dep,percent_allq,percent_hrq,percent_de
pq,idxH,idxD,roHR,roDEP,sumHRerr,sumDEPerr] = crossval;

    patientHR = [patientHR idxH];
    vectorHR = [vectorHR roHR];
    sumerrorHR = [sumerrorHR sumHRerr];

    patientDEP = [patientDEP idxD];
    vectorDEP = [vectorDEP roDEP];
    sumerrorDEP = [sumerrorDEP sumDEPerr];
end

numHR = 10;
numDEP = 11;

allerrorHR = zeros(1,numHR); % number of HR patient
allerrorDEP = zeros(1,numDEP); % number of DEP patient
CHR = zeros(1,numHR); % how many HR times patient show up
cDEP = zeros(1,numDEP); % how many DEP times patient show up

% HR
for k = 1:testrun
    for l = 1:3
        for g = 1:numHR %10 patient for HR
            if patientHR(l,k) == g
                allerrorHR(g) = allerrorHR(g) + sumerrorHR(l,k);
                CHR(g) = CHR(g) + 1;
            end
        end
    end
end
end

% --- ratio/percentage
HRvec = [9 9 7 7 19 18 15 24 1 14];
```

```

allpatientHR = CHR.*HRvec;
percentErrorHR = (allerrorHR./allpatientHR)*100; %if 100%, all wrong

% DEP
for k = 1:testrun
    for l = 1:3
        for g = 1:numDEP %10 patient for DEP
            if patientDEP(l,k) == g
                allerrorDEP(g) = allerrorDEP(g) + sumerrorDEP(l,k);
                cDEP(g) = cDEP(g) + 1;
            end
        end
    end
end
end

% --- ratio/percentage
DEPvec = [9 14 8 9 12 3 5 13 10 1 11];
allpatientDEP = cDEP.*DEPvec;
percentErrorDEP = (allerrorDEP./allpatientDEP)*100; %if 100%, all wrong

```

## percentMean.m

```
% Code written by Nik Nur Wahidah and Wan Ahmad Hasan, Spring 2011

function [mean_all, mean_hr, mean_dep, mean_allq, mean_hrq, mean_depq]
= percentMean
clear;clc;
testrun = 100;

all = [];
hr = [];
dep = [];

allq = [];
hrq = [];
depq = [];
for j = 1:testrun

[percent_all,percent_hr,percent_dep,percent_allq,percent_hrq,percent_de
pq] = crossval;
    all = [all percent_all];
    hr = [hr percent_hr];
    dep = [dep percent_dep];

    allq = [allq percent_allq];
    hrq = [hrq percent_hrq];
    depq = [depq percent_depq];
end

% mean percentage
mean_all = mean(all);
mean_hr = mean(hr);
mean_dep = mean(dep);

mean_allq = mean(allq);
mean_hrq = mean(hrq);
mean_depq = mean(depq);
```

## REFERENCES

- [1] Jiaquan Xu, Kenneth D. K., Betzaida T., “Deaths: Final Data for 2007”, National Vital Statistics Reports, Volume 58, Number 19, May 2010, website, [http://www.cdc.gov/NCHS/data/nvsr/nvsr58/nvsr58\\_19.pdf](http://www.cdc.gov/NCHS/data/nvsr/nvsr58/nvsr58_19.pdf)
- [2] AFSP (American Foundation for Suicide Prevention), January 2011, website, <http://www.afsp.org>
- [3] R. Maris, A. Berman, J. Maltzberger, R Yufit, “Assessment and prediction of suicide”, (pp.183-201), New York, Guilford Press, 1992.
- [4] James R. Rogers, “Suicide Risk Assessment”, Wiley e book, (pp259-263), 2001.
- [5] Man JJ., “The Neurobiology of Suicide”, Nature Medicine 4:25-30, 1998.
- [6] France, D.J., “Acoustical properties of speech as indicators of depression and suicidal risk”, Ph.D, Thesis, Vanderbilt University, August, 1997.
- [7] Ozdas, A., “Analysis of Paralinguistic Properties of Speech for Near-term Suicidal Risk Assessment”, Ph.D, Thesis, Vanderbilt University, May, 2001.
- [8] Yingthawornsuk, T., “Acoustic Analysis of Vocal Output Characteristics for Suicidal Risk Assessment”, Ph.D, Thesis, Vanderbilt University, December, 2007.
- [9] Elaine P.S., Marsal S., Maria Cristina C.F., Gisela O., Mara B., “Vocal Analysis of Suicidal Movie Characters”, Sao Paulo, Brazil, March, 2010.
- [10] Louis A.G., Goldine D.G., “The Measurement of Psychological States through the Content Analysis of Verbal Behavior”, University of California Press
- [11] France D.J., Shiavi R.G., Silverman S., Silverman M., Wilkes D.M., “Acoustical Properties of Speech as Indicators of Depression and Suicidal Risk”, Vol.47, July 2000.
- [12] Ververidis D., Kotropoulos C., “Emotional Speech Recognition: Resources, Features and Methods”, Aristotle University of Thessaloniki, April, 2006.
- [13] Keskinpala H.K., Yingthawornsuk T., Wilkes D.M., Shiavi R.G., Solomon R.M., “Screening for High Risk Suicidal States using Mel-Cepstral Coefficients and Energy in Frequency Bands”, 15<sup>th</sup> European Signal Processing Conference, September, 2007.

- [14] Salisbury D.F., “Researchers measure distinct characteristics in speech of individuals at high risk of suicide”, The online research journal of Vanderbilt University, October, 2000.
- [15] Tolkmitt F., Helfrich H., Standke R., Scherer K.R., “Vocal Indicator of Psychiatric Treatment Effects in Depressive and Schizophrenics”, Journal of communication disorders, Vol.15, pp.209-222, 1982.
- [16] Moore II E., “Evaluating Objective Feature Statistics of Speech as Indicator of Vocal Affect and Depression”, Ph.D Thesis, Georgia Institute of Technology, November, 2003.
- [17] Scherer K.R., “Vocal Communication of Emotion: A Review of Research Paradigms”, Speech Communication 40, pp.227-256, 2003.
- [18] Deng L., O’Shaughnessy D., “Speech Processing: A Dynamix and Optimization-Oriented Approach”, Signal Processing and Communication Series, pp.203-261, 2003.
- [19] Yannakoudakis E. J., Hutton P.J., “Speech Synthesis and Recognition Systems”, pp.16-26, 1987.
- [20] Rabiner L. R., Schafer R.W., “Digital Processing of Speech Signals”, Prentice-Hall Signal Processing Series, pp38-105, 1978.
- [21] Flanagan J.L., “Speech Analysis, Synthesis, and Perception, 2<sup>nd</sup> ed., Springer-Verlag, New York, 1983.
- [22] Ingo R.T., “The Myoelastic Aerodynamic Theory of Phonation “, PhD, pp.26-45, January, 2006.
- [23] Tools for Electrolottographic Analysis: Software, Documentation and Databases, website, <http://voiceresearch.free.fr/egg>
- [24] Digottex G., “Glottal source and vocal tract separation: Estimation of Glottal Parameters, Voice Transformation and Synthesis using a Glottal Model, PhD, University Pierre and Marie Curie, December, 2010.
- [25] Atal B. S. and Hanauer S. L., “Speech Analysis and Synthesis by Linear Prediction of the Speech wave,” J. Acoust. Soc. Am., Vol. 50, No. 2 (Part 2), pp. 637-665, August 1971.
- [26] Fairbanks G., Voice and Articulation Drillbook. Harper & Row, New York, 1960.

- [27] Theodoridis S. and Koutroumbas K., Pattern Recognition 4<sup>th</sup> Edition, Elsevier, pp. 1-84, 2009
- [28] Principal Component Analysis,  
<http://www.scribd.com/doc/36925637/10/Minimally-Adequate-Sample-Size>
- [29] Wikipedia, Human Voice, [http://en.wikipedia.org/wiki/Human\\_Voice](http://en.wikipedia.org/wiki/Human_Voice)
- [30] K. Bob., Mastering Audio, The Art and The Science, Focal Press, pg. 63