ACOUSTIC ANALYSIS OF SPEECH BASED ON POWER SPECTRAL DENSITY FEATURES IN DETECTING SUICIDAL RISK AMONG FEMALE PATIENTS

By

Wan Ahmad Hasan Wan Ahmad Sanadi

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 D. Mitchell Wilkes, Ph.D.

Associate Professor Ronald M. Salomon, M.D.

ACKNOWLEDGEMENTS

First of all, thank you God for giving me the strength and perseverance to finish my work. May all my effort be counted as good deeds worthy of Your reward.

I would like to express my deepest gratitude to my advisor, Dr. Mitch Wilkes for his dedication and patience in guiding me from the beginning until the end of my graduate program. My appreciation also goes to the second reader of my thesis, Dr. Ronald Salomon, who has been really helpful especially throughout the period of data acquisition for my research. I would also like to thank the department of Electrical Engineering for providing me the financial support thus enabling me to focus on my graduate work without worries.

Special thanks to my research partner, Nik Nur Wahidah Nik Hashim, without whom I would not be able to successfully finish my thesis. Your tireless effort and sacrifice would never be forgotten. I also thank my other lab and office mates, who have not just advised me on my graduate work, but also on life in general.

My greatest appreciation goes to both of my parents, Wan Ahmad Sanadi Wan Ali and Nor Hairiah Md Top who have raised me with love and care until I became what I am today. To other members of my family especially my siblings, thank you very much for being my inspiration.

Lastly, to all my friends at Vanderbilt and elsewhere that I could not mention their names one by one here, your support and encouragement were a driving force for me to finish my work. Your presence makes my life meaningful and enjoyable. Every bit of your help in any aspect of my life is deeply appreciated.

TABLE OF CONTENTS

Page

Page

Appendix

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER I

INTRODUCTION

Suicide is a major health problem in the United States. The latest statistics available on the Centers for Disease Control and Prevention (CDC) website have shown that a rate of 11.26 suicides out of every 100,000 population was recorded in 2007 and it was the highest since 1999. From 1991 to 2006, the suicide rate was consistently higher among males while among females, it has been gradually increasing from 2000 to 2006. According to 2007 data, suicide is the 11th leading cause of death in the United States and the number of suicides was twice that of homicides. More than 34,000 suicides were committed which is equivalent of 94 suicides per day [1]. Psychiatric disorders are prevalent among those who commit suicides. Shaffer et al. reported that more than 90% of the individuals who committed suicide were diagnosed with at least one psychiatric disorder [2]. Studies have also reported that depression, a typical psychiatric illness, is the most common antecedent to suicide [3, 4]. This close relationship explains the great amount of work conducted in eradicating depression because the hope is, by treating depression, suicide can be prevented. In essence, depression treatment is in fact suicide prevention.

Preventing suicide has been the subject of extensive study by researchers, and a number of published works emphasize the role of clinicians in the task [5, 6]. The public expects clinicians to predict suicide attempts and to prevent suicide death from happening. However, this task is complicated and there is no guarantee that their

predictions will always be correct. Nevertheless, clinicians are responsible to make their best effort to predict and prevent suicide [7, 8].

A common problem in predicting and preventing suicide is to determine the degree of suicidal risk in an individual patient. This process requires a considerable amount of commitment from the clinicians. They have to gather information about the clinical features of the patient, document this information, and then use it to formulate decisions on the patient's suicidal risk and the plan for treatment. The information gathering process is laborious because clinicians have to maintain regular interaction with the patient to acquire as much and as current information as possible. All the relevant clinical information will be used to evaluate the degree of suicidal risk for that particular patient. Important information includes the history of the patient's conditions, psychological testing records, self-report data, reports by other people, and also the current condition based on a clinical interview with the clinician. However, the decision made on the degree of suicidal risk of patients after all the information is acquired is still typically based on the experience and intuition of the clinician [9, 10, 11].

The procedure to assess the suicidal risk of patients is time-consuming and most of the time, clinicians cannot make drastic decisions in an instant. Important information needed to diagnose a patient may not always be available in urgent situations calling for immediate clinical judgment. Recent advancements have seen that computer-based diagnostic tools are able to provide additional data that can be useful to clinicians in making clinical judgment [9, 10, 11]. However, it is important to note that clinicians cannot use these computer-based diagnostic tools as a standalone instrument to make clinical judgment on patients.

The information about a patient's risk of committing suicide is crucial in deciding whether or not the patient needs to be hospitalized. The risk assessment is also equally important in determining if a patient is safe to be discharged from hospital. Additional tools that can indicate suicidal risk can help to permit hospitalization of a patient whose risk might have been inaccurately diagnosed and to boost assessment accuracy before patient release. As a result, patients' welfare is protected and the hospital image is safeguarded. Ultimately, suicide is more likely to be prevented [9, 11].

One of the cues that can be used to help in making clinical decisions is a patient's voice. The human voice is powerful as it can mirror the speaker's physical, mental and emotional state. Studies have shown that non-content speech acoustics are able to reflect the level of a person's psychological state [12]. Experienced clinicians have been using these vocal cues as symptoms in diagnosing abnormal behaviors or emotional conditions of patients [13]. Evidence has shown that changes in emotional state can also alter the speech production mechanism, namely the respiratory, phonatory and articulatory processes. These changes are then encoded into the acoustic signal that eventually can be heard. In short, vocalization reflects the many different features of the functioning neurophysiological structures of the human body [14].

Realizing the potential of the human voice in the clinical field, Drs. Stephen and Marilyn Silverman proposed the study of vocal properties in investigating suicidal conditions. The vocal properties of three different subject groups were studied. The sample for their study consisted of near term suicidal patients, major depressed patients, and non-depressed control subjects. Vocal features were extracted from the speech recordings and were analyzed to develop a classifier. The quality of a classifier is

determined by its differentiating ability, whether it can clearly separate between different subject groups [15]. Some of the vocal features that have been used in the non-content speech studies for determining suicidal states are mel-cepstral coefficients, power spectral density, formant frequency, vocal jitter, and glottal spectral slope [9, 10, 11, 15].

This thesis attempts to investigate some human vocal features to analyze their ability to distinguish between depressed female patients and high-risk suicidal female patients. The chosen vocal feature for our work is the power spectral density (PSD) of human speech. This paper is a continuation of Yingthawornsuk's work [11] and others [12, 42, 9, 10, 43], where they have already analyzed the effectiveness of vocal features including the PSD in identifying high-risk suicidal patients from depressed patients.

This thesis is organized as follows. Chapter 2 covers the background of the study starting with the physiology of the speech production process, followed by its working model. The background is capped with an explanation of how emotional arousal can affect the physiology of human speech thus enabling researchers to benefit from it.

Chapter 3 elaborates on previous work done related to speech and emotion. Earlier studies investigate the correlation between human speech and emotions, particularly depression. Further research in this area resulted in the idea of identifying near-term high-risk suicidal patients by analyzing their vocal characteristics. The chapter ends with relaying the significance of this work in developing additional diagnostic tools to aid physicians in preventing suicides in order to protect human life.

Chapter 4 explains the methodology implemented in acquiring and analyzing vocal features. The process of acquiring data starts with recording the interview session, followed by the pre-processing of the audio files to prepare the data for classification.

Types of classifications as well as statistical analysis methods being applied are also explained in this chapter.

Results and analyses are presented in the final chapter, which is chapter 5. The percentages of correct classifications based on various statistical sampling methods are tabulated. These recorded outputs are analyzed to measure the effectiveness of the classifications. Ways to enhance classification results and possible future work are also discussed. The overall conclusion drawn from the study is finally stated at the end of this chapter. The paper concludes with a list of references and MATLAB code in the appendix.

CHAPTER II

BACKGROUND

2.1 The Physiology and the Process of Speech Production

The speech waveform is an acoustic sound pressure wave, produced by autonomous movements of abdominal structures that form the human speech production system. Figure 1 illustrates the anatomical structures involved in the process of speech production.

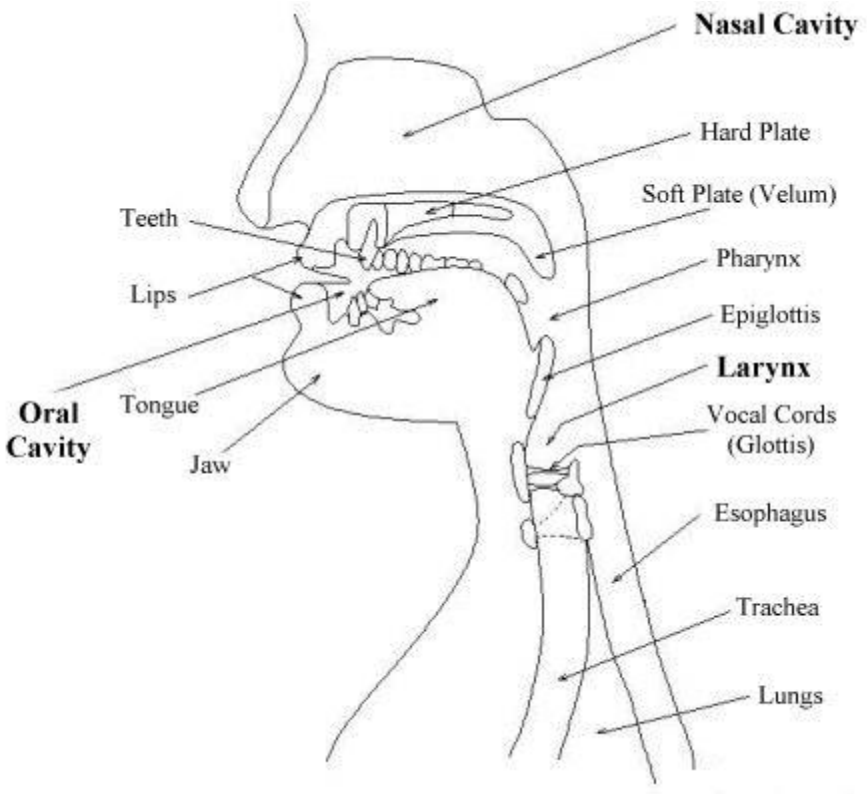


Figure 1: *The anatomical structures involved in the speech production system [16].*

The vocal tract is between the vocal cords (or vocal folds) and the lips. It is composed of the pharynx (throat or pharyngeal cavity, which is the path from the esophagus to the mouth) and oral cavity. The cross-sectional area of the vocal tract varies depending on the positions of the velum, lips, jaw, and tongue. The nasal tract is the path between the velum and the nostrils. The nasal tract and the vocal tract are combined when the velum opens to produce nasal sounds of speech. Inevitably sometimes, in general, the term vocal tract represents these two tracts combined to accommodate nasal sounds and to simplify explanation.

Figure 2 shows a simplified representation of the speech production mechanism. The three main paths that construct the speech production system are the pharyngeal cavity (or pharynx cavity), the nasal cavity, and the oral cavity. The speech production mechanism originates with the air inside the lungs from the normal breathing mechanism. First, the associated muscles apply force, thus pushing the air from the lungs through the trachea and bronchi. In Figure 2, this step is simplified as a piston-cylinder mechanism. The flow of air from the trachea heading up causes the tensed vocal cords within the larynx to vibrate. Because of the vibration, the air flow is split into quasi-periodic pulses for which the frequencies are adjusted when advancing through the pharyngeal cavity, the oral cavity and possibly the nasal cavity. Different sounds are produced depending on the positions of the articulators which include the jaw, velum, lips, tongue, and teeth [18].

The vibration of the vocal cords produces voiced speech sounds such as the vowel sounds. On the other hand, the unvoiced sounds are produced when vocal cords are in a relaxed position and the air pushes its way through a tightened vocal tract causing a turbulent flow. Truncated transient sounds like “-ch” at the end of “peach” or “-ck” at the

end of “pack” have a unique mechanism which starts when pressure increases behind a total closure point anywhere in the vocal tract. Abrupt release of such pressure by opening the closure point produces these sounds [18].

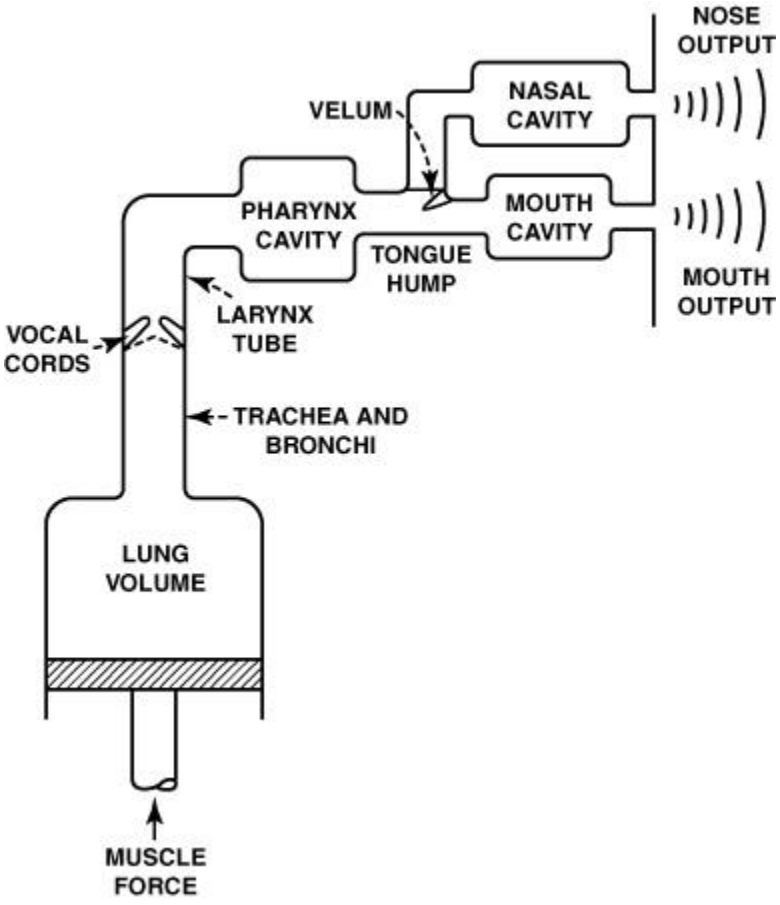


Figure 2: *Simplified representation of the functional components for speech production [17].*

Vibration of the vocal cords produces quasi-periodic air pulses, known as the glottal airflow waveform which is shown in Figure 3. Two commonly accepted theories that explain how the phonation (the vibration of the vocal cords) is initiated are the myoelastic theory and the aerodynamic theory. Van den Berg [20] was noted as the

originator of these theories which he called “the myoelastic-aerodynamic theory” in his paper.

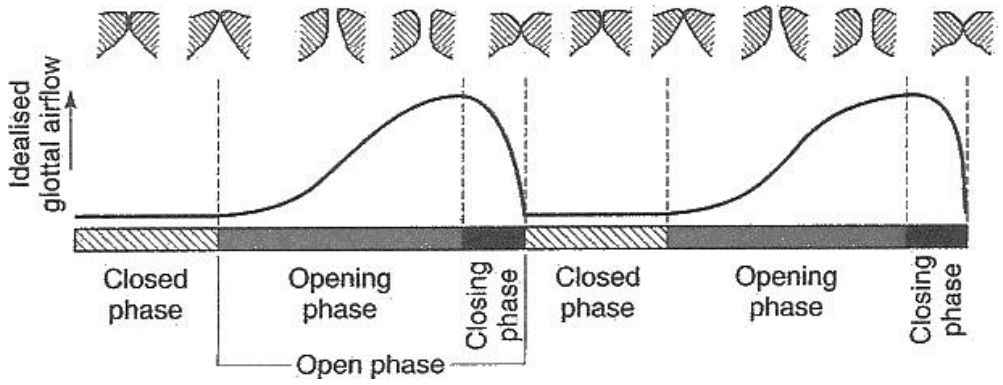


Figure 3: *Glottal flow waveform with the corresponding phases of vocal cords [19].*

The myoelastic theory explains that when the sub-glottal (below glottis – the glottis is the vocal cords together with the space in-between the cords) air accumulates until enough pressure is acquired; it pushes the vocal cords causing them to open. The escape of sub-glottal air reduces the pressure, thus the muscle tension recoil causes the vocal cords to converge back again. Then, the sub-glottal air pressure accumulates again, and the whole cycle repeats. The aerodynamic theory, which is based on the Bernoulli law, states that when the air flows through the glottis and overcomes the muscle of the vocal cords, it creates a push-pull effect that induces an oscillation. The push effect occurs when the vocal cords are opening from a closed position and the pull effect happens when the vocal cords are converging back from the glottal opening position. The airflow is cut off during glottal closure but the sub-glottal air pressure will push the vocal cords apart and the airflow starts up again, thus repeating the cycle. Figure 4 shows the grayscale images of vocal cords in the open and closed positions [21].

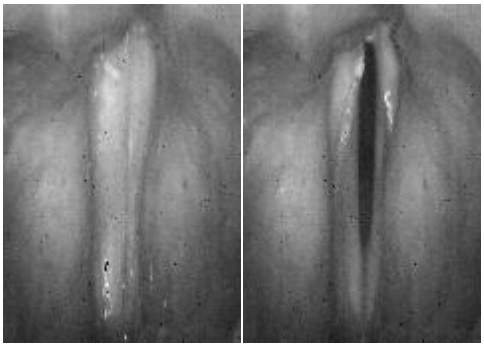


Figure 4: *Vocal cords in closed position (left) and open position (right) [22].*

2.2 The Source-Filter Model of Speech Production

Speech production can be explained with a simple source-filter model. At the most primitive level, the source-filter model can take the form shown in Figure 5.

Figure 5: *A basic speech model [23].*

An example to illustrate this model is when a person blows a saxophone. The air pressure from the mouth is the source and the saxophone itself is a filter. The sounds made are equivalent to speech in the speech model. Figure 6 shows a more refined model of speech production. The voiced speech and unvoiced speech are produced by the impulse train generator and the random noise generator respectively. The switch position points to the normalized excitation source depending on the characteristics of the voice (voice/unvoiced).

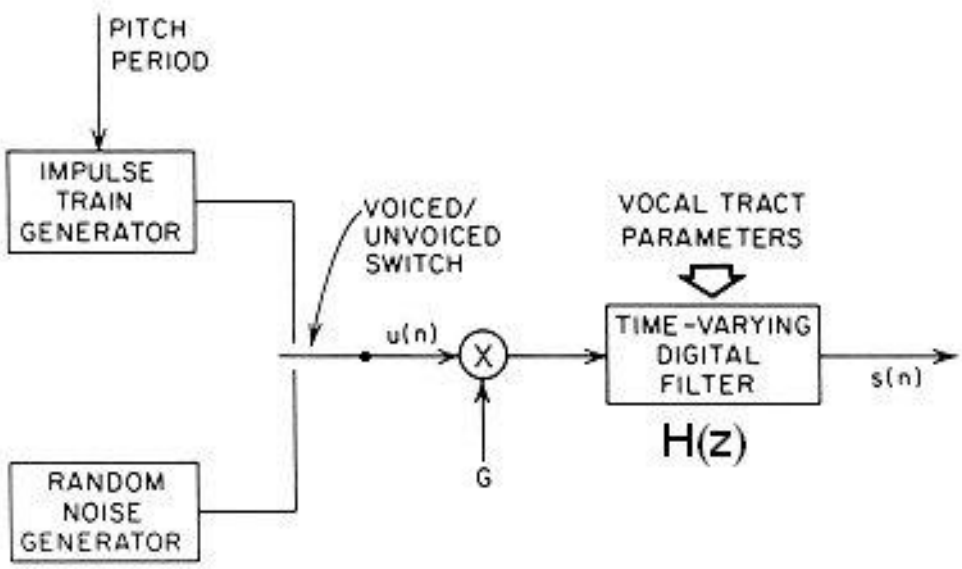


Figure 6: *A more refined source-filter model tied to Linear Predictive Coding (from [18]).*

A gain factor (*G*) is measured from the speech signal and is used to scale the normalized excitation source, *u(n)*. The scaled source then goes through the time-varying digital filter. The filter represents the vocal tract which has varying cross-sectional area throughout the tract. In the source-filter model, this varying size corresponds to the different values of the vocal tract parameters. The output of the digital filter is the speech signal, *s(n)*.

The use of the source-filter model to represent speech production is often linked with the linear predictive coding (LPC) model. LPC is a source-filter analysis-synthesis technique that estimates the generation of sound as an excitation source that goes through an all-pole resonant filter. The details of LPC are explained in many places such as in Rabiner's textbooks [18, 24], Bradbury's paper [25] and in Howitt's Otolith homepage

[26]. Here, the basic idea behind the LPC model is described based on Rabiner's textbook [18].

$$s(n) = \sum_{i=1}^{p} a_i s(n-i) + Gu(n) \qquad (2.1)$$

With Figure 6 as reference, equation (2.1) shows that a speech sample at time *n* is equal to a linear combination of the previous *p* speech samples with an added excitation term, *Gu(n)*, where *u(n)* is a normalized excitation and *G* is the excitation gain. The a_i terms are assumed to be constant coefficients over the speech analysis frame. Speech signals vary with time, so this process is conducted on short chunks of the speech signal called frames. 30 to 50 frames per second is normally used as that is enough to give intelligible speech with good compression [26]. Converting the equation using the *Z*-transform, equation (2.2) is obtained:

$$S(z) = \sum_{i=1}^{p} a_i z^{-i} S(z) + GU(z) \qquad (2.2)$$

Rearranging the terms from the equation above produces a transfer function *H(z)* shown in equation (2.3):

$$H(z) = \frac{S(z)}{GU(z)} = \frac{1}{1 - \sum_{i=1}^{p} a_i z^{-i}} = \frac{1}{A(z)} \qquad (2.3)$$

The transfer function *H(z)* represents the time-varying digital filter shown in Figure 6. This is basically an all-pole, autoregressive (AR) model of speech production, where the vocal tract is represented by non-uniform cylindrical tubes concatenated together as shown in Figure 7. The terms a_i are the filter coefficients that can be calculated using LPC analysis and p is the number of poles. Further in-depth analysis on LPC can be found in [18, 24, 25, 26].

2.3 The Effects of Emotion on the Physiological Structure of Speech Production

The respiratory, phonatory, and articulatory movements involved in speech production are mainly controlled by the respiratory organs, the laryngeal muscles and the various articulators. The neocortex is the part of the brain that mostly controls specific motor commands producing the corresponding muscle movements leading to the desired speech sequence [31]. On the other hand, the effects of emotional arousal that can influence the speech production mechanism, even against the speaker's will, are controlled mainly by the limbic system. Emotional arousal effects the speech production via the activation of the somatic nervous system and the autonomic nervous system. The latter consists of the sympathetic and parasympathetic nervous systems (SNS and PNS) [13, 14, 32]. Changes in the activation of SNS and PNS result in variations in blood pressure, heart rate, muscle tension, respiratory patterns, and motor coordination. All these variations eventually modify the respiratory, phonatory, and articulatory systems in speech production [33]. Figure 7 shows a simplified version of the emotional arousal effect on speech production. Clearly the physiology of speech production can be altered by changes in emotions.

The activation of SNS and PNS increases the possibility of changes happening in speech acoustic characteristics, and these changes can be captured by extracting some speech parameters. Modification in respiratory patterns can cause differences in sub-glottal pressure and this, together with changes in muscle tension can alter the pattern of vocal cord vibrations and the articulation process. Besides that, disturbances in the coordination of muscular activity involved in producing speech can also result in variations which can be reflected by measurable speech parameters [34].

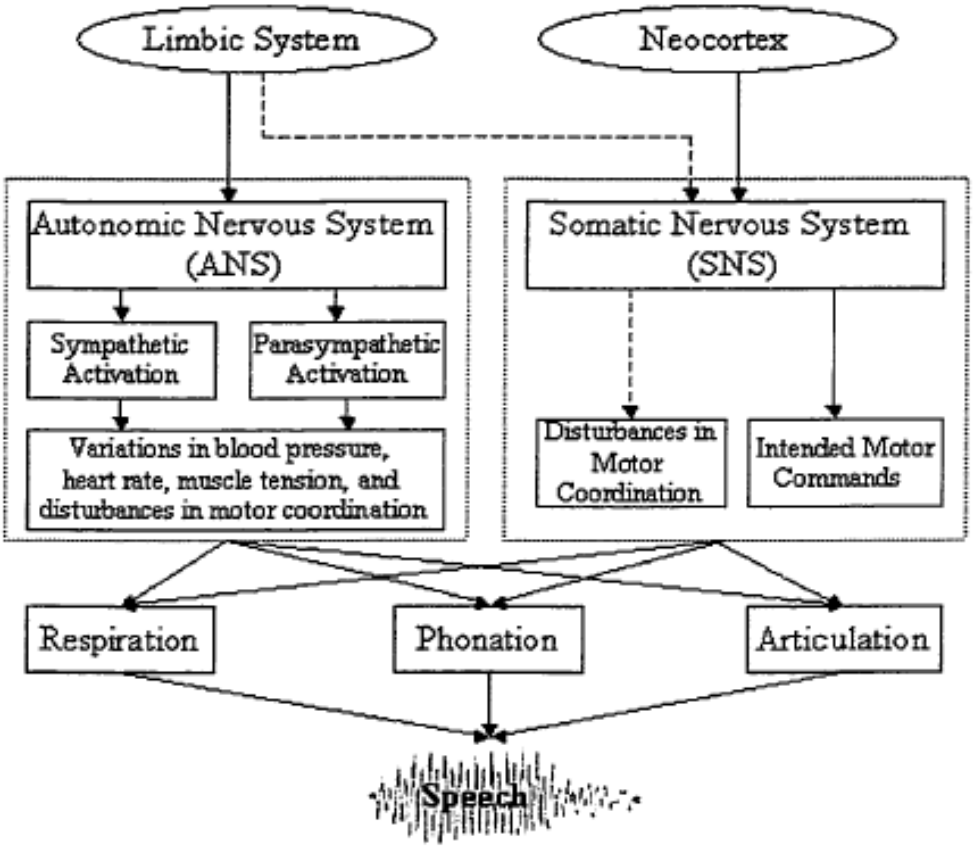


Figure 7: *The emotional arousal effect on speech production [10].*

The respiratory, phonatory, and articulatory systems are handled by neuromuscular control that has to be tuned accordingly to ensure smooth vocal cord vibration and seamless adjustments between articulatory positions. Changes in respiratory muscles, coordination, and laryngeal musculature can alter the shape of glottal flow waveform. Disturbances in coordination and phonatory muscles could also lead to changes in fundamental frequency, irregularities in the successive glottal cycle durations (vocal jitter), and variations in intensity (shimmer). Changes in articulatory musculature such as increased muscle tone would cause tenseness in the structure of the vocal tract (such as vocal tract resonance walls) and articulators which would eventually affect the resulting frequency spectrum of the speech. These, together with increased tension in the laryngeal musculature, were suggested to cause higher energy in the upper harmonics. Lack of coordination in the articulatory structures on the other hand would decrease the precision of articulation thus producing relatively narrower formant ranges. This is due to the inability of articulators to reach their targets smoothly [10, 13, 35].

The validity of this research is supported by the known effects of emotional arousal on speech production physiology. Serious suicidal thoughts represent a major change in a human's mental condition. This change includes a wide range of complex emotions and thus, the suicidal vocal patterns are expected to be different from non-suicidal [10].

CHAPTER III

RELATED WORK AND SIGNIFICANCE

3.1 Correlation between Speech and Depression

The study of speech characteristics in psychiatry has been done by many researchers since the thirties. In 1938, Newman and Mather [36] explored the effect of a number of psychiatric disorders to the speech of patients. There were three groups of patients whose speech were studied: patients having classical depression, patients with states of dissatisfaction, self-pity, and gloom, and patients with manic syndromes. Two types of speech, spontaneous and non-spontaneous, were recorded by doing interview and reading sessions (whenever possible), the same way our study was done. There were many forms of speech characteristics that were studied such as articulatory movements, pitch range, and speech tempo. The result from this study verifies that human speech can be affected by psychiatric disorders.

Hargreaves and Starkweather [37] used power spectrum analysis in their study as parameters to characterize the speech. The mood of eight patients with depression syndrome was tracked by observing the power spectrum. The result suggests that the power spectrum of the speech changes as the mood sways. The study also showed that power spectrum analysis is reliable in evaluating changes in mental and emotion states.

There were a number of studies that also have used the distribution of the energy spectrum as one of the speech characterizing properties [13, 38, 37, 39, 40, 9]. These studies have shown that the overall energy is low in the speech of patients with

depressive state and high for patients whose depression has been treated. However, in terms of the energy level increase across the speech frequencies (after the depression has been treated), studies have shown less consistent results. Some studies [38, 39] reported that greater energy increase occurred in the low frequency bands (less than 500 Hz), while other studies [37, 36] said that it occurred in the higher formants. France et al. [9] observed greater energy in the higher frequency range for depressed patients, compared to healthy patients where the greater energy resides in lower frequency.

Another study conducted by Tolkmitt et al. [39] revealed that the formant frequencies of speech for depressed patients before treatment were closer to neutral formant frequencies (500 Hz, 1500 Hz, and 2500 Hz) that are normally found when the vocal tract is in its usual resting position. This finding shows that before treatment, the patients' speech was made with less articulatory effort, which was mirrored by their first format frequencies being closer to 500 Hz. After treatment, greater articulatory effort is present, which means that the vocal tract shape varies more frequently and significantly during speech. As a result, the formant frequencies reach the expected values. Tolkmitt's study concurred with the suggestion that disturbance in muscular coordination of articulatory structures reduces articulatory precision producing narrower formant frequency ranges. This is due to the inability of the articulators to move and reach the appropriate positions needed to shape crisp vowel sounds [13, 35].

Moore et al. [41] selected prosody, formants, and glottal ratio/spectrum as classifying features in their study to discriminate between depressed and non-depressed subjects. The best separation was obtained with the application of glottal ratio/spectrum and formant bandwidths as discriminating features. The glottal ratio/spectrum analysis

resulted in 97.3% and 97.8% accuracy, while the formant bandwidths analysis produced 98.7% and 98.9%, for male and female respectively in each analysis.

3.2 Correlation between Speech and Suicidal Risk

The study of speech in psychiatry has advanced a step further from the analysis of depression to the analysis of suicide. The first investigation of vocal correlates of suicidal risk was initiated by Drs. Stephen and Marilyn Silverman [12], who have been involved in the treatment of severely depressed patients as well as suicidal patients for more than forty years since the sixties. They obtained the recordings of suicide notes and interviews made shortly before the patients attempted suicide. The result of the study showed that speech characteristics can provide important information on immediate mental and emotional state. It was discovered that the vocal qualities of the depressed patients change significantly when they move into the near-term suicidal state. This proposes the idea of near-term suicidal patients having their own set of speech characteristics that are different from the depressed patients, as a result of the change in articulation and speech production mechanism. Apart from providing tape recordings for the earlier study database, the Silvermans also contributed financially. Their efforts made continuing research on this area possible.

Campbell [42] continued the investigation by doing acoustical analysis to determine the suicidal risk of patients. She studied the statistical properties of the fundamental frequency distribution of 1 female patient and 2 male patients. The speech recordings were made at the time when these patients were considered suicidal and also at times when they were considered non-suicidal, so they were actually acting as their

own experimental and control subjects in this particular study. The statistical properties and variations in fundamental frequency distribution served as the discriminating features in this study and a result of 22.7% misclassification error was obtained. This promising result was the base for further statistical study of various other acoustical properties in speech.

France [9] investigated multiple acoustical properties of speech in his study: fundamental frequency, amplitude modulation (AM), formants, and PSD. Different features were extracted and their discriminating abilities were analyzed. Those features are mean, variance, range, skewness, and kurtosis of fundamental frequencies and amplitude modulation, locations and bandwidths of the formants, as well as PSD ratio analysis similar to our study. The result of the male study has shown that AM and PSD ratio analysis were effective in discriminating suicidal subjects from major depressed subjects. On the other hand, normal subjects can be differentiated from major depressed subjects and high-risk suicidal subjects with formant and PSD features.

Ozdas [10] divided her feature analysis into source domain analysis and filter domain analysis. The source domain method analyzed the effectiveness of vocal jitter and glottal flow spectrum in detecting depression and near-term high risk suicidal risk while the filter domain analysis investigates vocal tract characteristics including the mel-frequency cepstral coeffients (MFCC). The source domain glottal flow spectrum analysis resulted in 75% correct classification between major depressed and near-term suicidal patients. On the other hand, in the MFCC analysis where Ozdas employed a Gaussian mixture model (GMM), 80% correct classification was obtained. Combining the source

domain and filter domain features resulted in tremendous improvement, where a total of 90% correct classification was successfully reached.

Keskinpala et al. [43] did a follow-up to Ozdas' study where a new set of data with a controlled recording environment was investigated. The previous database included recordings from suicide notes left and interviews of patients who had actually attempted suicide. The new set of data was from clinical interviews where a practitioner would have greater control of the recording environment. During these interview sessions, usually both spontaneous (interview) and non-spontaneous speech (passage reading) were recorded. This is the type of data that has been used in future consequent studies including that in this paper.

Yingthawornsuk [11] continued the PSD-based study where he also used features extracted from a new proposed method of GMM spectral modeling in his analysis. In the male reading speech PSD ratio only analysis, four 500 Hz PSD ratios were used to build the classifier and a result of 82% correct classification was obtained between depressed and high risk suicidal patients. When the PSD ratio features were combined with the features from the GMM model, 86% classification accuracy was obtained in depressed-suicidal analysis for both male and female interview speech. Reading speech classification produced 88.50% and 90.33% for male and female subjects respectively. These accuracy rates obtained in the analysis of integrated features were obtained by the statistical cross validation approach.

3.3 Significance of the Paper

Suicide is a major health problem that has caused a lot of deaths in Unites States as proven by the facts in the introduction section. In order to save lives and protect human beings, more effort is needed to prevent suicide. The conservative method to assess a patient's suicidal risk is laborious and time-consuming, urging a need for additional tools that can aid and expedite the risk assessment process. Previous studies have shown that vocal characteristics can reflect the psychological state of patients and can be used as cues for determining suicidal risk. Therefore, studying the acoustic features extracted from the speech of depressed and suicidal patients could lead to a development of an objective diagnostic tool that can assess suicidal risk in a short amount of time. This tool can be used to aid physicians in making quick but precise clinical judgments on potentially suicidal patients.

This study represents a small but significant effort in the development of the desired diagnostic tool. The main focus of this paper is the analysis of acoustic features extracted from the speech of patients to determine whether the high-risk suicidal patients can be distinguished from the depressed patients.

CHAPTER IV

METHODOLOGY

Our original database consisted of four different categories: depressed patients, near-term high-suicidal risk patients, patients who are remitted from depression, and patients who had suicide ideations. For simplification, these groups are called depressed, high-risk, remitted, and ideation respectively. Out of all groups, only the first two groups mentioned are used in the experiment, as we are focusing on the relationship between depressed and high-risk in this paper. From this point forward, whenever the database is mentioned, it excludes the ideation group and the remitted group.

For each patient in every category, there are two types of speech samples: speech samples from an interview which represent spontaneous speech, and speech samples from a text-reading session which represent automatic speech. Usually both types of samples were obtained during the same meeting time between the patient and the interviewer. The interviewer would have some questions for the patient to answer and then, after the interview session is finished, the interviewer would ask if the patient could read a standardized text called "The Rainbow Passage" [27]. This passage is very widely used in language studies, articulation trainings, and many other parts of speech science. The reason for its popularity is that it is phonemically balanced (some say phonetically balanced), where the ratios of assorted phonemes mirror the ones in normal speech (A phoneme is the smallest unit of sound that forms meaningful variations between utterances [28]). The passage also contains all the usual sounds in spoken English. By

doing both interview and reading, different kinds of speech were able to be extracted and used in the risk analysis.

The interview and the text-reading sessions were conducted in the Vanderbilt University Psychiatric Hospital with its medical personnel as interviewers. All sessions were conducted with consent from the patients where their identities and privacy were guaranteed to be protected. Some patients declined to do the reading session, which explains why the number of patients is different for the two sessions, as shown in Table 1. These three different categories of patients were specified by experienced physicians, where the decision on a patient was made after a procedural analysis.

Table 1: *The sample size for each category of patients.*

Group	Number of patients	
	Interview	Reading
High-risk	12	10
Depressed	20	18

Because these data have been accumulated over time since 2003, the same audio acquisition system has been employed to ensure consistency and for convenience. The audio acquisition apparatus used consists of a Sony VAIO laptop, audio signal software and a microphone. The laptop specifications are as follows: Pentium IV 2GHz CPU, 512 Mb memory, 60 GB hard drive, 20x CD/DVD read/write unit, 250 GB external hard drive, and Windows XP OS. The software used are the ProTools LE digital audio editor

and Digital Audio Mbox for audio signal acquisition, while the microphone used is the Audix SCX-one cartiod. Before the interview, an experienced practitioner will setup the audio acquisition apparatus properly to ensure clean recording. The patient was asked to count from one to a certain number (usually more than ten) at a normal speaking rate and while this happened, the practitioner adjusted the recording system accordingly to make sure that the recording volume was consistent with previous recordings.

4.1 Data Pre-processing

A 32-bit analog to digital converter with 44.1 kHz sampling rate was used to digitize the speech samples. The speech samples were then edited using Audacity, where the long pauses (silent period that is more than 0.5 seconds) and unwanted sounds such as the interviewer's voice, the sound of coughing or sneezing, and background noise of people talking were removed from the raw audio. To avoid abrupt transitions between speech segments that can cause unwanted spurious frequency artifacts, the starting point and the ending point of each segmentation were done at zero crossings where no speech is present. Basically, the raw audio files were edited to eliminate unwanted sounds and silent periods, leaving a clean audio ready for further processing and analysis. The resulting audio data was then processed through a voiced-unvoiced detector that separates them into voiced data and unvoiced data. Only the voiced data was kept while the unvoiced data was removed. Then, the voiced data was detrended by subtracting the mean signal, and the output was divided into 20-second segments. The last segments that are less than 20-seconds were removed from the database. The whole process up to this point is called pre-processing [11].

4.2 Extracting Features

The features chosen for this study are PSD ratios as its reliability has been shown in previous studies [9, 10, 11, 15]. A simple method of calculating the periodogram was used to estimate the PSD. For every patient there was a set of 20-second segments. The number of 20-second segments varies depending on the length of the original interview and reading session. The longer the interview and reading session, the more segments that patient would have.

Each 20-second signal segment of a patient is divided into frames using MATLAB with 40-millisecond non-overlapping windows. As a result, 500 frames with size of 1764 points were obtained for each segment. Next, the PSD was calculated for all 500 frames. For our purpose, we only kept the PSD region spanning from 0 to 2000 Hz (considered as the total PSD region) for analysis as this range was typically found to provide enough required acoustic information. Eight PSD bands based on frequency segments of the same size were extracted from the total PSD region, which means each band would cover a span of 250 Hz. Only the first seven bands were used from this point forward. The energy of each band and the total energy from 0 Hz to 2000 Hz were then calculated by finding the area under the curve. The mean energy for each band was calculated across 500 frames as well as the mean energy of the total PSD. After that, the energy ratio of each band to the total PSD was calculated by dividing each band's mean energy with the mean energy of the total PSD. The final result is seven energy ratios denoted as PSD_1, PSD_2 until PSD_7. The procedure to extract PSD features (modified based on [11]) is summarized as follows:

1) Extract only the voiced part of each patient's speech sample by running the original audio through a voiced/unvoiced detection filter.

2) Subtract the mean signal from the voiced speech signal to detrend it and separate the output signal into 20-second segments.

3) Divide each 20-second segment into 500 frames with a 40-millisecond non-overlapping window.

4) Calculate the PSD of each frame with the periodogram method.

5) Divide the PSD region within the frequency range of 0-2,000 Hz into eight equal 250 Hz bands and only use the first seven bands in the next steps.

6) Calculate the energy (area under the PSD curve) in 0-2,000 Hz range and also the energy in each 250 Hz band.

7) Calculate the energy ratio of each 250 Hz band to the total energy from 0 to 2,000 Hz.

8) Repeat step #4 to step #7 until all 40-msec frames of the signal have been analyzed.

9) Calculate the mean energy ratio of each band across all 500 frames for the present 20-second segment and then store them for further analysis.

10) Repeat from step #3 until all 20-second segments have been analyzed for the current patient.

11) Repeat from step #1 for the next patient's speech sample.

Figure 8 shows the flowchart of PSD feature extraction from a speech sample.

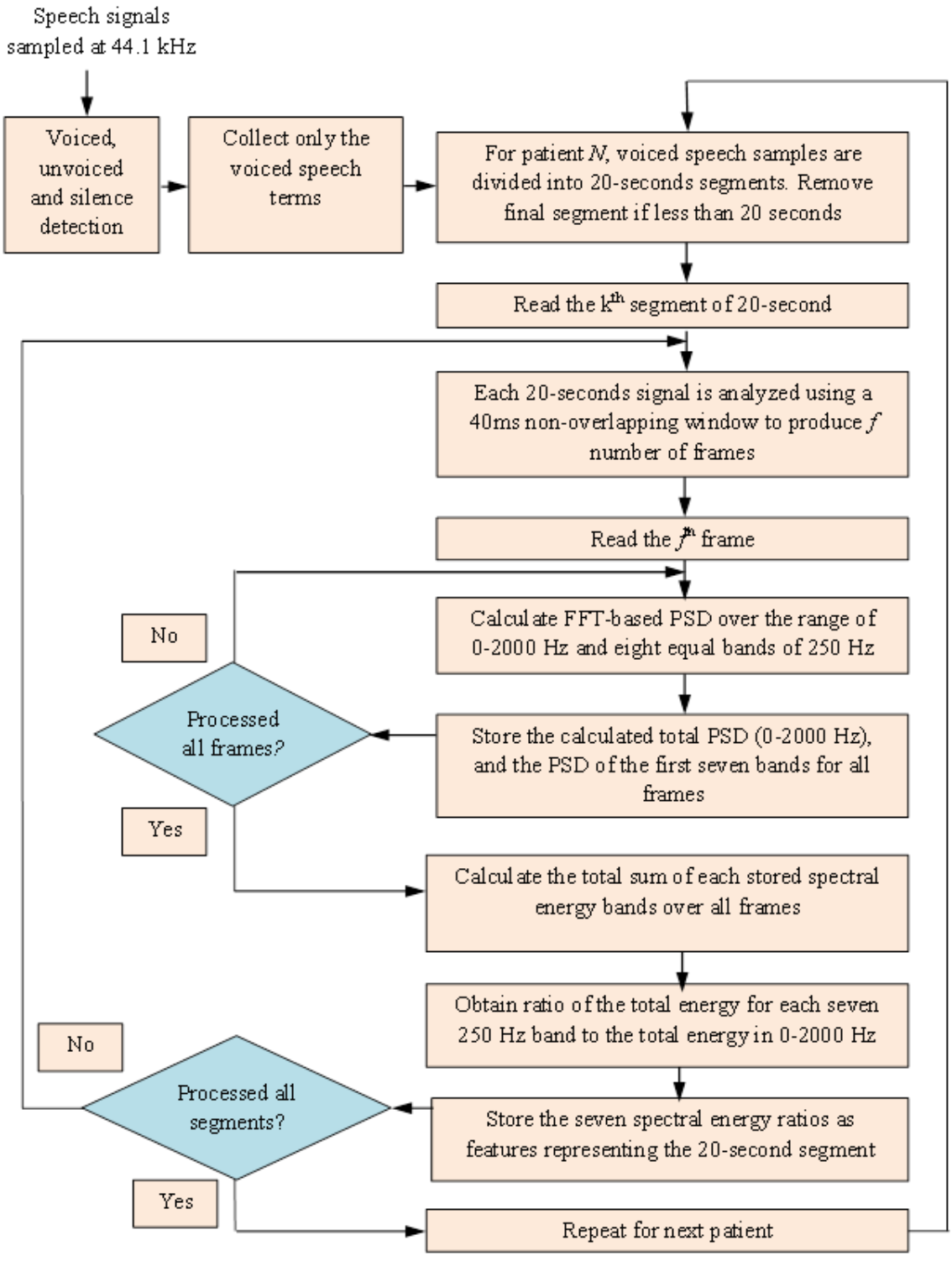


Figure 8: *Flowchart of PSD features extraction.*

4.3 Statistical Analysis of Features

Seven acoustic features (PSD_1 through PSD_7) that have been acquired from each 20-second segment of a patient were stored in a 1x7 row-vector. For each patient, there are a certain number of these row-vectors depending on the number of segments. These row-vectors are stacked according to their sequence of segments. The row-vector that was produced by the first 20-seconds segment would be at the top and the row-vector that was produced by the next segment would be under the first row-vector. The same process was done until the last row vector that was produced by the last 20-seconds segment, which would reside at the bottom of the stack. Doing this for all patients would produce a representative matrix NxM for each patient, where N is the number of vectors (representing the number of 20-seconds segments) for a particular patient, and M is the number of PSD ratios which in our case would be seven for all patients. The accumulated total number of vectors in our database is listed in Table 2.

Table 2: *The accumulated total number of vectors in database.*

	Number of vectors	
Condition / Speech	Depressed	High-risk
Interview	194	77
Reading	42	26

The statistical classification method used in this study is Fisher's Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA) that are represented by the linear and quadratic classifier included in the MATLAB Statistics

Toolbox under the command “classify”. These two classifiers are derived from a Bayesian classifier and the only significant difference between them is that in QDA, the covariance is assumed to be different for every class, while the covariance is assumed to be the same for all classes in LDA. Generally in statistical data analysis, QDA allows more flexibility for the covariance matrix and as a result, it may fit the data better than LDA. However, since the covariance matrices are different (which means it has more than one covariance matrix), the number of parameters to estimate increases. In QDA, more classes means more covariance matrices and this might not be the best option in the case of many classes with a few sample points, because it can be computationally expensive and numerically unstable. In short, there is a trade-off between having a simple model (represented by LDA) and fitting the data well. Sometimes a simple model can produce the same result as a complicated model does. Even if the simple model does not fit as well, it might be better for the data because it is more robust, faster, and computationally cheaper [29].

In this study, both LDA and QDA were used in order to see which one would give the best result. There were three statistical analysis approaches in the implementation of LDA and QDA in this study: cross-validation, jackknife, and the same test-train all-data method. Before diving into these approaches, the meaning of “training and testing” data should be explained as they are basic terminology in statistical classification.

Training data is a set of data that are already labeled with their own corresponding classes (which means we know which data belongs to which class), that are being used in a classification process. On the other hand, testing data is an unlabeled set of data, which we want to classify based on their relationship with the given training data, in a

classification process. The relationship between these two types of data, whether they belong to the same class or not, is determined by the parameters describing that particular class. In order for the classifier to make the comparison, the number of describing parameters must be the same in the testing data and in the training data. In our database, all the data are pre-labeled with their class, whether they belong to depressed or high risk. Depending on the approach, we could "unlabel" some or all of these data, making them into a testing data set in the classification process. Based on the basic concepts and terminology explained, the different approaches are described next.

The same test-train all-data method, as the name suggests, is using the same data in the training data set and the testing data set. All available data are used in both testing and training in this method. Because of that, the result of the classification would be the same regardless of how many times it is being done. Since the same data is used for training and testing, the results are typically overly optimistic.

The cross validation method separates all data into desired proportions of testing and training data. For example, a researcher can pick 20% of all data as testing data, and the remaining 80% as the training data. These proportions of data are randomly sampled from all the available data each time the classification process is done and because of that, the result would be different for each iteration. To get proper results, the classification is done repeatedly in many iterations and the average result is obtained as the final output.

The jackknife method, also known as the hold-one-out method, can also be considered as a subset of the cross validation method but with a few differences. It uses only one sample as the testing data and the remaining samples as training data. In our

case, during each classification process, one patient was chosen for testing and the others would make up the training data. One by one each patient would be classified until there were no more patients to classify, and then the results are accumulated. Using the Jackknife method, the randomness associated with the cross validation method is not present.

All three resampling methods were implemented and coded with MATLAB. For the classification process (LDA and QDA), the built-in MATLAB function "classify" was used and the detailed documents regarding how it can be used can be found on the MathWorks website [30].

CHAPTER V

RESULTS AND ANALYSES

5.1 Result Evaluation

Yingthawornsuk [11] in his work analyzed three out of four bands of PSD ratios and he obtained a good classification result separating different groups of patients. He also concluded that the more bands used, the better the results. Continuing from his observation, this paper investigates the effect of using eight bands (i.e., eight PSD ratios) in discriminating between depressed and high-risk suicidal female patients. The spontaneous speech (Interview) and the automatic speech (reading) are analyzed separately.

5.1.1 Classification of Spontaneous Speech

In our investigation of the spontaneous speech, our analysis is divided into two: classification including all available patients and classification by excluding some patients. The reasons for this division are explained later on in this chapter.

5.1.1.1 Analysis of All Data

The mean and standard deviations of all PSD ratios are listed in Table 3. The general trend shows that the band mean decreases when going from depressed to high-risk, except for PSD_1 where the opposite happens, and PSD_6 where the value stays the same. This shows that more energy from the overall spectrum resides within 250 Hz for

high-risk suicidal speech compared to depressed speech. Generally, above 250 Hz, the energy level of suicidal speech is low compared to depressed speech.

Table 3: *PSD ratio mean and standard deviation for high risk and depressed patients.*

	Depressed		High-risk	
	Mean	Std. Dev.	Mean	Std. Dev.
PSD_1	0.382	0.212	0.484	0.140
PSD_2	0.377	0.118	0.366	0.100
PSD_3	0.171	0.100	0.098	0.052
PSD_4	0.040	0.025	0.025	0.017
PSD_5	0.010	0.007	0.008	0.008
PSD_6	0.006	0.005	0.006	0.007
PSD_7	0.007	0.005	0.005	0.005

Research on statistics have shown that, to obtain reliable results, the minimum number of subjects providing useful data for analysis should be five times the number of the variables being analyzed [44]. Using all seven PSD ratios would give better results but our sample size (i. e., 32, the total number of depressed and high-risk patients) is not big enough to be adequately analyzed by the seven variables. Using only one PSD ratio on the other hand would produce low accuracy results. We decided to analyze the results of not more than three bands combined together to compensate between dimension size and accuracy. We examined the performance for all possible 2 and 3 band combinations.

The result of same test-train all data can serve as an indication to whether the cross validation and jackknife method may give good results. The assumption is that the

cross validation result and the jackknife result will give lower correct classification rate than the same test-train all data method because more known information about the data is present in the testing data set. However, the same test-train all data approach gives a measure of the separability of the data. Therefore the first classification analysis is done with the same test-train all data approach. However, the outcome is not as convincing as what we have expected. The top three results using linear and quadratic classifiers are shown in Table 4. Note that the "All" row is the overall correct classification rate, while the "High Risk" row denotes sensitivity, which is the ability of the classifier to correctly classify high-risk suicidal patients from depressed patients. The "Depressed" row is the specificity of the classification, which is the rate of correctly identifying depressed patients among all high-risk suicidal patients.

Table 4: *Percentage (%) of correct classifications between depressed and high-risk suicidal patients.*

Band	PSD_3, PSD_5, PSD_6	PSD_1, PSD_2, PSD_6	PSD_4, PSD_6, PSD_7
Classifier	Linear	Quadratic	Linear
All	69.00	69.74	70.48
High-risk	79.22	77.92	81.82
Depressed	64.95	66.49	65.98

We decided that there may be data outliers, and that we might achieve better results by eliminating some patients' data. These data outliers may have affected the result. One possible explanation is the existence of a subpopulation whose vocal qualities

do not reflect the actual condition diagnosed by physicians. For example, a patient who was diagnosed as high-risk suicidal by a physician but the way the patient speaks would always have closer resemblance to that of depressed patients, causing the patient to be misclassified. Another possibility is damage to some of the organs involved in speech production such as the vocal cords, where they cannot accurately reflect the emotional condition of the patient, leading to misclassification. However it is important to note that these are just possibilities and we do not have enough information to make any conclusion.

In an effort to find the possible outliers, we proceed by investigating deeper into the best same-test all data result which is given by bands 4, 6, and 7. These bands gave the highest overall and high-risk classification percentage accuracy. Using the same bands, we run the cross validation a hundred times to obtain an error histogram. Our cross validation approach takes 3 patients randomly from each category (depressed and high-risk) and uses them as the testing data set. The proportions of testing and training data sets from the overall data are about 19% and 81% respectively. From this process, we recorded the number of times that a particular patient shows up in the testing data set and what is the misclassification error rate. These two parameters make the error histogram. The error histogram of high-risk patients that resulted from the cross validation of the PSD_4, PSD_6, and PSD_7 combination is shown in Figure 9.

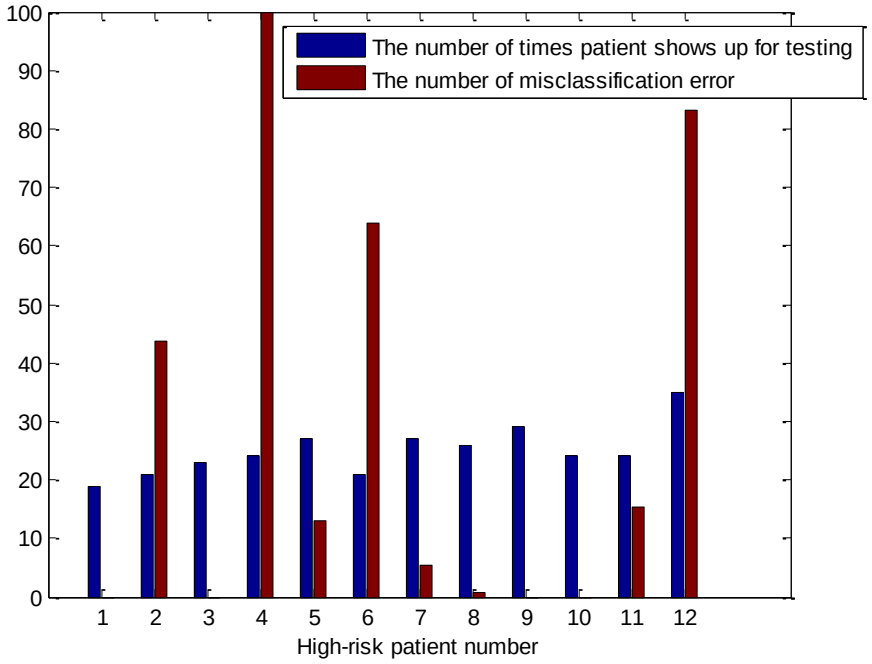


Figure 9: *Error histogram of high-risk patients resulted from cross validation of PSD_4, PSD_6, and PSD_7 combination (spontaneous speech).*

As we can see clearly from the error histogram above, the two high-risk patients with the most misclassification errors are patient 4 and patient 12. The error histogram of depressed patients is not shown here because none of the patients has a clear distinction in terms of having significantly more errors than the others. Therefore we can only pick possible outliers from the high-risk patients based on the error histogram, and those are patient 4 and patient 12.

We proceeded in the effort of detecting possible outliers by using the jackknife method. The jackknife method is run for the same bands combination and the result was observed. Table 5 shows the number of total vectors, correctly classified vectors, and misclassified vectors as a result of jackknife classification.

Table 5: *The number of total vectors, correctly classified vectors, and misclassified vectors as a result of jackknife classification.*

Depressed patients	Total	Correct	Wrong
Patient 1	7	5	2
Patient 2	5	5	0
Patient 3	7	3	4
Patient 4	17	16	1
Patient 5	28	21	0
Patient 6	13	0	13
Patient 7	6	0	6
Patient 8	13	0	13
Patient 9	9	3	6
Patient 10	10	10	0
Patient 11	10	8	2
Patient 12	10	8	2
Patient 13	5	0	5
Patient 14	4	4	0
Patient 15	8	8	0
Patient 16	15	12	3
Patient 17	9	7	2
Patient 18	3	1	2
Patient 19	9	6	3
Patient 20	6	6	0

High-risk patients	Total	Correct	Wrong
Patient 1	5	5	0
Patient 2	6	4	2
Patient 3	11	11	0
Patient 4	6	0	6
Patient 5	4	4	0
Patient 6	5	2	3
Patient 7	2	2	0
Patient 8	5	5	0
Patient 9	5	5	0
Patient 10	4	4	0
Patient 11	6	6	0
Patient 12	18	3	15

Depressed patient with most misclassified vectors

High-risk patient with most misclassified vectors

The classification of depressed patients using jackknife approach produced a useful result. From Table 5, we can observe that patient 6 and patient 8 have the most vectors classified wrongly. These two patients can now be considered as possible outliers for the depressed group. The result of the jackknife approach for high-risk patients also gave an encouraging result where patient 4 and patient 12 have the highest number of vectors with wrong classifications. This agrees with our previous high-risk patients'

analysis using the error histogram method. Having four overall possible outliers, we proceeded by redoing the classification process excluding the four patients.

5.1.1.2 Analysis of Data Excluding Four Patients

The mean and standard deviation were calculated and recorded again after removing the four patients. The result is shown in Table 6.

Table 6: *The mean and standard deviation of PSD ratios after four patients were removed.*

	High-risk		Depressed	
	Mean	Std. Dev.	Mean	Std. Dev.
PSD_1	0.509	0.151	0.337	0.178
PSD_2	0.360	0.109	0.397	0.100
PSD_3	0.099	0.061	0.188	0.097
PSD_4	0.017	0.009	0.045	0.024
PSD_5	0.004	0.003	0.011	0.007
PSD_6	0.003	0.002	0.006	0.005
PSD_7	0.003	0.001	0.007	0.005

The trend of PSD ratio mean after four patients have been removed is mostly the same as when all patients' data were used where only PSD_1 decreases going from high-risk to depressed while others increase. However, the mean difference between each PSD ratio value is bigger after the four patients have been removed than before, as demonstrated in Table 7. Based on this fact alone we can at least already predict that the separation will probably be better if we redo the classification.

Table 7: *Mean difference for each PSD ratio when going from high-risk to depressed.*

	Before removing 4 patients	After removing 4 patients
PSD_1	-0.102	-0.172
PSD_2	0.011	0.036
PSD_3	0.073	0.089
PSD_4	0.015	0.028
PSD_5	0.002	0.007
PSD_6	0	0.004
PSD_7	0.002	0.005

We continued by redoing the linear and quadratic classifications and were able to obtain much better results in all three approaches (same test-train all data, jackknife, and cross validation). The quadratic classifier has shown to produce better overall results than the linear classifier. Table 8 shows the best results using 2 band and 3 band combinations for each approach and they were obtained using the quadratic classifier. For 2 bands, the combination of bands 5 and 7 produced the best results while bands 4, 5, and 7 yielded the best results for the 3 bands combination.

Table 8: *Best correct classification results (%) for 2 bands and 3 bands after removing four patients.*

Method \ Bands		**2D (**PSD_5, PSD_7**)**	**3D (**PSD_4, PSD_5, PSD_7**)**
Same test-train all data	All	82.81	84.62
	High Risk	94.34	98.11
	Depressed	79.17	80.36
Cross validation	All	81.43	81.41
	High Risk	86.66	80.13
	Depressed	77.40	80.66
Jackknife	All	80.54	82.35
	High Risk	86.79	88.68
	Depressed	78.57	80.36

The good results obtained with low dimensionality enabled us to visualize the classification. From the classification of the two bands, 5 and 7, we can produce a 2D scatter plot that can help us observe the distribution of data and the line separating the two classes. Figure 10 shows the scatter plot of the bands 5 and 7 all data classification using the quadratic classifier. Figure 11 is a zoomed view of the same scatter plot. Based on these two plots, we can clearly see the separation between high-risk and depressed patients.

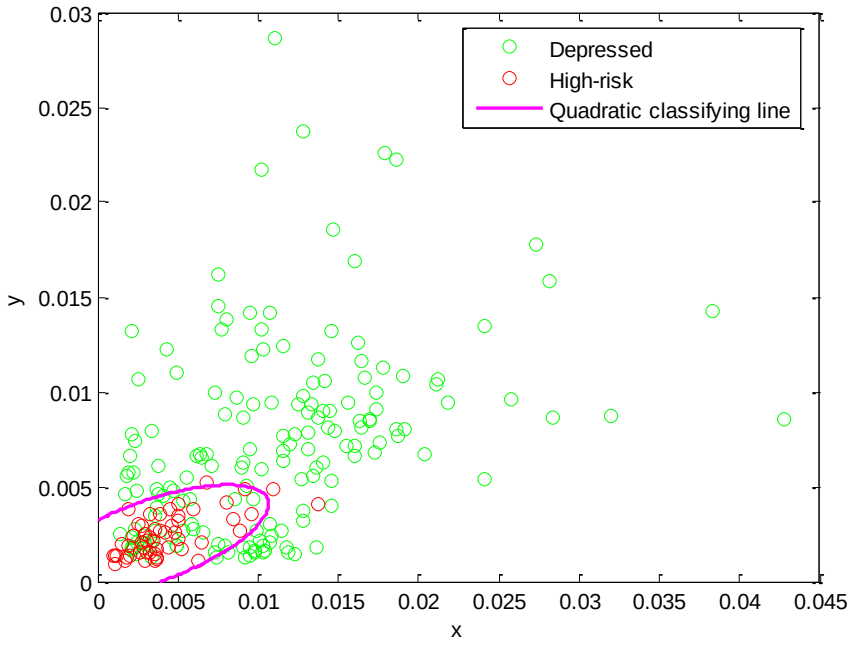


Figure 10: *Scatter plot of band 5 and 7 data classification (spontaneous speech).*

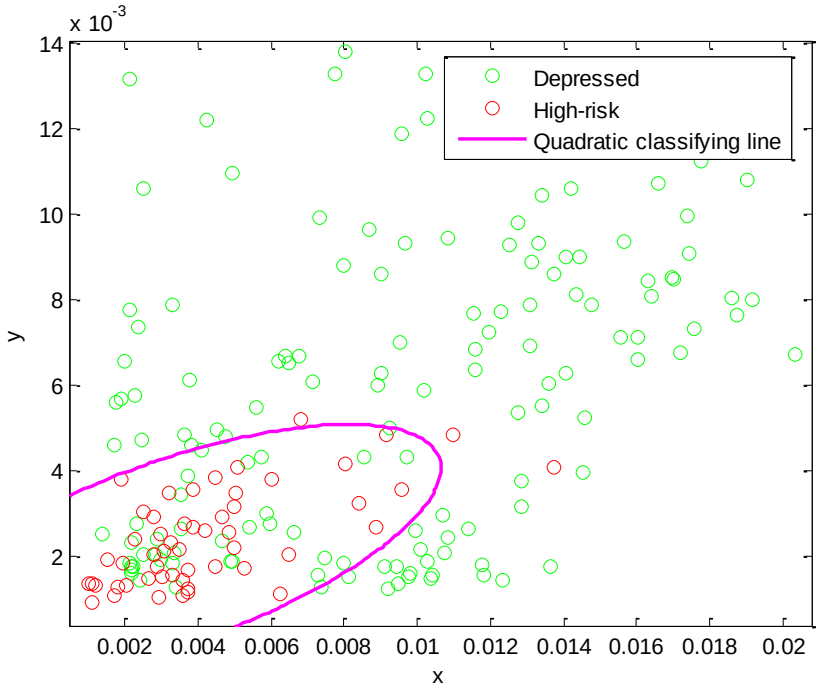


Figure 11: *Zoomed view of band 5 and 7 data classification (spontaneous speech).*

5.1.2 Classification of Automatic Speech

Automatic speech, represented by the passage reading, is speech made without the patient having to think about what to say beforehand. The reason that the automatic speech is being analyzed independently is because the brain mechanism activated during automatic speech is different from the spontaneous speech. Intuitively, patients might not put as much emotion in reading a written passage compared to when they tell their own stories.

The same test-train all data classification produced reasonable results for some of the two and three bands combination. The best results for each of two band and three band combinations are shown in Table 9. The combination of bands 2 and 7 produces the best result in 2D while the 3D best result is obtained from bands 2, 3, and 7 combined together. Because these results fulfill our expectations, we assume that there are no outliers in the dataset thus it is not necessary to remove patients and redo classification as we did in the interview speech analysis.

Table 9: *The best results in 2 and 3 bands combination for automatic speech using the same test-train all data method.*

Band	2D ($PSD_{2,}$ PSD_{7})	3D ($PSD_{2,}$ $PSD_{3,}$ PSD_{7})
Classifier	Linear	Linear
All	76.47	80.88
High-risk	76.92	80.77
Depressed	76.19	80.95

The above results show that the classification is very consistent as it produces nearly identical numbers for all classification, high-risk classification, and depressed classification. The scatter plot for the 2D result is shown in Figure 12.

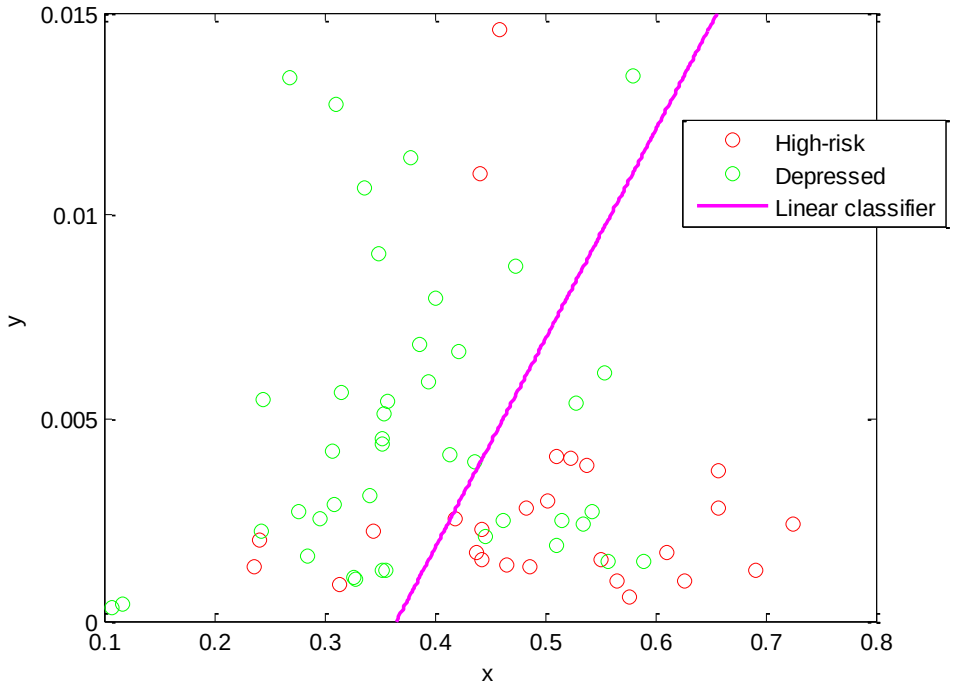


Figure 12: *The scatter plot from 2D (2,7) linear classification (automatic speech).*

The final goal is to obtain appropriate results using the cross validation method because they would represent the data very well. Because there is no need to remove any possible outliers, we can immediately implement the cross validation approach of classification. The top three results are listed in Table 10.

Table 10: *The top three correct classification percentages from cross validation approach in automatic speech.*

Bands	PSD_2,PSD_3,PSD_4	PSD_2,PSD_4	PSD_2,PSD_7
Classifier	Linear	Linear	Linear
All	72.05	73.20	74.20
High-risk	76.69	76.59	75.83
Depressed	65.29	68.02	70.21

Based on the results, the linear classifier works best for automatic speech in contrast to spontaneous speech where the best result is produced when using the quadratic classifier.

5.2 Discussion and Conclusion

The spontaneous speech and automatic speech have different characteristics thus they were analyzed separately. We first observed the results of classification for spontaneous speech using all the available patients' data. The highest percentage of correct classification through the same test-train all data approach is 70.5%. After the removal of some patients' data as possible outliers, classification was redone and the best result of 81.4% correct classification was successfully obtained through the cross validation method. Eliminating some of the patients' data in spontaneous speech analysis enabled us to gain better results. There is a possibility that these data were outliers because their removal improved the classification result tremendously. However, there is simply not enough information (i. e., a large enough database) to make a definite conclusion that these data are truly outliers.

The classification results of automatic speech data were satisfactory. The highest percentage of correct classification using same test-train all data approach is 80.9%. The cross validation method yielded 74.2% correct classification. These results were very reasonable thus there was no need to find and remove outliers.

If the patients' data removed in spontaneous speech classification were not considered as outliers, the overall result from the classification of both types of speech indicates that automatic speech simply has a stronger ability to discriminate between depressed and high-risk female patients rather than spontaneous speech. However, it is highly possible that the removed data were outliers, because including them in the classification yielded result which is below expectation. The possibility of them being outliers is also supported by the fact that there is one patient whose audio file was present in the spontaneous speech dataset but not in the automatic speech dataset, while most of the other patients have audio files in both datasets. The below expectation initial result yielded from spontaneous speech classification might be also caused by patients sometimes switching between depressed vocal patterns and high-risk vocal patterns during the same interview session. This is based on the observations of Drs. Marilyn and Stephen Silverman [5] where they explained that a patient cannot always be in near-term high-risk suicidal state all the time during his or her speech, and they do change between states during the same session. These changes of states might also influence the classification result, as demonstrated in the initial outcome of spontaneous speech analysis.

In the future, more data should be collected in order to obtain better classification results. The strictness of the patient labeling or patient grouping procedure can also be

increased, especially in deciding whether a patient is in near-term high-risk class. The Silvermans' [5] data was obtained under a strict procedure where only patients who have actually attempted suicide were labeled as near-term high-risk suicidal. Improving this procedure may increase the discriminating ability of data because the assignment of any patient to a certain class is more definite and specific. An obvious idea for future work is to include a combination of different types of features as Yingthawornsuk [11] and others did. Vocal jitter, formants, and PSD are just some of the possible different features that can be combined together in the classification process. Another possibly interesting future work is to conduct a longitudinal study where the same patient is analyzed for a defined period of time which is typically long. The last suggestion for future work is to identify the high-risk period and the non-high-risk periods in an interview session.

As an overall conclusion, automatic speech has more discriminating power than the spontaneous speech, provided that the removed patients are not outliers. If they are really outliers, then both types of speech can be utilized confidently in classifying depressed and near-term high-risk patients. This work has completed its objective by showing that the female speech can indicate high-risk suicidal patients from depressed patients and vice versa. This result can be improved on with future studies to make the conclusion more concrete.

APPENDIX A

SELECTED SPONTANEOUS SPEECH CLASSIFICATION RESULTS

Same test-train all data classification result using the data of all available patients

	Band:	1:7	1:6	1:5	1:4	1:3	1:2	1	2:3	2:4
Same test-train (Linear)	All	71.22	68.63	66.42	65.68	62.36	62.73	58.67	61.62	66.05
	High Risk	76.62	79.22	79.22	79.22	75.32	72.73	66.23	77.92	77.92
	Depressed	69.07	64.43	61.34	60.31	57.22	58.76	55.67	55.15	61.34
Same test-train (Quadratic)	All	74.54	71.96	73.06	67.53	66.42	64.21	63.84	57.20	61.62
	High Risk	89.61	87.01	87.01	77.92	89.61	89.61	84.42	87.01	89.61
	Depressed	68.56	65.98	67.53	63.40	57.22	54.12	55.67	45.36	50.52

2:5	2:6	2:7	2	3:4	3:5	3:6	3:7	3	4:5	4:6	4:7
66.42	68.27	70.48	54.98	67.53	66.79	68.27	69.00	63.84	64.21	64.94	71.59
80.52	79.22	79.22	51.95	81.82	77.92	72.73	74.03	84.42	81.82	76.62	83.12
60.82	63.92	67.01	56.19	61.86	62.37	66.49	67.01	55.67	57.22	60.31	67.01
66.79	70.11	70.48	49.82	60.15	63.47	67.16	67.16	58.67	60.52	63.10	66.42
84.42	80.52	90.91	75.32	87.01	83.12	81.82	85.71	89.61	90.91	87.01	89.61
59.79	65.98	62.37	39.69	49.48	55.67	61.34	59.79	46.39	48.45	53.61	57.22

4	5	5:6	5:7	6	6:7	7	1,3	1,4	1,5	1,6	1,7
59.78	57.93	56.46	63.47	48.71	58.67	58.30	62.73	62.73	59.04	65.68	57.56
74.03	71.43	75.32	80.52	25.97	83.12	79.22	81.82	74.03	66.23	63.64	58.44
54.12	52.58	48.97	56.70	57.73	48.97	50.00	55.15	58.25	56.19	66.49	57.22
54.24	61.62	56.46	61.99	66.79	53.87	60.52	61.25	66.05	64.58	66.42	62.36
83.12	66.23	80.52	90.91	19.48	92.21	72.73	84.42	83.12	84.42	41.56	81.82
42.78	59.79	46.91	50.52	85.57	38.66	55.67	52.06	59.28	56.70	76.29	54.64

2,4	2,5	2,6	2,7	3,5	3,6	3,7	4,6	4,7	5,7	1,2,4	1,2,5
59.78	56.46	52.77	58.30	63.10	65.31	63.10	64.94	58.67	57.93	63.10	64.58
72.73	70.13	54.55	79.22	83.12	77.92	83.12	76.62	74.03	77.92	72.73	79.22
54.64	51.03	52.06	50.00	55.15	60.31	55.15	60.31	52.58	50.00	59.28	58.76
58.67	57.20	68.63	63.84	58.67	61.62	58.67	62.73	55.72	54.24	60.89	67.53
83.12	71.43	31.17	64.94	88.31	84.42	89.61	83.12	90.91	85.71	89.61	84.42
48.97	51.55	83.51	63.40	46.91	52.58	46.39	54.64	41.75	41.75	49.48	60.82

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
66.42	62.73	63.10	64.58	61.99	62.73	64.94	59.41	56.09	57.93	61.62	68.27	67.16
77.92	79.22	80.52	80.52	79.22	80.52	76.62	70.13	62.34	77.92	84.42	72.73	80.52
61.86	56.19	56.19	58.25	55.15	55.67	60.31	55.15	53.61	50.00	52.58	66.49	61.86
69.74	68.63	63.10	66.42	63.84	64.21	68.27	66.05	64.58	63.84	62.73	64.94	60.89
77.92	88.31	85.71	79.22	85.71	81.82	71.43	88.31	61.04	88.31	89.61	81.82	84.42
66.49	60.82	54.12	61.34	55.15	57.22	67.01	57.22	65.98	54.12	52.06	58.25	51.55

3,5,6	3,5,7	3,6,7	4,5,7	4,6,7
69.00	63.84	66.42	63.47	**70.48**
79.22	83.12	80.52	80.52	**81.82**
64.95	56.19	60.82	56.70	**65.98**
66.42	61.25	60.52	64.21	62.73
85.71	87.01	85.71	90.91	89.61
58.76	51.03	50.52	53.61	52.06

Classification result using all three approaches after removing the data of selected patients

a) Linear Classification

Linear	Band:	1:4	1:3	1:2	2:4	2:5	3:4	3:5	3:6	4:5
Same test-train										
	All	75.11	74.21	70.59	75.57	75.57	74.66	76.02	76.02	72.85
	High-risk	90.57	90.57	84.91	94.34	88.68	92.45	88.68	90.57	92.45
	Depressed	70.24	69.05	66.07	69.64	71.43	69.05	72.02	71.43	66.67
Cross-val										
	All		72.55	69.01	70.56		76.49	72.61		74.27
	High-risk		84.89	76.70	77.24		86.29	84.68		91.84
	Depressed		63.60	62.06	63.99		69.58	63.81		62.95
Jackknife										
	All		71.95		72.40		72.85	72.85		71.49
	High-risk		86.79		84.91		86.79	86.79		92.45
	Depressed		67.26		68.45		68.45	68.45		64.88

4:6	4:7	4	5:6	5:7	6:7	7	**1,4**	**1,5**	**2,4**	**2,7**	**3,5**	**4,6**
73.76	78.28	71.49	71.04	71.95		70.59	74.21	70.59		72.85	71.04	74.66
92.45	96.23	92.45	86.79	90.57		94.34	90.57	84.91		96.23	86.79	92.45
67.86	72.62	64.88	66.07	66.07		63.10	69.05	66.07		65.48	66.07	69.05
69.63		72.83	70.25	73.01		76.01					70.86	73.97
88.38		90.52	85.03	90.81		95.40					80.35	92.11
57.01		60.81	60.69	61.36		63.19					63.73	62.97
68.78		71.04										
92.45		90.57										
61.31		64.88										

4,7	**5,7**	**1,2,4**	**1,2,5**	**2,3,5**	**2,4,5**	**2,4,6**	**2,4,7**	**2,5,6**	**2,5,7**	**3,4,6**	**3,4,7**	**3,5,6**
75.11	69.23	75.11	71.49	71.04	75.11	74.21	75.11	72.85		74.66	75.57	71.49
94.34	92.45	92.45	88.68	86.79	90.57	92.45	94.34	86.79		92.45	96.23	86.79
69.05	61.90	69.64	66.07	66.07	70.24	68.45	69.05	68.45		69.05	69.05	66.67
76.26	72.56	74.31			73.12	71.78	73.54			71.84	75.08	
94.32	90.05	84.23			88.53	86.27	92.14			87.81	89.55	
63.69	61.33	65.47			63.35	61.96	61.05			60.25	65.18	

4,5,7	4,6,7	1,3,4	1,4,5	1,4,6	1,4,7	1,5,6
75.11	76.47	74.21	75.11	74.21	76.47	73.76
94.34	96.23	90.57	90.57	90.57	98.11	86.79
69.05	70.24	69.05	70.24	69.05	69.64	69.64
77.19	75.62					
94.11	96.35					
65.99	62.08					

b) Quadratic Classification

Quadratic	Band:	1:4	1:3	1:2	2:4	2:5	3:4	3:5	3:6	4:5
Same test-train	All	76.47	75.57	69.68	74.21	79.19	73.76	76.92	83.71	78.73
	High-risk	96.23	92.45	86.79	92.45	92.45	88.68	94.34	92.45	94.34
	Depressed	70.24	70.24	64.29	68.45	75.00	69.05	71.43	80.95	73.81
Cross-val	All		68.54	64.47	69.37		74.92	72.45		75.62
	High-risk		66.90	66.82	70.77		80.47	70.53		80.53
	Depressed		68.24	61.81	66.84		70.47	72.13		71.83
Jackknife	All		68.33		69.68		71.49	71.49		75.57
	High-risk		67.92		79.25		81.13	71.70		81.13
	Depressed		68.45		66.67		68.45	71.43		73.81

4:6	4:7	4	5:6	5:7	6:7	7	**1,4**	**1,5**	**2,4**	**2,7**	**3,5**
84.16	86.88	73.30	74.21	82.81		72.40	75.11	75.57		72.40	73.76
94.34	98.11	90.57	90.57	94.34		94.34	92.45	88.68		94.34	84.91
80.95	83.33	67.86	69.05	79.17		65.48	69.64	71.43		65.48	70.24
80.63		73.41	75.23	80.97		76.05					70.99
79.45		86.77	84.48	85.10		92.23					74.97
80.82		63.72	69.03	77.80		65.23					66.93
81.45		72.85	72.40	80.54	73.30	72.85	71.49	72.85	71.04		71.49
83.02		88.68	83.02	86.79	88.68	94.34	83.02	79.25	81.13		75.47
80.95		67.86	69.05	78.57	68.45	66.07	67.86	70.83	67.86		70.24

4,6	**4,7**	**5,7**	**1,2,4**	**1,2,5**	**2,3,5**	**2,4,5**	**2,4,6**	**2,4,7**	**2,5,6**	**2,5,7**	**3,4,6**	**3,4,7**
77.83	80.09	**82.81**	74.66	73.76	71.95	81	78.28	80.54	73.76	82.81	74.66	80.54
96.23	94.34	**94.34**	92.45	86.79	84.91	94.34	96.23	96.23	94.34	94.34	92.45	98.11
72.02	75.60	**79.17**	69.05	69.64	67.86	76.79	72.62	75.6	67.26	79.17	69.05	75.00
76.63	78.96	**81.43**	71.40			77.85	76.61	75.25		77.00	72.75	76.63
85.27	83.53	**86.66**	73.23			82.17	80.26	80.90		81.13	78.63	75.60
71.19	73.97	**77.40**	67.95			74.67	74.22	70.54		73.58	68.07	75.66
75.57	78.28	**80.54**				76.47	76.02	76.92		79.64		76.47
86.79	86.79	**86.79**				81.13	86.79	86.79		81.13		81.13
72.02	75.60	**78.57**				75.00	72.62	73.81		79.17		75.00

3,5,6	3,5,7	3,6,7	4,5,7	4,6,7	1,3,4	1,4,5	1,4,6	1,4,7	1,5,6	1,5,7
77.38	81.00	74.21	**84.62**	81.90	75.57	81	78.73	80.54	75.11	82.35
90.57	90.57	92.45	**98.11**	98.11	92.45	94.34	96.23	96.23	94.34	92.45
73.21	77.98	68.45	**80.36**	76.79	70.24	76.79	73.21	75.6	69.05	79.17
71.63	78.65	73.70	**81.41**	77.11		74.35	75.15	76.54	72.59	78.73
74.14	79.89	85.44	**80.13**	86.30		75.42	78.45	80.45	78.78	78.8
68.46	77.01	64.79	**80.66**	71.06		73.02	72.58	72.54	68.33	77.13
72.85	78.73		**82.35**	78.28		74.21	73.3	77.83		79.19
75.47	81.13		**88.68**	86.79		77.36	83.02	86.79		81.13
72.02	77.98		**80.36**	75.60		73.21	70.24	75		78.57

The bolded and underlined numbers denote those that were presented in the result and analysis chapter. The grayed spaces represent results that are not as good as what we were expecting. These results were observed but not recorded.

APPENDIX B

SELECTED AUTOMATIC SPEECH CLASSIFICATION RESULTS

Selected classification result using the data of all available patients

a) Linear Classification

Linear	Band:	1:7	1:6	1:5	1:4	1:3	1:2	1	2:3	2:4
Same test-train	All	79.41	73.53	73.53	79.41	73.53	75.00	51.47	75.00	75.00
	High Risk	76.92	76.92	76.92	80.77	76.92	84.62	61.54	76.92	76.92
	Depressed	80.95	71.43	71.43	78.57	71.43	69.05	45.24	73.81	73.81
Cross-val	ALL						71.04		70.65	72.05
	High Risk						74.23		72.32	76.69
	Depressed						64.95		66.74	65.29

2:5	2:6	2:7	2	3:4	3:5	3:6	3:7	3	4:5	4:6	4:7
73.53	73.53	77.94	73.53	67.65	67.65	67.65	73.53	67.65	64.71	66.18	70.59
76.92	76.92	76.92	76.92	80.77	73.98	80.77	88.46	73.08	80.77	88.46	84.62
71.43	71.43	78.57	71.43	59.52	64.29	59.52	64.29	64.29	54.76	52.38	61.90

4	5	5:6	5:7	6	6:7	7	1,3	1,4	1,5	1,6	1,7
61.76	61.76	66.18	66.18	57.35	61.76	64.71	73.53	70.59	58.82	51.47	64.71
88.46	80.77	88.46	80.77	76.92	84.62	84.62	76.92	80.77	73.08	73.08	76.92
45.24	50.00	52.38	57.14	45.24	47.62	52.38	71.43	64.29	50.00	38.10	57.14

2,4	2,5	2,6	2,7	3,5	3,6	3,7	4,6	4,7	5,7	1,2,4	1,2,5
73.53	69.12	72.06	76.47	67.65	64.71	64.71	61.76	69.12	66.18	77.94	76.47
80.77	69.23	73.08	76.92	69.23	65.38	73.08	84.62	92.31	80.77	76.92	84.62
69.05	69.05	71.43	76.19	66.67	64.29	59.52	47.62	54.76	57.14	78.57	71.43
73.20			**74.20**								
76.59			**75.83**								
68.02			**70.21**								

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
75.00	77.94	73.53	73.53	**80.88**	69.12	70.59	73.53	69.12	73.53	76.47	66.18
84.62	84.62	76.92	76.92	**80.77**	69.23	73.08	73.08	69.23	76.92	76.92	76.92
69.05	73.81	71.43	71.43	**80.95**	69.05	69.05	73.81	69.05	71.43	76.19	59.52
	71.65						70.72		71.75	70.41	
	76.20						70.20		71.14	67.81	
	64.74						69.30		70.69	70.87	

3,4,7	3,5,6	3,5,7	3,6,7	4,5,7	4,6,7	1,3,7
66.18	66.18	61.76	64.71	70.59	67.65	
80.77	76.92	65.38	76.92	92.31	92.31	
57.14	59.52	59.52	57.14	57.14	52.38	
						71.05
						72.01
						68.37

b) Quadratic Classification

Quadratic	Band:	1:7	1:6	1:5	1:4	1:3	1:2	1	2:3	2:4
Same test-train	All	79.41	77.94	79.41	70.59	70.59	76.47	48.53	72.06	69.12
	High Risk	100.0	96.15	100.0	84.62	84.62	92.31	65.38	84.62	84.62
	Depressed	66.67	66.67	66.67	61.90	61.90	66.67	38.10	64.29	59.52
Cross-val	ALL									
	High Risk									
	Depressed									

2:5	2:6	2:7	2	3:4	3:5	3:6	3:7	3	4:5	4:6	4:7
73.53	77.94	79.41	70.59	64.71	67.65	69.12	76.47	63.24	64.71	64.71	75.00
96.15	96.15	100.0	65.38	92.31	96.15	92.31	100.0	80.77	88.46	88.46	96.15
59.52	66.67	66.67	73.81	47.62	50.00	54.76	61.90	52.38	50.00	50.00	61.90

4	5	5:6	5:7	6	6:7	7	1,3	1,4	1,5	1,6	1,7
60.29	55.88	64.71	69.12	51.47	55.88	60.29	72.06	61.76	58.82	51.47	61.76
88.46	80.77	88.46	84.62	92.31	80.77	92.31	80.77	88.46	84.62	76.92	76.92
42.86	40.48	50.00	59.52	26.19	40.48	40.48	66.67	45.24	42.86	35.71	52.38

2,4	2,5	2,6	2,7	3,5	3,6	3,7	4,6	4,7	5,7	1,2,4	1,2,5
67.65	66.18	72.06	75.00	72.06	63.24	63.24	60.29	63.24	66.18	70.59	70.59
88.46	69.23	73.08	76.92	84.62	76.92	76.92	88.46	88.46	80.77	88.46	84.62
54.76	64.29	71.43	73.81	64.29	54.76	54.76	42.86	47.62	57.14	59.52	61.90
			71.02								
			73.43								
			66.23								

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
73.53	76.47	69.12	72.06	75.00	64.71	69.12	69.12	72.06	72.06	67.65	63.24
84.62	88.46	80.77	80.77	84.62	88.46	88.46	84.62	80.77	76.92	80.77	88.46
66.67	69.05	61.90	66.67	69.05	50.00	57.14	59.52	66.67	69.05	59.52	47.62

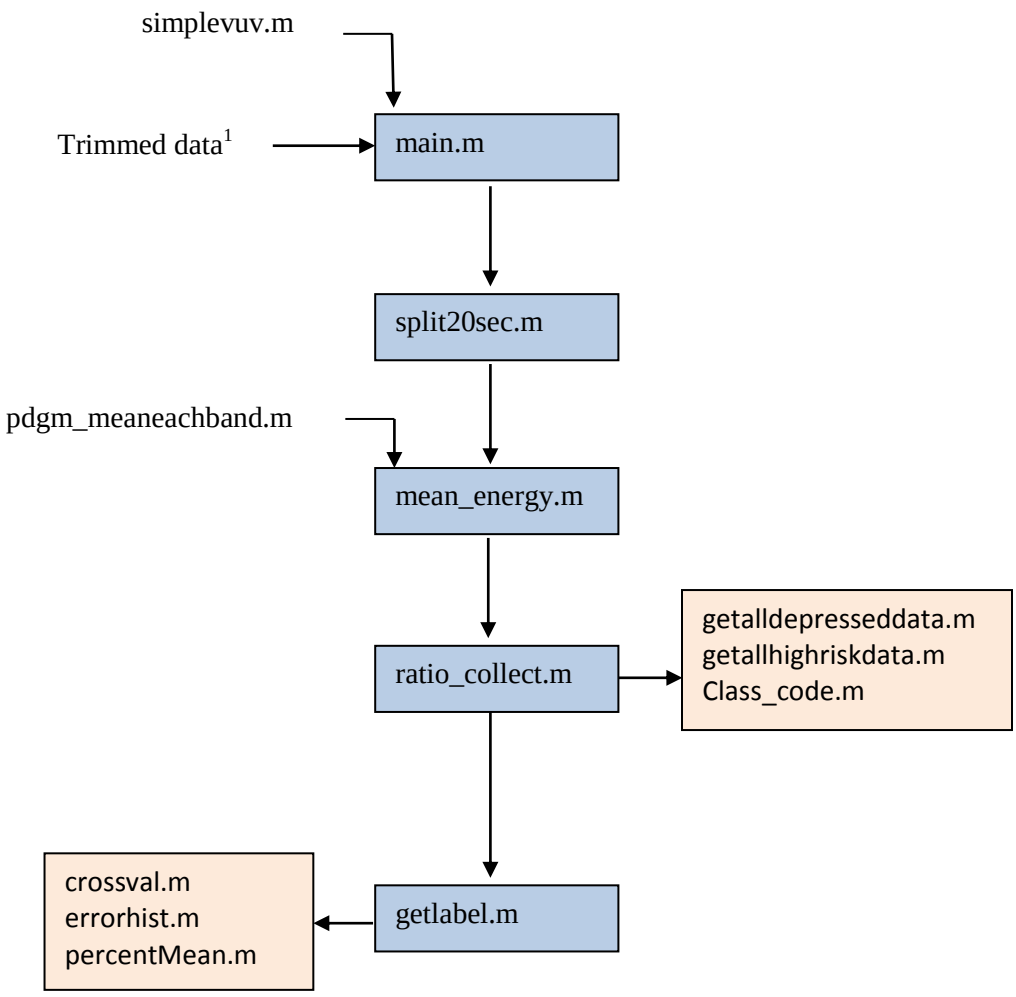
3,4,7	3,5,6	3,5,7	3,6,7	4,5,7	4,6,7
66.18	75.00	73.53	67.65	70.59	70.59
92.31	88.46	80.77	92.31	92.31	96.15
50.00	66.67	69.05	52.38	57.14	54.76

The bolded and underlined numbers denote those that were presented in the result and analysis chapter. The grayed spaces represent results that are not as good as what we were expecting. These results were observed but not recorded.

APPENDIX C

EXAMPLE OF MATLAB CODES IMPLEMENTED

Steps for features extraction and analysis using Matlab:



[1]Trimmed data are speech samples that were collected after removing interviewer's voice, cutting long pauses and removing other irrelevant noises such as door slam, coughing, and sneezing.

simplevuv.m: Identifying voiced and unvoiced data

```
%Code by Mitch Wilkes, modified by Nik Nur Wahidah Nik Hashim and
%Wan Ahmad Hasan Wan Ahmad Sanadi
%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 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);

%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 onlyl)

%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
% m = 1;
% for n = 1:Nwins
%     if unvoiced(n) == 1;
%         unv(:,n) = X(:,n);
%         m = m+1;
%     end
% end
%
% %collecting silence part
% u = 1;
% for t = 1:Nwins
%     if silent(t) == 1;
%         sil(:,t) = X(:,t);
%         u = u+1;
%     end
% end
```

main.m: Collecting voiced data

```
%Code by Nik Nur Wahidah Nik Hashim and
%Wan Ahmad Hasan Wan Ahmad Sanadi
%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
```

split20sec.m: Dividing voiced data into 20-second segments

```
%Code taken from:
%http://www.mathworks.com/matlabcentral/newsreader/view_thread/292920
%Spring 2011
%------------------------------------------------------------------------
---

% function split20sec(fileName,fileNamewav)
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 shortG %turns off scientific notation
    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 desired ranges
        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 example
1-8001
        wavefn=strcat(fileName, num2str(k)); %build filename dynamiclly
        wavwrite([wavefilesplit],Fs,32,strcat('D:\niknwan\3vuv 20 sec
segments only\male interview\',wavefn));
    end;
end
```

pdgm_meaneachband.m: Extracting PSD ratio

```
%Code by Nik Nur Wahidah Nik Hashim and
%Wan Ahmad Hasan Wan Ahmad Sanadi
%Spring 2011
%-------------------------------------------------------------------------
---

function [mean_energy] = pdgm_meaneachband(s,fs)
% [s,fs] = wavread('011706nt1_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);
psd = [];

for i = 1:Xc
    % 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
%
% % %Plot of 1-2000hz
% figure(1),plot(freq(2:2001,:),psd(2:2001,:))
% %
% % %Plot of 1-500hz
% figure(2),plot(freq(2:500,:),psd(2:500,:))
% %
% % %Plot of 500-1000hz
% figure(3),plot(freq(501:1000,:),psd(501:1000,:))
% %
% % %Plot of 1000-1500hz
% figure(4),plot(freq(1001:1500,:),psd(1001:1500,:))
% %
% % %Plot of 1500-2000hz
% figure(5),plot(freq(1501:2001,:),psd(1501:2001,:))

%assigning variables for 4 bands and full range
```

```
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;

allband1area=[];
allband2area=[];
allband3area=[];
allband4area=[];
allband5area=[];
allband6area=[];
allband7area=[];
allband8area=[];
totalallarea=[];

for j=1:Xc;

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

    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(ftotal,psdtotal);
```

```
    band1area=trapz(fr1,psdr1);
    band2area=trapz(fr2,psdr2);
    band3area=trapz(fr3,psdr3);
    band4area=trapz(fr4,psdr4);
    band5area=trapz(fr5,psdr5);
    band6area=trapz(fr6,psdr6);
    band7area=trapz(fr7,psdr7);
    band8area=trapz(fr8,psdr8);

    allband1area=[allband1area; band1area];
    allband2area=[allband2area; band2area];
    allband3area=[allband3area; band3area];
    allband4area=[allband4area; band4area];
    allband5area=[allband5area; band5area];
    allband6area=[allband6area; band6area];
    allband7area=[allband7area; band7area];
    allband8area=[allband8area; band8area];

    totalallarea=[totalallarea; totalarea];

end

meanband1=mean(allband1area);
meanband2=mean(allband2area);
meanband3=mean(allband3area);
meanband4=mean(allband4area);
meanband5=mean(allband5area);
meanband6=mean(allband6area);
meanband7=mean(allband7area);
meanband8=mean(allband8area);
meantotalarea=mean(totalallarea);

energy1ratio=meanband1/meantotalarea;
energy2ratio=meanband2/meantotalarea;
energy3ratio=meanband3/meantotalarea;
energy4ratio=meanband4/meantotalarea;
energy5ratio=meanband5/meantotalarea;
energy6ratio=meanband6/meantotalarea;
energy7ratio=meanband7/meantotalarea;
energy8ratio=meanband8/meantotalarea;

mean_energy=[energy1ratio energy2ratio energy3ratio energy4ratio
energy5ratio energy6ratio energy7ratio energy8ratio];
```

mean_energy.m: Collecting all PSD ratios

```
%Code by Nik Nur Wahidah Nik Hashim and
%Wan Ahmad Hasan Wan Ahmad Sanadi
%Spring 2011
%-------------------------------------------------------------------------
---

function mean_energy

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

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

    [mean_energy] = pdgm_meaneachband(s,fs);
    [path, name, ext] = fileparts(files(i).name);
    filename = fullfile(path, [name []]);
    filename = strcat('C:\Users\Desktop\4psdratio\female interview pdgm
6band\DEP\',filename);
    save(filename,'mean_energy');
end
```

ratio_collect.m: Grouping the PSD ratios of each patient and label based on their category.

```
%Code by Nik Nur Wahidah Nik Hashim and
%Wan Ahmad Hasan Wan Ahmad Sanadi
%Spring 2011
%-------------------------------------------------------------------------
---

function ratio_collect

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

load(files(1).name);
[path, name, ext] = fileparts(files(1).name);
[a b] = strread(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] = strread(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(:,1:7);
        filename = strcat('C:\Users\Desktop\5psd combine\female rea
8band\','d',c);%d for depressed
        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] = strread(name, '%s %s', 'delimiter','_');
c=char(names(length(files)-1));
ratiolist=[ratiolist;mean_energy];
ratiolist=ratiolist(:,1:7);
filename = strcat('C:\Users\Desktop\5psd combine\female rea
8band\','d',c);%d for depressed
save(filename,'ratiolist');
```

getalldepresseddata.m: Loading all depressed data into a matrix

```
%Code by Mitch Wilkes
%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: Loading all high-risk data into a matrix

```
%Code by Mitch Wilkes
%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
```

Class_code.m: Classification by same test-train all data method and jackknife method

```
%Code by Nik Nur Wahidah Nik Hashim,
%Wan Ahmad Hasan Wan Ahmad Sanadi, and Mitch Wilkes (jackknife part)
%Spring 2011
%-------------------------------------------------------------------------
---

clear;clc
getalldepresseddata
getallhighriskdata

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

% two
a=1 ; 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(194,1); zeros(77,1)];

% plot3(data(1:123,1), data(1:123,2),
data(1:123,3),'ro',data(124:218,1), data(124:218,2),
data(124:218,3),'bo');
% [C,err,P,logp,coeff]
[idxl,err,P,logp,coeff] = classify(data,data,lab,'linear');
all = sum(idxl==lab)/length(lab)*100
hr = sum(idxl(1:109)==lab(1:109))/123*100
dep = sum(idxl(110:204)==lab(110:204))/95*100

[idxq,errq,Pq,logpq,coeffq] = classify(data,data,lab,'quadratic');
all = sum(idxq==lab)/length(lab)*100
hr = sum(idxl(1:109)==lab(1:109))/123*100
dep = sum(idxl(110:204)==lab(110:204))/95*100

% figure,plot(data(1:194,3), data(1:194,4),'bo',data(195:271,3),
% data(195:271,4),'ro');

figure,plot(data(1:109,1),data(1:109,2),'ro',data(110:204,1),data(110:2
04,2),'o')
hold on

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.
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 HR-DEP
clear;clc
getalldepresseddata
getallhighriskdata
myData = [Hmean_energy(:,5); Dmean_energy(:,5)];
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(dtotal);
all = hall + dall;

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

% quadratic
htotal=[];dtotal=[];

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(dtotal);
allq = hallq + dallq;

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

getlabel.m: Appending a label column in the data for cross validation and error histogram

```
%Code by Nik Nur Wahidah Nik Hashim and
%Wan Ahmad Hasan Wan Ahmad Sanadi
%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('C:\Users\wanahmwa\Desktop\6crossval\female int
6band\',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('C:\Users\Desktop\6crossval\female int
6band\',Dvec(i).name);
    save(filename,'ratiolist');
end
```

crossval.m: Classification by cross validation and identifying classification error

```
%Code by Nik Nur Wahidah Nik Hashim and
%Wan Ahmad Hasan Wan Ahmad Sanadi
%Spring 2011
%-------------------------------------------------------------------------
---

%  function
[percent_all,percent_hr,percent_dep,percent_allq,percent_hrq,percent_de
pq,countHR,countDEP] = crossval
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
% function
[percent_all,percent_hr,percent_dep,percent_allq,percent_hrq,percent_de
pq] = crossval % for crossval evaluation only
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
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('d*.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
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 = 7;
% class = classify(alltest(:,band), alltrain(:,band), alltrain(:,8));
% classq = classify(alltest(:,band), alltrain(:,band),
alltrain(:,8),'quadratic');
%
% a=3;
% b=4;
% 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=4; b=6; 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 class(j) == 0
```

```
            counterrHR = [counterrHR; 1];
        elseif class(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 class(j) == 1
            counterrDEP = [counterrDEP; 1];
        elseif class(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: Calculating error histogram from cross validation iterations

```
%Code by Nik Nur Wahidah Nik Hashim and
%Wan Ahmad Hasan Wan Ahmad Sanadi
%Spring 2011
%-------------------------------------------------------------------------
---

% function [patientHR, patientDEP, vectorHR, vectorDEP, sumerrorHR,
sumerrorDEP,allpatientHR,percentErrorHR] = errorhist
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 = 12;
numDEP = 20;

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 %12 patient for HR
            if patientHR(l,k) == g
                allerrorHR(g) = allerrorHR(g) + sumerrorHR(l,k);
                cHR(g) = cHR(g) + 1;
            end
        end
```

```
    end
end

% --- ratio/percentage
HRvec = [5 6 11 6 4 5 2 5 5 4 6 18];
allpatientHR = cHR.*HRvec;
percentErrorHR = (allerrorHR./allpatientHR)*100; %if 100%, all wrong

% DEP
for k = 1:testrun
    for l = 1:3
        for g = 1:numDEP %20 patient for DEP
            if patientDEP(l,k) == g
                allerrorDEP(g) = allerrorDEP(g) + sumerrorDEP(l,k);
                cDEP(g) = cDEP(g) + 1;
            end
        end
    end
end

% --- ratio/percentage
DEPvec = [7 5 7 17 28 13 6 13 9 10 10 10 5 4 8 15 9 3 9 6];
allpatientDEP = cDEP.*DEPvec;
percentErrorDEP = (allerrorDEP./allpatientDEP)*100; %if 100%, all
wrong9
```

percentMean.m: Calculating mean classification result from cross validation method

```
 %Code by Nik Nur Wahidah Nik Hashim and
%Wan Ahmad Hasan Wan Ahmad Sanadi
%Spring 2011
%-------------------------------------------------------------------------
---

% average percentage

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] Centers for Disease Control and Prevention (CDC), Web-based Injury Statistics Query and Reporting System (WISQARS), National Center for Injury Prevention and Control, CDC, 2010, <www.cdc.gov/injury/wisqars/index.html>.

[2] Shaffer, D., Gould, M. S., Fisher, P., Trautman, P., Moreau, D., Kleinman, M., & Flory, M., "Psychiatric diagnosis in child and adolescent suicide", Archives of General Psychiatry, vol. 53, pp. 339-348, Apr. 1996.

[3] Roy, A., "Risk factors for suicide in psychiatric patients", Archives of General Psychiatry, vol. 39, pp. 1089-1095, Sept. 1982.

[4] Guze, S.B, Robins, E., "Suicide and primary affective disorders", The British Journal of Psychiatry, vol. 117, pp. 437-438, 1970.

[5] Blumenthal, S.J., "Youth suicide: The physician's role in suicide prevention", JAMA, vol. 264, pp. 3194-3196, 1990.

[6] Schulberg, H.C., Bruce, M.L., Lee, P.W., et al., "Preventing suicide in primary care patients: The primary care physician's role", General Hospital Psychiatry, vol. 26, pp. 337–345, 2004.

[7] Cornelius, J.R., Clark, D.B., Salloum, I.M., Bukstein, O.G., Kelly, T.M., "Management of suicidal behavior in alcoholism", Clinical Neuroscience Research, vol. 1, pp. 381-386, 2001.

[8] Sher, L., "Preventing suicide", QJM, vol. 97, pp. 677–680, 2004.

[9] France, D.J., "Acoustical properties of speech as indicators of depression and suicidal risk", Ph.D. Thesis, Vanderbilt University, Aug. 1997.

[10] Ozdas, A., "Analysis of paralinguistic properties of speech for near-term suicidal risk assessment", Ph.D. Thesis, Vanderbilt University, May 2001.

[11] Yingthawornsuk, T., "Acoustic analysis of vocal output characteristics for suicidal risk assessment", Ph.D. Thesis, Vanderbilt University, Dec. 2007.

[12] Silverman, S.E., "Vocal parameters as predictors of near-term suicidal risk", U.S. Patent 5 148 483, Sept. 1992.

[13] Scherer, K., "Nonlinguistic vocal indicators of emotion and psychopathology", in C.E. Izard (Ed.), "*Emotions in personality and psychopathology*", pp. 493-529, Plenum Press, New York, 1979.

[14] Scherer, K.R., "Vocal correlates of emotional arousal and affective disturbance", In H. Wagner and A. Manstead (Ed.) Handbook of social psychology, pp. 165-197, Wiley, New York, 1989.

[15] Ozdas, A., Shiavi, R.G., Wilkes, D.M., Silverman, M.K., Silverman, S.E., "Analysis of Vocal Tract Characteristics for Near-term Suicidal Risk Assessment", Methods of information in Medicine, vol. 43, pp. 36-38, 2004.

[16] Picture by Beng Tiong Tan, taken from Nikos Tsourakis' webpage, "Speech Production and Acoustic-Phonetics", Technical University of Crete, <www.telecom.tuc.gr/~ntsourak/tutorial_acoustic.htm>.

[17] Flanagan, J.L., *Speech Analysis, Synthesis, and Perception*, Springer-Verlag, Berlin, 1965.

[18] Rabiner, L., Juang, B.H., *Fundamentals of Speech Recognition*, Prentice Hall, New Jersey, 1993.

[19] Picture taken from Charles Feilding's webpage, "College of Santa Fe Auditory Theory", < www.santafevisions.com/csf/html/lectures/017_instruments_IV.htm>.

[20] Berg, J. v.d., "Myoelastic-aerodynamic theory of voice production", Journal of Speech and Hearing Research, vol.1, pp. 227–244, Sept. 1958.

[21] McKinney, J., C., *The Diagnostic and Correction of Vocal Faults*, Genevox, Nashville, 1994.

[22] Shue, Y.,L., "The Voice Source in Speech Production: Data, Analysis and Models", Ph.D. Thesis, UCLA, 2010.

[23] Hutchinson, M., "The Source Filter Model of Speech" Connexions module, <http://cnx.org/content/m12470/latest/>.

[24] Rabiner, L., Schafer R., *Digital Processing of Speech Signals*, Prentice Hall (Signal Processing Series), 1978.

[25] Bradbury, J., "Linear Predictive Coding", Dec. 5, 2000, <http://my.fit.edu/~vkepuska/ece5525/lpc_paper.pdf>.

[26] Howitt, W., "Linear Predictive Coding (LPC)" tutorial, Otolith homepage, <http://www.otolith.com/otolith/olt/lpc.html>.

[27] Fairbanks, G., Voice and Articulation Drillbook, Harper & Row, New York, 1960.

[28] International Phonetic Association, "Phonetic description and the IPA chart", *Handbook of the International Phonetic Association: a guide to the use of international phonetic alphabet,* Cambridge University Press, 1999.

[29] Applied Data Mining class (STAT 897D) website, Lesson 10 - Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA), The Pennsylvania State University, 2011, < https://onlinecourses.science.psu.edu/stat857/node/80>.

[30] MathWorks' MATLAB Statistics Toolbox, Classify function, <http://www.mathworks.com/help/toolbox/stats/classify.html>.

[31] Borden, G. J., Harris, K. S., Lawrence J. R., *Speech Science Primer: Physiology, Acoustics and Perception of Speech*, William and Wilkins, Baltimore, MD, 1994.

[32] Darby, J. K. (Ed), *Speech Evaluation in Psychiatry*, Grune & Stratton, Inc., New York, 1981.

[33] Scherer, K. R., "Speech and emotional states". In J. K. Darby (Ed.), *Speech Evaluation in Psychiatry*, pp. 189-220, Grune & Stratton, New York, 1981.

[34] Scherer, K. R., "Vocal indicators of stress". In J. K. Darby (Ed.), *Speech Evaluation in Psychiatry*, pp. 171-187, Grune & Stratton, New York, 1981.

[35] Scherer, K. R., "Vocal affect expression: A review and a model for future research", Psychological Bulletin, vol. 99, no. 2, pp. 143-165, 1986.

[36] Newman, S., Mather, V., "Analysis of spoken language of patients with affective disorders", American Journal of Psychiatry, vol. 94, pp. 912-942, 1938.

[37] Hargreaves, W. A., Starkweather, J. A., "Voice quality changes in depression", Language and Speech, vol. 7, pp. 84-88, 1964.

[38] Ostwald, P. F., Soundmaking: The Acoustic Communication of Emotion. Charles C. Thomas, Springfield: Illinois, 1963.

[39] Tolkmitt, F., Helfrich, H., Standke, R., Scherer, K., "Vocal indicators of psychiatric treatment effects in depressives and schizophrenics", Journal of Communication Disorders, vol. 15, pp. 209-222, 1982.

[40] Kuny, S., Stassen, H. H., "Speaking behavior and voice sound characteristics in depressive patients during recovery", J. Psychiat. Res., vol. 27, no. 3, pp. 289-307, 1993.

[41] Moore, E., Clements, M., Peifer, J. and Weisser, L., "Comparing Objective Feature Statistics of Speech for Classifiying Clinical Depression", IEEE Int. Conf. (EMBS), pp. 17-20, 2004.

[42] Campbell, L., "Statistical Characteristics of fundamental frequency distributions in the speech of suicidal patients", Master's Thesis, Vanderbilt University, 1995.

[43] Keskinpala, H. K., Yingthawornsuk, T., Wilkes, D. M., Shiavi, R. G., Salomon, R. M., "Screening for high risk suicidal states using mel-cepstral coefficients and energy in frequency bands", 15th European Signal Processing Conference (EUSIPCO), 2007.

[44] O'Rourke, N., Hatcher, L., Stepanski, E. J., "A step-by-step approach to using SAS for univariate & multivariate statistics", SAS Institute and Wiley, 2005.